Philosophical Studies Series

Neil Roughley

Wanting and Intending

Elements of a Philosophy of Practical Mind

Springer

Philosophical Studies Series

Volume 123

More information about this series at http://www.springer.com/series/6459

Neil Roughley

Wanting and Intending

Elements of a Philosophy of Practical Mind

Springer

Neil Roughley
Department of Philosophy
Duisburg-Essen University
Essen, Germany

Philosophical Studies Series
ISBN 978-94-017-7385-0 ISBN 978-94-017-7387-4 (eBook)
DOI 10.1007/978-94-017-7387-4

Library of Congress Control Number: 2015959478

Springer Dordrecht Heidelberg New York London

Printed on acid-free paper

Springer Science+Business Media B.V. Dordrecht is part of Springer Science+Business Media (www.springer.com)

Acknowledgements

In a much earlier incarnation, this book was a ‘Habilitationsschrift’ accepted by the University of Constance. The research out of which that first version grew was made possible by a generous grant from the Swiss National Science Foundation (SNF), to whom I would like to express my sincere gratitude. I’d also like to thank the members of the Zurich Ethics Centre, the Constance Philosophy Department and the members of the latter university’s interdisciplinary Centre for Research on Intentions and Intentionality for many years of exciting and constructive discussions.

I have had the pleasure to discuss these issues with many colleagues over the years, and I should like to follow tradition in thanking all those who have, in varying degrees of intensity, contributed to this book taking on the form it now has. I have profited from discussions on these matters with Colin Allen, Rüdiger Bittner, Ronnie de Sousa, Ezio Di Nucci, Harry Frankfurt, Hanjo Glock, Barbara Guckes, Richard Holton, Geert Keil, Anna Kusser, Anton Leist, Hannes Rakoczy, Peter Schaber, Julius Schälike, Michael Smith, Thomas Spitzley, Peter Stemmer, Ralf Stoecker, Galen Strawson, Bernhard Thoele, Raimo Tuomela and Jay Wallace. Larger blocks of text were read and commented on helpfully by Anja Achtziger, Peter Baumann, Christoph Fehige, Dunja Jaber, Michael Schmitz and Holmer Steinfath. I am particularly grateful to Michael Bratman, Peter Gollwitzer, Al Mele and especially Gottfried Seebass, all of whom took the time and trouble to read versions of the entire manuscript and furnish it with detailed and extremely useful comments.

Essen, Germany
June 2015

Neil Roughley

Contents

Part II Intending

Introduction

This book aims to answer two simple questions: what is it to want and what is it to intend? Because of the breadth of contexts in which the phenomena picked out by the two terms are implicated and the wealth of views that have been expounded in order to account for them, providing the answers is not quite so simple. Doing so requires an examination not only of the relevant philosophical theories, but also of the rich empirical material that has been provided by work in social and developmental psychology. In order to give the reader an idea of what awaits her, this introduction begins with an overview of the resulting theory in broad brushstrokes, together with indications of some key methodological considerations at work in its development. I then provide a chapter-by-chapter summary of the book's structure.

I

According to what is sometimes labelled 'the standard story' of action (Smith 2010; Hornsby 2010), intentional actions result from the causal interaction of beliefs and 'desires' of agents. According to what has been influentially labelled an 'add-on' model of emotion, emotions can be understood as beliefs and 'desires' of agents supplemented by affective components (Goldie 2000, 39ff.; Deigh 2010, 37f.). And according to various 'internalism' requirements in meta-ethics or practical reason, reasons facts or moral judgements entail that the reasons' bearer or the maker of the judgement is, or would under certain conditions be, 'motivated' to act in accordance with the relevant fact or judgement (Finlay and Schroeder 2012).

A whole series of central philosophical issues depends on the roles played by 'desires' or 'motivation' in the relevant contexts. It seems clear that the suitability for the relevant roles of the states picked out by these expressions will depend on how the expressions are precisely to be understood. If the advocates and deniers of any of the positions just named are picking out different sets of conditions in talking about 'desires', then they may well be talking past one another.

Now, the progenitor of the most widely discussed version of the ‘standard story’ of action, Donald Davidson, introduced a complication into the account, after it became clear that talk of beliefs and desires leaves out the clearly distinguishable mental state of intending. Davidson suggested that intending is an unconditional value judgement concerning the desirability of some future action of the agent, in view of his beliefs (Davidson 1978, 98 ff.). In contrast to his conception of the role of beliefs and desires, this claim met with very little positive resonance. However, the diagnosis that there is a mental state here that calls out for analysis was widely accepted and helped to trigger a period of intensive theorising. The position that emerged as the most influential is undoubtedly Michael Bratman’s ‘planning theory of intention’ (Bratman 1987; 1999).

The key move in Bratman’s theory is to characterise intending in terms of two kinds of ‘function’, one causal and one normative. Intentions, he claims, are states that dispose agents to act and deliberate in certain ways and which are subject to norms requiring their embedding in particular structures of deliberation. This proposal involves a specification and extension of the functionalist view of mental states, a view that is adopted by prominent advocates of the standard account of intentional action, such as Michael Smith, to characterise beliefs and desires.

Functionalism, both in the standard causal variant and in Bratman’s causal-cum-normative version, explicitly abandons the aim of naming conditions that pick out individual desires or intentions. It follows that functionalist theories of wanting or intending are unable to name features of either attitude in virtue of which they have the causal or normative consequences the theory pinpoints. However, the good reason for avoiding any such claim, namely, the multiple realisability of mental states by different physical states, does not carry over to the relationship between states characterised in psychological terms.

Thus, intention, for instance, might be definable in terms of other psychological states, where the satisfaction of the psychological conditions named in the definition suffices to explain, or at least contributes significantly to explaining the causal and normative phenomena associated with intention. Moreover, this might all be true even if functionalism were to be the correct theory of the relationship between mind and body. The burden of Part II of this study is to show that it is indeed true.

The theory developed there is inspired by a certain reading of Aristotle’s concept of ‘prohairesis’, which he defines as ‘bouleutikē orexis’, sometimes translated as ‘deliberative desire’. Both the sense in which intention is deliberative and the sense in which it is a ‘desire’ will require extensive explanation. Importantly, my inspiration is ‘a certain reading’, which abstracts both from the fact that Aristotle’s own conception only covers the generation of a subset of those attitudes we think of as intentions and from the question as to whether all forms of Aristotelian ‘orexis’ represent their object under the ‘guise of the good’, as the Thomist tradition of Aristotle interpretation assumes.

Independently of the latter hermeneutic question, I argue that making room for intention precisely involves – pace both Thomist Aristotelians and the later Davidson – making room for an action-controlling attitude that is constitutively independent of evaluation and reasons judgements. Whilst the conception developed here agrees with Bratman on this point, the inspiration of a – perhaps truncated or alternative – Aristotle supports a view according to which the mechanisms decisive for intending are located upstream from the possession of the attitude, rather than, as the functionalist planning theory would have it, downstream. The key, I claim, is the *mark of deliberation* that all intentions bear, rather than the fact that they structure deliberation when we play host to them. As not all intentions are deliberatively generated, the theory's plausibility depends significantly on showing how the mark of deliberation is also conferred where the agent has not deliberated on the specific action intended. The way to do this, I argue, is to develop a disjunctive upstreamist or genetic theory.

Thus viewed, the causal and normative phenomena that Bratmanian functionalism sees as essentially associated with intending are features for which a theory of intending should offer some kind of explanation – although not a conceptual one. As the prime symptoms of intention turn out to be not its causal but its normative consequences, it seems that the theory's litmus test must be its capacity to explain why intention is subject to the requirements of practical rationality and why they have the hold on us that they do. This means that understanding intention also involves answering the question prominently raised by John Broome as to whether rationality is reason-providing (Broome 2013, 204). The genetic disjunctive theory of intention proposes an affirmative answer for practical rationality, an answer that doesn't rely on putative doxastic or cognitive features in order to derive practical from theoretical rationality. The proposal, like Bratman's, sees practical rationality as irreducibly practical. However, whereas normative functionalism cannot explain the relevant requirements and their applicability to intention in terms of intrinsic features of intending, the upstreamist, truncated Aristotelian can, I claim, do precisely that.

The requirements of practical rationality turn out to be grounded in the essence of intentional agency, which is in turn only explicable in terms of the significance of deliberation. One consequence of this view is that young children, who have not yet developed the capacity for minimal deliberation, are not yet intenders, although they may have goals. Intention is 'constructed', that is, a product of our being socialised in a way that generates practical deliberators. Moreover, there are good empirical reasons to think that this constructive process is itself dependent on normative features of our life form, in particular on practices of holding responsible and the concern they generate for anchoring attributability.

According to the theory, then, intention is not, as Searle has claimed, the 'biologically primary' form of practical intentionality (Searle 1983, 36). Children may be programmed to develop goals under relatively minimal conditions of adult care, but intentions seem to require a considerably richer set of practices.

Goals I take to be kinds of wants. More precisely, I take them to be kinds of wants*, where the asterisk indicates that we are dealing with a generic phenomenon that includes everyday wants, but also longings, cravings, mere wishes and intentions. The central claim of the analysis of wanting developed in Part I of this study is that the core of all the attitudes just named is articulated by sentences of the form 'Let it be the case that *p*'. This is a modified version of a claim originally advanced by Anthony Kenny back in 1963, which I think has been unjustly confined to the archives of the philosophy of mind. The claim grounds in the methodological assumption that the basic attitudes of belief and 'desire' are only explicable on the psychological level by means of their expressive articulation, that is, in terms of sentences with analogous structures. This assumption in turn grounds in what appears to be the best explanation of Moore-paradoxicality. Wants*, according to the resulting theory, are essentially *optative attitudes*.

Kenny himself saw an analogy model of the basic mental states as restricted to humans, as a result of their apparently unique capacity to perform actions under particular linguistic descriptions (Kenny 1989, 35 ff.); animal 'desires', he thought, have some other, perhaps purely, dispositional structure, which differs essentially from that of human 'volitions'. In this linguistically transformed Aristotelian model of the mind's striving faculty, it remains mysterious why these behaviour-favouring states of human and non-human animals should both be classified as 'wants'. This reproduces the puzzle of the unity of 'orexis' in Aristotle.

Such a view unjustifiably prioritises the linguistic side of the analogy between the mental and the linguistic. Certainly, we can only develop the analogy because we have access to both mental and linguistic structures. However, the key claim to which the analogy model gives rise is separable from the presence of language in the bearers of mental states. This is the claim that to want* is to set subjective standards, whereas to believe is to represent with a view to meeting the objective standard of truth. Once this is clear, it opens the way to a new view of animal mind, according to which some higher animals may possess wants*, although they don't possess beliefs. This is plausibly the case because the subjective setting of standards is considerably less demanding than developing the concept of an objective standard to which one's mental representations are answerable.

A model of this kind can be seen as a contribution to a more Aristotelian view of the natural continuity between non-human and human animal minds. It should be noted, however, that the model marks the phylogenesis of wanting* as the moment at which a form of subjective normativity comes into being, something which naturalists of certain hues may worry about. It is, however, central to the optative conception of wanting* that wanters* are necessarily subjective standard setters, just as believers are necessarily beings that orient themselves to an objective standard. Attempts by functionalists to avoid this conclusion by giving the direction of fit metaphor a purely dispositional reading fail, as they are either circular or end up changing the subject.

Large portions of Part I of this study consist in a defence of the optative theory against various objections as well as a presentation of its advantages relative to rival conceptions. One reason why wanting* cannot be essentially a matter of being motivated is that we want* many things that are not even candidates to be brought about by our own action. In such cases, the presence of wants* is often indicated by hedonic experiential features of various kinds. Hedonic dispositions have, alongside motivational dispositions, been the historically most popular candidates in terms of which wants might be analysed. The optative analysis allows us to assign both kinds of dispositions their proper place, as frequent, but unnecessary, qualifications of optative attitudes. What we call ‘desires’ in everyday language generally have both features, whereas yearnings, intentions, wishes or whims may have only one or possibly neither.

Two further rival conceptions of ‘desiring’ result from the normative interests in play in discussions of practical reasoning and moral psychology. According to the first, the guise-of-the-good strategy, ‘desires’ are themselves evaluative; according to the second, the mere ascription view, talk of wanting should frequently be given a minimal reading that renders it explanatorily irrelevant. Both strategies designed to insulate action for reasons against the vagaries of contingent motivation underestimate the resources of a model of wanting as subjective standard setting.

A consistent methodological feature of this investigation is the recourse to the results of empirical psychological studies, in particular studies from social, developmental and comparative psychology. Whilst philosophical psychology essentially remains an armchair discipline, it cannot, and should not, avoid making empirical claims. Therefore, the philosophical psychologist should see to it that he is sufficiently informed as to what empirical investigations of his topic have brought to light, whilst bearing in mind that such investigations themselves work with more or less explicit conceptual assumptions.

This study takes the functionalist claim seriously that an overview of the causal environment of mental states is decisive for understanding those states themselves. With this in mind, I collate the various empirical features that cluster around the two phenomena, dubbing them ‘the optative’ and ‘the intentional syndrome’, respectively. For the former, recent work on motivation, as well as the older studies of the ‘New Look’ psychology by Bruner and colleagues, is particularly relevant. For the latter, Heckhausen’s Rubicon model of intention formation, Gollwitzer’s studies on goal intentions and implementation intentions, as well as work on procrastination are significant. Work on goal priming and automaticity by Bargh and others provides important input for the discussion of nonconscious wanting* and nondeliberative intending. Both studies by Wellman, Bartsch and Astington on children’s development of concepts of wants, purposes and intentions and comparative studies on animal mind by Matsuzawa, Kacelnik, Clayton, Brian Hare and Call/Tomasello provide material on borderline or transitional cases, thus giving clues as to how the attitudes are embedded in other features of an agential life form. Finally, Rakoczy’s work on the developmental psychology of normative attitudes supports the thesis that the concern shaping the concept of intention is the provision of an anchor for attributability.

There are philosophers who think that such empirical data are of little or no importance for philosophical claims. There are also philosophers – likely not to be the same ones – who think that philosophical psychology should get by without any substantial reference to consciousness (except perhaps where qualia are in play). In contrast, I claim that neither wanting nor intending can be adequately understood without reference to conscious attitudinising. In the case of wanting*, the tokening of a conscious thought in the optative mode is sufficient, but not necessary; in the case of intending, a conscious thought of this form turns out to be necessary, but insufficient.

A last point concerns normativity. As already remarked, prominent positions claim that 'desire' is a normative attitude, necessarily presenting its object as desire-worthy; intention has been taken to be an all-out value judgement; and intention has also been taken to be part-definable by its subjection to the norms of rationality. I reject all three claims, the first two because both wanting* and intending are essentially noncognitive attitudes, the third because intention is defined in terms of its genesis, not its normative, or other consequences.

However, dimensions of normativity, I argue, are at work in both concepts. As a matter of subjective standard setting, wanting* is essentially a normative attitude, albeit normative in a weak sense that involves no claims on agents, including the want's* bearer, concerning the conditions of the standard's realisation. Unlike wanting*, to which non-human animals with no conception of transsubjective norms may play host, intending, I claim, plausibly requires socialisation into a system involving practices of holding responsible. However, intention isn't only empirically dependent on such normative practices. Its inner structure depends on the general, and generally accepted, normative demand that agents deliberate on the question of their willingness to see themselves as realisers of their motivationally unrivalled wants*. A culture devoid of this demand would not, I will be claiming, have our concept of intention.

II

The investigation is carried out in two parts, dedicated to wanting and intending, respectively, and consisting of five chapters apiece. My broad strategy is the same in each case. I begin by collating and structuring the material that the analysis should be explaining. This consists of the phenomena we encounter in our everyday lives, the intuitions and connections sedimented in our language and the results of empirical research under controlled laboratory conditions. These steps prepare the way for a constructive analysis, whose capacity to account for the decisive phenomena is put to the test and defended against competing conceptions.

Wanting

In the *first chapter*, I approach the topic of practical mind via a brief survey of a number of important positions in the history of philosophy. The founding question for a philosophy of practical mind is raised by Aristotle when he asks what it is in the soul that originates movement. I discuss the answers to this question proposed by Plato, Aristotle himself, Hobbes and Hume, before rounding off the historical survey with a glance at the introduction of the notion of 'pro-attitude' in the last century. The key question put to the various proposals concerns their capacity to give a unitary account of what it is that 'moves' agents. Put in terms of the last of the suggestions discussed: is there a single pro-component that unites the diverse ways of being for something under one genus? And if so, is that pro-component the feature that moves us?

None of the positions surveyed provide adequate answers to these questions. Plato's modular model of the soul turns out to presuppose the notion of a motivational state. The motivational conception of unity proposed by Hobbes depends on his crude materialist conflation of very different features of mental states and Hume's hedonistic position runs into problems of coherence at the moment at which he attempts to explain how affect can be motivationally decisive. The twentieth-century notion of pro-attitudes is so formal as to provide no criterion for membership. The most demanding of the theories surveyed, Aristotle's attempt to establish a criterion of motivational unity under the term 'orexis', also appears, in spite of his explicit aim to the contrary, to be no more than nominal. The relationship between the three forms of orexis, as well as the category's relationship to both the emotions and to 'prohairesis', remains unclear. Here, as in Plato, the attempt to keep ethical and nonethical motivation separate remains a serious stumbling block. Nevertheless, the introduction of 'prohairesis', that is, of a deliberatively transformed motivational state, is a groundbreaking move, which calls out for integration within a systematic motivational theory.

A first sketch of a systematic answer to Aristotle's question is developed in *Chapter 2*. As I take it that this is the founding question not only of a philosophy of practical mind but also of empirical motivational psychology, I approach the topic with an eye to how motivational psychologists circumscribe their discipline. The chapter proposes a skeletal understanding of motivated behaviour, the type of 'movements' with which Aristotle's question is plausibly concerned. On the basis of a non-standard understanding of 'behaviour', I propose a three-factor analysis of the kind of state to which creatures must be playing host when behaving in a manner that can count as motivated. According to the proposal, which urges recognition of 'the Frege point's' importance here, motivational states require a representational component, including a form of primitive self-reference, alongside a marker of attitudinal mode and the functional feature of motivational force. On this basis, I argue that James' ideomotor theory, which has gained considerable popularity in psychology, cannot be a theory of motivated behaviour. I also examine the relationship between motivational force and arousal and offer arguments as to why, pace the Logical Connection Argument, the conceptual relation between motivational states and motivated behaviour does not exclude the relation also being explanatory.

The chapter closes with a discussion of whether there are nonlinguistic animals whose behaviour we should classify as motivated. Various empirical phenomena that are best explained by flexibility in the animals' adaptation of means to ends suggest that the answer is affirmative. Moreover, a case of what I call 'one-way triangulation' provides an indication that, in contrast to what Davidson thought, motivational states may be possessed by a creature in the absence, not only of language but also of doxastic states.

The three-factor conception of motivational states opens the way for a move that severs any necessary connection that may be thought to exist between the ‘modal’ and representational features of motivational states, on the one hand, and the physiological mechanisms brought together under the functional concept of motivational force, on the other. It also allows us to see that other attitudinal features, specifically beliefs, are generally involved when we say someone is ‘motivated’ to do something. Factoring out both ‘energising’ and doxastic components allows us to abstract a ‘purified’ concept I label *wanting**. This is the attitudinal core also present in those compound states generally referred to by everyday terms such as ‘want’, ‘desire’, ‘concern’ and ‘interest’. *Chapter 3* argues, firstly, that wants* are not merely motivational states, as they are also responsible for a whole syndrome of characteristic effects, ‘the optative syndrome’, and, secondly, that they don’t necessarily motivate. Against functionalist positions in the philosophy of mind, I argue that any attempt to define wanting* in terms of motivational or other effects distorts the primarily first-person and irreducibly practical character of desire’s attitudinal core. These points of criticism prepare the ground for a formulation of the requirements on a constructive theory of wanting*. The following two chapters attempt to meet them.

Chapter 4 develops the idea of an expressive explication of the attitudes, which grounds in the claim that there is an essential structural analogy between mental states and linguistic utterances. The strengths of the conception are first demonstrated by showing how it explains the phenomenon of Moore-paradoxical sentences for beliefs. Applied to wants*, it reveals them as essentially optative attitudes, that is, as mental states articulated by utterances of the form ‘Let it be the case that p’. The optative analysis is then confronted with two competing proposals stemming from the field of moral psychology. According to the first, axiological theory, ‘desires’ entertain the same relation to the good as beliefs do to truth. The main argument for the view, Anscombe’s hermeneutic vertigo argument, is shown to conflate the putative incoherence of a non-axiological concept of wanting with the incomprehensibility of an agent’s reasons for wanting. According to the second proposal, the mere entailment view, which revives the main premise of the Logical Connection Argument, talk of ‘desires’ is, in at least certain important cases, simply a way of characterising an action as intentional, a characterisation that makes no substantial contribution to its explanation. I distinguish three reasons for this view and show why none of them justify the claim that wants* are mere ascriptions.

In an appendix to the chapter, the optative analysis is related to the metaphor of ‘direction of fit’. I argue against reductive attempts to rid us of the metaphor, claiming instead that it marks an irreducibly normative feature of attitudinising. At the close, the chapter returns to the suggestion at the end of Chapter 2, that there may be creatures that play host to motivational states without being believers. This possibility turns out to be entailed by the conception of wanting* as the setting of subjective standards, which, unlike the objective standard required by belief, don’t require the capacities for full Davidsonian triangulation.

The *last chapter of Part I* discusses the relations between optative attitudinising, consciousness and affect. The explication of wanting* in terms of its linguistic expression suggests that conscious thoughts of the appropriate form are sufficient for their bearer to be the bearer of the corresponding attitude. I argue that this is indeed so, although no such thought is necessary for a want's* correct ascription. Cases of what I call 'subintentional action', of an agent's motivated inaccessibility to her own motivation as well as experimental evidence from goal priming all demonstrate the lack of any such necessity. Nevertheless, although features of the optative syndrome that we would otherwise be at a loss to explain can warrant the ascription of non-conscious wants*, there can be no strict criteria for their correct ascription.

Only with conscious wanting* do we have such a criterion. For this reason, it remains the phenomenon that allows us to get a conceptual grip on our optative attitudes. This is the case, even though agents may sometimes be mistaken as to whether they 'really want' something, a mistake whose possibility depends on the fact that the latter phrase generally picks out wanting* in conjunction with further features. One such further feature that has frequently been seen as decisive for wanting in one way or another is that of pleasure or displeasure. The chapter closes with a survey of the various relationships between wanting* and affect. I offer arguments as to why, in spite of a long philosophical tradition to the contrary and its revival at the hands of Galen Strawson, none of these relationships qualify as constitutive of the generic concept a theory of wanting should be trying to reconstruct.

Intending

The natural first step on the road to an adequate systematic understanding of intending, taken in *Chapter 6*, is a discussion of its relation to belief. This is natural for two reasons. First, the linguistic means of intention expression have a grammatically assertoric form, and, second, belief may appear to be precisely what needs adding to optative attitudinising in order to generate intention. After discussing forms of intention's expression, and noting that the English language provides different means of expressing intention's formation and its possession, I discuss and reject attempts to understand intention's 'supra-optative' commitment component in doxastic terms, which take the assertoric form of typical intention expressions literally. I also reject the positions of Anscombe and Velleman, for whom intention can be identified with either an epistemic or doxastic attitude. Rather, intention, I claim, in agreement with Bratman and Mele, has only an extremely weak negative doxastic condition. What is decisive is, however, the explanation of this condition. It is here that I follow Aristotle, for whom the doxastic condition on prohairesis derives from the identical doxastic condition on the practical deliberation through which the attitude is generated. The key here is the practical character of practical deliberation, that is, its performance in order to generate an action-controlling attitude. This, I claim, accounts for closely related conceptual and rational doxastic conditions on deliberative intending. These are then distinguished from intending's typical doxastic symptoms, which, according to social psychological studies, may be at least in part caused subpersonally by our decisions or by our planning to implement decisions. I conclude that the discussion of belief, rather than uncovering the completing conditions that transform an optative into an executive attitude, on the contrary, indicates that the step into commitment will have to be explained by different means.

The following *seventh chapter*, on ‘the intentional syndrome’, details intention’s other prominent accompaniments, dividing them into two groups: causal consequences of, and normative requirements applicable to intending. The causal features – motivational strength, pervasion of an agent’s mental life and persistence – may frequently be more pronounced than in cases of other wants*, but they are not specific enough to warrant postulating a distinct kind of mental state. The real challenge for the reductionist is posed by requirements of rationality to which intenders are unavoidably subject. The main burden of Chapter 7 is the development and fine-tuning of explicit formulations of the relevant principles.

Here, I discuss various moves in the recent debates on rationality, arguing for a view of the intention-consequential (*IC*) requirements that is largely wide-scope, although with certain restrictions. I also argue against Wedgewood’s and Broome’s claim that rationality necessarily supervenes on the mind. One reason why the supervenience thesis is false is that perhaps the most basic *IC* requirement concerns executive consistency, i.e. is a principle one of whose relata is not an attitude, but an action. I defend this claim both against the supervenience thesis and the assertion that any such requirement would be conceptual rather than rational. Specific formulations of standards for subordinate intending and for the eschewal of intention-undermining intentions are also proposed. The first involves a rejection of Broome’s restrictive assumption that the doxastic premise should concern the necessity of the agent’s own intending and its effects, rather than that of his performing a certain action. Although the second is close to a formulation of Bratman’s, I take issue with his explanation of the requirement in terms of agglomerativity, which I argue has a mere epistemic role. Finally, I also propose a formulation of a requirement of deliberative intention persistence, which specifies the doxastic conditions under which it is irrational not to uphold one’s intentions. These concern the relative deliberation-conduciveness of the conditions under which the intention was formed and those given at a later point in time. It is significant for an understanding of the structure of intention that this latter principle only applies to deliberative cases.

The last three chapters contain my systematic proposal as to how intentions can be reductively understood whilst accounting for the specificity of the intentional syndrome, in particular whilst allowing us to understand the force of the requirements of intention rationality. The proposal is disjunctive and genetic: I claim that intentions are optative attitudes on which a contextually unique practical status has been conferred, a status that can be conferred by one of two aetiological mechanisms.

The first of those mechanisms is decision, the most salient form of intention acquisition. *Chapter 8* aims to clarify what it is about deciding that enables it to confer that status. I begin here by looking at cases in which deciding goes beyond motivational and evaluative phenomena, cases for which decision theory leaves no room and which are metaphysically overdramatised by existentialism. In order to get a handle on the feature at work in these cases, I compare the different ways deciding and judging depend on at least minimal forms of theoretical inquiry and practical deliberation, respectively. The decisive difference concerns the relationship between the motivations for the relevant form of reflection: whereas it is an essential, internal feature of inquiry that it be powered by a want* which specifies the conditions under which some belief content can count as an answer to the question that inaugurated that inquisitive episode, no such criterial specification of the motivation internal to minimal deliberation can be given. Instead, all that is required here is that the answer to the question that inaugurated deliberation be the deliberator's own answer. An extended discussion of conceptions of the strong ownership of attitudes (Frankfurt, Korsgaard, Velleman) and of the relationship between conclusive reasons judgements and decisions supports this conclusion. After rejecting the claim that decisions might be actions, I conclude with a constructive proposal, according to which a decision is a conscious optative occurrence seen by its bearer as his own answer to the optative uncertainty that triggered an episode of minimal deliberation and which for that reason causes him to desist from further deliberation on the matter.

Chapter 9 then takes us from the analysis of decision to an analysis of intention. I begin by arguing for conditions which ensure the products of certain decisions persist as decisional intentions. I then turn to nondecisional intentions, distinguishing five distinct kinds: spontaneous, gradual, specifying, habitual and judgement-derivative. Nondecisional intentions, it turns out, don't only differ from their decisional conspecifics in the question of their non-subjection to a persistence requirement but also in the matter of their non-subjection to belief constraints. This difference supports the claim that the word 'intention' doesn't always pick out the same constellation of features. I go on to develop the claim that the contextually unique practical status of nondecisional intentions ('being set') grounds in a combination of motivational strength, conscious wanting* and not seeing the question of one's realisation of the want's* content as an 'open question'. My defence of this claim involves arguing for the necessity of each of these three components.

First, of the five types of nondecisional intentions, those nondeliberatively generated from conclusive reasons judgements raise the farthest-reaching questions for a motivational analysis. In a discussion that supplements the parallel treatment of decisional cases in Chapter 8, I argue that they are covered by a broad principle relating various kinds of action-controlling agential states to conclusive reasons judgements.

Second, the claim that nondecisional intending requires conscious wanting* is defended against the suspicion that there are whole clusters of cases in which intention needs nothing of the sort. Here I look at empirical literature on cases in which, as Gollwitzer puts it, action control is 'delegated to the environment', concluding that neither 'implementation intentions' nor habitual intentions get by without conscious optative thoughts. These cases are to be distinguished from routine components of composite actions, which may be goal-directed, but not intended.

Third, the metaphor of not leaving a question open is interpreted as either the non-triggering of, or unresponsiveness to the triggering of the dispositional want* to deliberate that is necessarily in the background of full intentional agency.

At the end of the chapter, the results of the discussion are brought together in a disjunctive definition of intending, which is defended against scepticism about the status of disjunctive definitions and supported by empirical psychological data concerning intention's ontogenesis.

The *final chapter* of the study argues that the disjunctive analysis provides a distinctive, and distinctively plausible explanation of the intention-consequential requirements of practical rationality. Consistent with an analysis that sees intentions as optative attitudes accompanied by only minimal, negative doxastic conditions, I reject explanations of the *IC* requirements that attempt to derive them from belief components of intending. Instead, the disjunctive analysis calls for a noncognitivist explanation of the requirements, which sees them as constraints on an agent's project of self-forging, a project that is in an important sense orthogonal to her responses to reasons. The chapter discusses the most prominent noncognitivist proposal of how such an explanation should be provided, Michael Bratman's grounding of the norms in the demands of self governance. I argue that Bratman's model is vitiated by its normative functionalist framework, which, because it cannot, on pain of circularity, ground its explanation in features of intending, has to postulate a specific reason behind intention's subjection to the requirements. However, because the norms are in place for all intentional agency, they cannot depend on specific aims or reasons, even reasons most of us have, but need to be grounded in intentional agency itself.

According to my constructive proposal, to postdeliberatively opt to perform certain actions is to 'take personal responsibility' for them, thus providing an anchor for their attributability. It is a concern to provide such an anchor, I claim, that has conferred on intention the shape it has. The reason for intending's subjection to the *IC* requirements, I then argue, is that their contravention renders responsibility-taking unintelligible: an agent who intends in violation of the *IC* requirements becomes opaque to himself, thus necessarily losing his hold on his own agency. This fact provides agents with a distinct kind of reason, one that is both stringent and, under certain circumstances, strikingly weak – a conjunction of features which, I argue, suffices as an answer to the scepticism expressed by Broome as to whether rationality can be shown to generate reasons. I then show how the notion of taking personal responsibility extends to cover cases of nondecisional intending. The explanation of this extension is, crucially, in part normative. It grounds in the deeply entrenched demand that persons deliberate on whether they are willing to see themselves as the realisers of their motivationally unrivalled wants*.

It turns out, then, that, although intention is a descriptive concept, the subjection of the full range of its decisional and nondecisional variants to the *IC* requirements presupposes a normative component in the understanding of their aptness for the role that explains that subjection. Indeed, intention is so strongly interwoven with our culture of normative address that it seems extremely unlikely that a creature without such a normative life form might develop a concept with anything like the contours that intention has for us.

Part I
Wanting

Chapter 1
The Question of Motivational Unity: Historical Preliminaries

1.1 Practical Mind: Aristotle's Question

We often think of ourselves as being *moved* to act. This thought seems to imply that our bodily movements are the effects of some kinds of events in us, events which themselves have the power to "move" us. The more technical term we use to capture this idea is "motivation", derived from the past participle of the Latin for "to move", "motus". In *de Anima* (DA 432a18-19), Aristotle wondered "what it is in the soul which originates movement".[1] In doing so, he raised the founding question both of motivational psychology and of a philosophy of practical mind. Whereas for the former the question is primarily empirical, for the latter it is primarily conceptual in character.

Understood in the latter sense Aristotle's question can be reformulated as asking what it is that makes certain of our mental states *potentially motivating*, or *motivational* attitudes: in virtue of what features are components of our psychology apt to move us to act? And what distinguishes them from those psychological elements that are unable to take on that role?

In a second step, the question allows of differentiation: are there *internal divisions* within our potentially motivating attitudes to which correspond differing roles in the production of our behaviour? If so, where are the dividing lines to be drawn? This line of investigation naturally leads to the further question as to the relationship between *unity* and *plurality* among the practical attitudes: are the divisions establishable here best understood as surface phenomena, merely ways of

[1] Although Aristotle's concept of the soul is much broader than the modern notion, including the vegetative and the visceral, he quickly excludes the latter features as irrelevant for the kind of "forward movement" that interests him (DA 432b15ff.).

N. Roughley, *Wanting and Intending*, Philosophical Studies Series 123,
DOI 10.1007/978-94-017-7387-4_1

distinguishing different manifestations of the same *generic* attitude or do various states qualify as motivational through the instantiation of *irreducibly different* properties?

Bound up with this ontological question is the epistemic or methodological one as to what means at our disposal allow a vindication of answers of one or the other kind. Is it possible to establish *criteria* here or must we content ourselves with adducing *paradigmatic cases* that we are able to recognise as falling on one or the other side of the relevant divide?

Either way, an essential step towards establishing the distinctions must be the consideration of the divisions established by the *language* in which we talk about our practical minds or, less reflectively, express our practical concerns. A philosophy of practical mind has to strike a balance here: on the one hand, it needs to develop an adequate awareness of the relevant distinctions provided, or imposed, by the language in which it is approaching the topic. On the other hand, it should retain an awareness of the fact that the divisions established by language are to a significant degree the products of historical and cultural contingencies.

A philosophical treatment of these issues should stay in touch with the divisions of everyday language. Nevertheless, philosophy is neither descriptive linguistics nor sociology of language. Rather, the conceptual questions here concern fundamental features of mental reality that are reflected in linguistic structures, themselves surely more variable than our basic mental apparatus. Note that making this distinction does not entail denying that the genesis of that apparatus is itself up to a point dependent on developmental processes of a cultural and linguistic nature (cf. Shore 2000, 100f.).

If the psychological reality investigated by a philosophy of practical mind is not to be equated with the linguistic structures of a particular community, it is also to be distinguished from the material, physical or physiological reality that underlies it. What used to be called “the mind-body problem” has tended to be the main focus of recent philosophy of mind. As important as this metaphysical issue is, it should not detract from the importance of clarifying the structures of our *inner-psychological reality*. There is certainly nothing to be gained here by a blanket refusal to allow talk of “reality” or “ontology” with reference to the inner-psychological sphere. Whether what we take to be the everyday reality of our intents, concerns, feelings, impulses and emotions is in a metaphysical sense “less real” than that of neurons or electrical impulses is a fairly distant concern. There is clearly a trivial sense in which this question is to be answered affirmatively. But this is of no more relevance for our self-understanding as beings emotionally and actively involved in the world than is a theory of quarks for an understanding of mechanics. Moreover, anyone who reserves talk of the ontology of the mental for neurophysiological states and events faces the question as to why we should stop there and not see as “real” only items individuated in microphysical terms.

In an initial approach to Aristotle’s question, I shall take a (no doubt too) brief look at some of the ways in which it has been tackled and answered in the history of philosophy. Before coming to Aristotle’s own answer (Sect. 1.3), I shall begin with the Platonic model he was rejecting (Sect. 1.2). Both conceptions struggle with the task of bringing together the phenomenal variety of what moves us to act with a unitary explanatory answer. In contrast, the seventeenth and eighteenth-century theories of Hobbes (Sect. 1.4) and Hume (Sect. 1.5) propose reductive explanations according to which one factor can be isolated that is essentially responsible for the movements we call our actions. Although both move beyond Aristotle in pressing for such a unitary form of explanation, they both pay the price of being unable to account for significant features or types of motivational phenomena. In a final section (Sect. 1.6), I turn to the introduction of the notion of "pro-attitude" in analytic meta-ethics and action theory. Although the term was designed to establish unity among potentially motivating mental states, it turns out to do so in complete abstraction from the question of *what* it is that allows their unification under one such term. I conclude that, on the evidence presented here, the question of motivational unity remains a challenge for a systematic philosophy of practical mind, a challenge that can only be met by mobilising resources not exploited by the main conceptions of our philosophical tradition.

1.2 Plato and the Tripartite Practical Mind

In Book IV of *The Republic* (436a-443b), Plato famously divides the soul into three. He draws the lines here entirely in motivational terms, a move that, as Aristotle notes (DA 432a26-b4), is both surprising and implausible if Plato's aim is a topography of the mental in its entirety. He proceeds by applying a principle of identity, a principle that encompasses both perceptual and overt behavioural criteria. However, when he comes to apply it, he does so by picking out purely motivational features. The principle, which has been labelled "the principle of contraries" (Irwin 1995, 204; Bobonich 2002, 223), specifies conditions under which the cause (or recipient) of certain effects cannot count as a single entity. This, Plato claims, cannot be the case where the effects brought about (or undergone) are contraries "with respect to the same thing" (Rep 436b) in so far as they occur at the same time. He doesn't concern himself further with the receptive features of the mind provided for in the principle, focussing exclusively on motivational effects. It is the impulses to approach or avoidance behaviour thus picked out that he classifies as contraries of the sort the principle mentions (Rep 437b).

Because of this restriction in the principle's application, we can say that Plato is effectively working with a principle of exclusion of motivational contraries. The basic idea is that no single psychological capacity could simultaneously move its bearer both to approach and to avoid the same thing. Plato cites examples which show convincingly that agents can be internally conflicted as regards an action they are considering. We can be drawn to do something out of hunger or greed, whilst resisting the action on the basis of a value judgement or because of the shame the thought of the action triggers. The conjunction of these phenomena and the motivational exclusion test lead Plato to argue that agents are moved by one of three potentially conflicting faculties: "to logistikon" ("reason"), "to thumoeides" ("spirit" or "anger") and "to epithumētikon" ("appetite" or "felt desire").

Plato's tripartite model is, among other things, an attempt to provide an explanation of akrasia that does not simply explain it away as Socrates had done in the *Protagoras*. Against the Socratic view that intentional action is necessarily the result of an optimising value judgement[2] (Prot 356b), Plato introduces two further loci of human agency. For those of us who are convinced that people can be genuinely conflicted and not just bad calculators, this is a move with some phenomenological plausibility. However, doubts are in order as to the argument by means of which Plato establishes the soul's partition and as to the status of the resulting three-way split.

The key question for the application of Plato's motivational exclusion test is what makes some behavioural impulse the contrary of another. The example that is supposed to be clear in this respect is that of a person who is thirsty and thus longs to drink the water in a particular receptacle, but who, knowing that the water is salty, judges it best not to satisfy her thirst. The agent is motivated both to perform and to refrain from performing the action. However, it is unclear why Plato thinks that such examples fail to pass his motivational exclusion test. As the properties of being salty and being water, or perhaps thirst-quenching, are distinct, one could maintain that these impulses, whose simultaneous realisations would indeed be incompatible, are nonetheless not impulses "with respect to the same thing". Indeed, if the divided motivation with respect to salt water fails Plato's test, then thirst coupled with the aversion to some specific taste, say of cherryade, will also fail it. But aversion to a particular taste should belong to the same – appetitive – category as thirst. Thus, Plato's principle seems either to prove too much or too little. It appears either to assign the two cherryade impulses to different parts of the soul or to be unable to establish any such difference in the salt water case.

A second reading of the principle could focus on Plato's description of Leontios as "getting indignant at [the] violence [of his desires]" (Rep 440b). Leontios' shame at his desire to view the corpses, the example by means of which Plato separates spirited from appetitive motivation, may lead us to think that what he means by "contrariety" is a relationship between forms of first- and second-order motivation. However, on this reading, the two "contrary" impulses would again have as contents – and thus be "in respect of" – something different. Leontios' "appetite" concerns his viewing the corpses, whereas his spirited reaction concerns his appetitive impulse. It could be replied that the second-order motivation generally gives rise to a first-order impulse whose content is the non-realisation of the content of the appetitive impulse. It might then be stipulated that such first-order impulses generated by relevant second-order attitudes are to count as contraries to the impulse at which the second-order attitude is directed. However, this stipulation would have

[2] I prefer this non-everyday terminology to the frequently used "best judgement", which is ambiguous as to whether the judgement is best as measured against external standards (the best judgement an agent could have made) or involves judging that a course of action would be the best all things considered. It is the latter that "optimising" – as opposed to "optimal" – "value judgement" is intended to pick out.

little to do with the ordinary sense of "contrary". There is, moreover, no good reason to think that all second-order motivation must have its source in what Plato would see as a different part of the soul from that occupied by the impulse it rejects. People are, for instance, sometimes ashamed at being angry under certain circumstances. Finally, the salt water example is not plausibly seen as having this structure.

Irwin suggests that in order to make Plato's test work, we indeed have to accept that his examples are unsuitable to make the point he was trying to make. Moreover, according to Irwin, the relation whose instantiation Plato takes as excluding the relata belonging to a single entity is indeed second-order in character. He sees it as essential to the relation that one of the motives involved entails the belief that the other motive belongs to a particular part of the soul. On the basis of this belief, the first motive essentially involves a rejection of action motivated by that part of the soul in the specific circumstances (Irwin 1995, 216f.). If this were to be correct, what is misleading called "contrariety" would be an attitudinal constellation involving a highly specific form of hierarchical motivation. Above all, it would be a form of motivation that presupposes at least the belief in the soul's partition that the motivational exclusion test is supposed to establish in the first place.

Plato's *argument* for the division of the soul into modular faculties is thus unconvincing. Nevertheless, as I remarked, his distinctions between forms of motivation do have a phenomenological basis. However, in spite of an overlap of his subdivisions with our categories of appetite, emotion and value judgement, there is no one-to-one match here. Independently of the principle of motivational non-contrariety, one wonders whether Plato has clear criteria for which forms of motivation would be classified as belonging to which part of the soul.

The attempts of contemporary interpreters to discover implicit criteria in the architecture of the Platonic soul can be usefully divided into those that do so along *formal* lines and those that localise the criteria in the states' *contents*. The latter attempts (Cross and Woozley 1966, 118ff.; Kenny 1973, 4ff.) raise more questions than they answer. As regards "epithumia": "acquisitiveness" and the "love of profit" hardly fit together nicely with the bodily appetites (580e-581a). The aims of "logos" – knowledge and ruling – require very specific background assumptions in order to be brought together under one head. Perhaps "thumos", a sub-class of what we call emotions and at first sight the most unfamiliar group, is most easily seen as unified: anger, courage and shame may all be seen as manifestations of a striving for something like self-esteem. And certain forms of agitation of small children and even animals might perhaps be seen as primitive forms of such manifestations (Cooper 1984, 14ff.). But even if some such internal unification can plausibly be attained, it certainly doesn't look as though all human motives have been encompassed and there remain large questions concerning the relations between the categories. The drive to "acquire" one thing or another, for instance, seems to play a central role in all three.

More systematic interpretations of Plato's divisions work with the formal criterion of the relation of the particular part of the soul to the concept of a value judgement. According to Irwin's early suggestion (1977, 192–5), the three forms of motivation differ according to the *extent* to which they are dependent on value judgements, the spectrum ranging from complete dependence (reason) over part-dependence (spirit) to complete independence (appetite). Irwin's later proposal (1995, 209–16), by and large seconded by Bonobich (2002, 248–54), sees the parts as differing according to the *purview* of the relevant value judgements, which concern either the agent's life as a whole (reason) or a smaller part of the agent's life (spirit). These proposals have the advantage of squaring with the apparently contradictory tendencies of emotions such as anger and shame both to run away with their bearers and to indicate what is important to them.

However, such conceptions raise the question of the status of the value judgements that play this criterial role. How, one might wonder, can judging what is best be the core of the reasoning part and yet also play a role in the spirited part? The answer lies in the fact that Plato's project is not that of dividing the soul into discrete psychological faculties or functions. Rather, each of the parts is itself kitted out with a set of varying psychological capacities (Bonobich 2002, 220). The name given to each of the parts therefore does not pick out the capacity that is exclusively constitutive of that part, but rather the capacity that dominates it. In Book IX (580d-e), Plato makes this structure explicit relative to "to epithumētikon", which is simply the part in which desires are particularly intense or salient. All three parts of the soul have "epithumiai". It is this structural characteristic of the parts of the soul that allows Plato to draw a relatively stringent analogy between them and the social classes in the state. Plato describes the rulers as characterised by simple and moderate desires that are guided by reason (431c). Here, "epithumiai" is used in a generic sense that simply picks out motivational states (Cooper 1984, 5) – of whatever content and whatever their relationship to value judgements. As has been frequently remarked (de Sousa 1987, 24–7; Irwin 1995, 217–22; Bonobich 2002, 221–3, 254–7), Plato's psychic partitions are agents within agents or homunculi. For this reason, he has to presuppose the concept of a motivational state in order to fit each of the parts out with it. Making sense of that concept is therefore no part of Plato's project. Indeed, Aristotle formulates his question as to what it is in the soul that generates movement as an explicit response to what he saw as Plato's failure to realise its importance.

1.3 Aristotle and the Problems of Motivational Unity

Aristotle takes it to be obvious that his question has to be answered in terms of a generic type of mental state, one not distributed among various faculties. "That which moves", he states, "is a single faculty", a faculty he terms "orexis" (DA 433a22). It is Aristotle's declared aim to establish conceptual unity here, a move in part accomplished by insisting on the separation of "orexis" from purely theoretical capacities such as "dianoia" (thought), which on its own "moves nothing" (NE 1139a36). According to Aristotle, the genus "orexis" is realised in three species (EE 1225b24-26; DA 414b2; MM 1187b36), two of which, "epithumia" and "thumos", are non-homuncular reworkings of the Platonic distinctions, while the third, “boulēsis”, is meant to do the motivational, although not the epistemic work Plato assigned to “logos”. In translation, “boulēsis” is often rendered, somewhat strangely, by “wish” or “rational wish”.

Motivation, then, is according to Aristotle a matter of the functioning of one particular faculty, itself confusingly rendered in English sometimes as “appetite”, sometimes as “desire”. Where one of these terms is used for the generic notion, the other tends to be used for its species, “epithumia”. But *what* is the feature of “orexis” in virtue of which its various species are able to move their bearers to act?

One answer that has been suggested (Tuozzo 1994, 535ff.; Cooper 1999, 243ff.) is that orectic states necessarily involve an *evaluative* component, namely seeing their objects as in some sense “good”.[3] This is uncontroversial for “boulēsis”, which Aristotle characterises explicitly as “desire for the good” (Rhet 1369a3) and sees as directed towards ends grounded in the ethical character of the attitude’s bearer. It is less clear that evaluation is constitutive of the other two forms of “orexis” (Mele 1984b, 147ff.), which Aristotle explicitly terms “irrational” (Rhet 1369a4). His argument for the claim that “thumos”-induced akrasia is less disgraceful than akrasia caused by “epithumia” (NE 1149a24-1149b3) is based on the – Platonic – assertion that “thumos” or anger “listens to” reason. The angry or indignant person has, often over-hastily, taken some action of another person to be a “slight” (cf. Rhet 1378a31-1378b2). Such a “taking” is clearly an evaluation. If this argument about the different statuses of “thumos”-induced and “epithumia”-induced akrasia speaks for an evaluative component of the former, it obviously tells against the same being true of the latter. Nevertheless, there are various passages in which Aristoteles, firstly, ties “epithumia” to the capacity for pleasure and pain (EE 1225b31-32; DA 414b4-6; Rhet 1370a17-18) and, secondly, characterises the pleasant as an “apparent good” (DA 431a8-14). Indeed, this appearance of goodness inherent in the experience of pleasure is supposed to explain “why the pleasant is desired” (EE 1235b26-27).

This is not the place to suggest a solution to what appear to be inconsistencies in Aristotle’s conception. Certainly, when Aristotle declares that “all actions due to ourselves either are or seem to be good *or* pleasant” (Rhet 1369b19-20[4]), it is unclear whether the second “or” is to be read inclusively. This hermeneutic unclarity is paralleled by the systematic question: in what sense can the “evaluation” that is supposed to be at work in the experience of pleasure be the same as that entailed by the “rational wishes” for the ends expressive of a person’s virtue?

Moreover, how precisely is the explanatory role of such evaluation to be understood? If “evaluation” *explains why* something is desired, then it is natural to

[3] The claim that “desire” necessarily aims at the good is frequently labelled the “guise of the good” thesis, a thesis that is in turn sometimes taken to be uncontroversially Aristotelian (e.g. Tenenbaum 2010b, 4). Katja Vogt argues persuasively that the claim in Aristotle concerns “background motivation” to lead a life that goes well, an orientation that may impose no more than side-constraints on many of our individual motivations (Vogt unpublished, 12ff.).

[4] My emphasis.

assume that it is not a *constituent* of what it explains. If that assumption is correct, then even the necessary presence of some evaluative component would still leave the question open as to what unifies the forms of "orexis" as motivational attitudes.[5]

The question of the unity of "to orektikon" is equally fraught with respect to the sub-category of "wish". As Nielsen emphasizes, certain passages in *De Anima* and *Topics* seem to support the claim that Aristotle thought of "boulēsis" as being located in the rational part of the soul, whereas others speak for the view that all forms of orexis belong to the soul's irrational part (Nielsen 2012, 49). The latter view seems to be implied by the claim in the *Politics* that "anger and wishing and desire are implanted in children from their very birth, but reason and understanding are developed as they grow older" (Pol 1134b20-25). The problem appears, at least in part, to derive from Aristotle's use of two principles of classification, one to pick out faculties and sub-faculties, the other to identify "parts" of the soul. He makes one explicit remark on the problem, according to which "that which produces movement, that is, the desiring part qua desiring part, must be one in species . . . , although the things that produce movement are in number many" (DA 433b5-13[6]). Whether this can count as a solution depends on what is involved in being a "species" here. If all that is required is verbal unification – perhaps on the basis of a common causal role – then explanatory unity has been sacrificed. If on the other hand, belonging to a species in this sense involves possessing or instantiating the same mechanisms, the aim of motivational unity is upheld, but it appears completely unclear how Aristotle thinks it might be realised. Only a couple of paragraphs earlier, he had maintained that "breaking up" "to orektikon" would be "absurd" (DA 4324–6). It is difficult to see how he has not done precisely this.[7]

Two further points show that Aristotle had difficulties achieving his aim of clarifying a generic concept of motivation. In both cases, the difficulties he encountered are tied up with his sensitivity to important distinctions within motivationally relevant mental phenomena. The first point is raised by parallel passages in the *Nicomachian* and *Eudemian Ethics* (NE 1105b21-24; EE 1220b11-15). In these, Aristotle lists what he calls the *passions* ("pathē"), among which both "epithumia" and "thumos"[8] are numbered, along with other types of what we call "emotions": "fear, confidence, envy, joy, love, hatred, longing, emulation, pity". This classification raises two questions: firstly, how the two "passions", appetite and spirit, can be of a kind with "wish", which is apparently a non-passionate form of "orexis"; and secondly, what the motivational status is supposed to be of the rest of the emotions named here. Their enumeration alongside appetite and spirit would appear to exclude their being forms of either. However, as they have been

[5] Boyle and Lavin (2010) argue that the guise of the good thesis in Aristotle indeed has the status of a conceptual claim.

[6] Here, I am following the translation used by C.D.C. Reeve (Reeve 2012, 26).

[7] Reeve certainly thinks he has. Cf. Reeve 2012, 26ff.

[8] In the former work, Aristotle entitles the second group of passions "orgē", which Cooper (1999, 251) interprets as synonymous with "thumos".

otherwise assigned no place amongst the forms of “orexis”, one is apparently led to the obviously false conclusion that they are not motivational states. And indeed, when Aristotle analyses the various passions in the *Rhetoric* (1378a21-1388b31) as states essentially involving experiences of pleasure or pain as a result of certain beliefs, only “anger” is assigned an unmistakeably orectic component – although shame does involve “shrinking from” disgrace (1384a24-25). Even fear (1382a21-1382a12) is missing any such feature.

The second point at which Aristotle’s phenomenology of practical mind threatens to undermine his unifying project is his introduction of “prohairesis”, generally translated as “choice” or “decision”. “Prohairesis” is the movement of mind in which practical deliberation (“bouleusis”) culminates and which, barring akrasia, results in corresponding action. As such one would expect it to either be or to give rise to a particular species of “orexis”. And, although Aristotle makes a prominent claim that appears to exclude this – “orexis” has only three species (EE 1225b24-26) – his various definitions of “prohairesis”, as “bouleutikē orexis” (NE 1113a11; 1139a24), “orexis dianoētikē” and “orektikos nous” (NE 1139b4-5), makes it clear that, if it is not a kind of orexis, it certainly issues from the combination of some orectic state with rational or deliberatively produced components.

By means of the systematic introduction of “prohairesis”, Aristotle makes room for the concept of what one can call an *eminently practical attitude*, that is, a kind of motivational state that is in some sense more intimately bound up with action than (other) forms of “orexis”. Kenny has argued that “prohairesis” should be seen as the first recorded attempt to conceptualise a notion of intention (Kenny 1973, 134). This seems right, although two qualifications are in order.

First, although it appears clear that the term stands for a mental event, its extension thus overlapping with that of our term “decision”, it is not so clear whether it is also stands for the product of that movement of mind, that is, what we call an intention. Second and more importantly, neither the mental event nor the attitude is equivalent to our related concept. This is because an Aristotelian “choice” is only made after deliberation that takes as its starting point a “rational wish” directed towards some overarching conception of the good life and which yields a judgement that a course of action is best all things considered (Anscombe 1965, 147; Kenny 1975, 18; Nielsen 2012, 50). “Prohairesis” is the end point of a particular kind of practical deliberation (“bouleusis”), namely such deliberation in as far as it enjoys a specific ethical status (Mele 1984a, 140ff.). For this reason, there is a curious gap in the Aristotelian action-theoretic landscape: neither the practical reflection beginning with an appetite or a “spirited” motivational state, nor the decision with which such reflection terminates is provided with a referring term or a theory.[9] Moreover, from a

[9] Aristotle does remark briefly that the bad or incontinent man will engage in “calculation” (“logismos”), perhaps effectively (EN 1142b17-20). He also names the instrumental faculty of “cleverness” (“deinotes”) that may be possessed by the former as the degenerate version of practical wisdom (“phronesis”), the latter being necessary for genuine deliberation (EN 1144a23-29). There is however, as far as I can see, virtually no elaboration of how reflection in these cases might proceed.

modern point of view, it seems odd that a difference in content or ethical quality of a mental process or state should be taken to qualify the phenomena thus differentiated as being of different *kinds*.

Nevertheless, Aristotle's introduction of the concept of "prohairesis" is an enormous step in the understanding of practical mind. It brings our attention to bear on a form of motivation that is dependent on deliberation on the agent's part and, as a result, conceptually distinct from any form of motivation directly bequeathed him by features of his previous state, whether these be appetites, emotions or other forms of "pathē".

All in all, Aristotle both insists on the necessity of a unified notion of motivational attitudes and at the same time identifies phenomena accounting for which poses a threat to such unification. These concern the relationship between *motivation and emotion*, the Platonic problem of the relationship between *ethical and appetitive motivation* and the question of how to make sense of the *transformation of motivation through deliberation*. Each of these issues is of central importance for any systematic treatment of motivation. One obstacle to their solution in Aristotle's theory is his tendency to run the latter two issues together.

1.4 Hobbes' Double Reductionism

Skipping across the centuries to Thomas Hobbes, we find a powerfully reductive mind intent on establishing unity among motivational attitudes, and on doing so independently of any phenomenological or normative considerations. Hobbes is of particular interest for both positive and negative reasons. On the positive side, he demonstrates admirable intellectual rigour in pursuing Aristotle's unifying project. The negative lesson to be learnt concerns the price to be paid for his particular method of unification.

According to Hobbes' psychological reductionism, all motivational states, including the emotions, are analysable in terms of one basic kind of state, namely "appetite", or its negative variant "aversion" (L VI). All the practical attitudes are constructed by combining this basic motivational building block with features such as beliefs about the probability of obtaining the relevant object, specifications of the object itself, temporal properties of the attitude and further attitudinal compounds. Hope, for instance, is appetite plus the expectation of attainment, self-confidence is "constant hope". The eminently practical attitude that the Scholastics had termed "will" is simply "the last Appetite in Deliberation", deliberation in turn consisting merely in the alternation of states of appetite and aversion on considering possible consequences. The machinery of Hobbesian psychological reduction, once set in motion, simply steamrolls flat all those phenomenal differences that had led

Aristotle into inconsistencies in the context of his unifying project. For Hobbes, reducibility is an a priori matter and he is clearly not going to let himself be distracted from its demonstration by mere matters of detail.[10]

It is striking that Hobbes' definitions of the emotions contain no mention of *affect*. Fear, for instance, is simply aversion plus "an opinion of Hurt from the object". The reason why he apparently sees no need to mention affective components is his assumption that hedonic experience somehow runs parallel to appetite. For Hobbes, pleasure and displeasure are simply what we feel when we desire or are averse; they are the "appearance" of the "motions" constitutive of appetite.[11]

In fact, Hobbes is not particularly interested in what role pleasure plays precisely. In the earlier *Human Nature* (HN VII, 1–2), instead of talking of "appearance", he identifies both appetite and pleasure with the same imperceptible motions of matter. In the later *De Homine* (DH XI, 1), pleasure and appetite are again conceived as in essence identical, the difference now being whether the "object" is present or only foreseen. These differences, which have significant consequences for how specific motivational phenomena are to be understood, are not subject to any investigation that could justify one variant rather than the other. The impression one gains is that the summary formulations are simply there to assign pleasure *some* place in Hobbes' reductive psychological scheme.

The reason for this lies in Hobbes' a priori view of what *must* be going on in human psychology. He thinks that it simply follows from his general materialist ontology, according to which all there *really is* is matter in motion. Matter in motion, then, is what psychological phenomena must be. And Hobbes thinks he can show this by picking out one basic psychological concept, which marks the transition from the physical micro-level of matter in motion to the psychological macro-level of motivation. That concept he names "endeavour", which consists in minimal, imperceptible motions within the relevant body. This provides the basis for his definition of "appetite" as a particular form of endeavour: as "small beginnings of Motion … towards something which causes it". It is in terms of *this* concept of appetite, or "desire" as he sometimes calls it, that all other motivational states are supposed to be reducible.

The construction is not one that anyone would be likely to defend today. Nevertheless, there is something to be gained by pointing to the basic confusions that allowed it to appear plausible in the first place. Firstly, the claim that the *object*

[10] Tom Sorell puts the point slightly more politely. He describes Hobbes as pursuing an empirical, rather than a semantical reconstruction. According to such a project, the precise way our linguistic terms fit together with the mechanisms that cause behaviour is relatively uninteresting (Sorell 1986, 90ff.). Still, Hobbes does seem to think that some terms from our everyday language do pick out the basic motivational concepts.

[11] A contemporary materialist theory of hedonic experience with a comparable structure is Tim Schroeder's conception of pleasure as the representation of net desire satisfaction relative to expectation (Schroeder 2001; 2004, 88ff.). On Schroeder's theory, see Roughley unpublished a.

of a motivating attitude is identical to its *cause*, motivated by the empiricist doctrine that whatever contents are in our heads must have found their way in there from outside, obviously cannot be upheld. Wanting to go to Heaven or to meet Father Christmas are presumably states caused neither by Heaven nor by Father Christmas. And, in spite of the fact that thirst and hunger can be triggered by the perception of some refreshing- or tasty-looking item, both are primarily caused by endogenous changes in their bearers.

Secondly, Hobbes' assertion that appetite involves a form of "motion towards" its object is equally false. Even if matter in motion were to be all that there is, the electrical motions along neurons and the chemical motions of neurotransmitters, which are no doubt characterised by some direction dictated by the lie of the fibres, are not in any literal sense "towards the object" the desire for which is sustained by such motions. In other words, "motion towards" is being used here in two different senses, one *physical*, the other *attitudinal*. The question of the relationship between the two levels is obscured by the ambiguous use of the expression.

Finally, not only does this confusion paste over the gaps between psychological and physiological levels. It also obscures a central issue within the psychological sphere. This is the question of whether *potentially motivating* attitudes necessarily involve "motion towards" their objects in the sense of being *actually motivating*. For Hobbes, having an appetite for something entails *already endeavouring* to get that thing, however imperceptibly. But as soon as one cancels the materialist definition of endeavour, it is obvious that this claim is wildly inaccurate. In the context of Hobbes' system, it appears to have some plausibility, because *every* mental event involves, in one way or another, movement of the matter that supports our mental life. But as this is equally true of perception, the existence of this kind of movement in no way implies that subactional beginnings of overt behaviour have been in any sense inaugurated. Otherwise all mental events would be motivational.

Hobbes thus runs together *three* things under the heading of "motion": the *movement* of material particles, the *directedness* of attitudes towards their object and *motivation* to act.[12]

The resulting construction is nothing if not bold. And it does have the virtue of attempting seriously to get beneath the surface of linguistic structures in order to determine the real mechanisms of human motivation. Nevertheless, in the final resort, it must be seen as a negative example of the price of attempting to solve the problem of *inner-psychological*, motivational unity by reductive *psychophysical* arguments.

[12] Peters and Tajfel argue that the theories of both Hobbes and the behaviourist psychologist C.L. Hull are invalidated by their failure to respect the logical difference between the physical concept of motion and the psychological concept of striving (Peters and Tajfel 1958, 33). On Hull, see below Section 3.3.1.

1.5 Hume and Hedonic Unity

Hume's psychology is in several respects a successor theory to that of Hobbes. Although Hume is untempted by psycho-physical reduction, he is very much concerned to explain the variety of our mental life as proceeding from the functioning of a small number of mechanisms. Like Hobbes, he sees one particular kind of state as the source of our motivation: Hume identifies motivational states with *passions*. His much-cited claim that "reason is ... the slave of the passions, and can never pretend to any other office than to serve and obey them" (T II, iii, 3) echoes a lesser known dictum of Hobbes, according to which "the Thoughts are to the Desires, as Scouts, and Spies, to range abroad, and find the way to the things Desired" (L VIII). Both are more or less explicit agreements with the Aristotelian contention that "the calculative faculty or what is called thought [cannot] be the cause of ... movement" (DA 432b26-27).

Hobbes replaces Aristotle's three- (or four-)fold orectic structure with his unidimensional, but idiosyncratic concept of "endeavour"; where Hobbes sidelines affect, Hume claims that it is the members of the Aristotelian class of "pathē" that are responsible for "moving" us. Like the Aristotelian "passions" (NE 1105b23-24; Rhet 1378a21-22), the Humean variants also constitutively involve pleasure or pain. For Hume, however, in contrast to Aristotle, it is precisely this affective component that is supposed to explain action.

Again like Aristotle, Hume offers componential analyses of the passions. Viewed as attempts to answer Aristotle's question in *de Anima*, Hume's analyses, like those of Hobbes, aim to reveal the compound structures within which the motivationally relevant psychological factor is at work. A significant difference between the motivational reductions of the two authors is that Hume's general motivational term "passion", unlike Hobbes' "appetite" or "desire", appears not to designate the *basic motivational feature* common to all motivational states. Rather, it at least usually stands for the particular *compound* that contains the relevant component.

Hume seems clear enough on what it is that moves us, although, as I shall argue in a moment, when one really wants to know how the conception works, it begins to wobble. The following appears to be a canonical expression of his position: "There is implanted in the human mind a perception of pain and pleasure as the chief spring and moving principle of all its actions" (T I, iii, 10). The passions are mental states (what Hume means by "perceptions") which, so Hume seems to say, move us because they involve us being *affected hedonically*.

Not every pleasure or pain is a passion. In particular, mere sensations of either kind don't qualify. Passions are one step removed from immediate sensation and thus belong to what Hume calls "secondary impressions" or "impressions of reflexion" (T I, i, 1; II, i, 1). He thinks of these as mental states in some way causally dependent on sensations, a dependence usually mediated by the intervention of some "idea". For Hume, the difference between "ideas" and "impressions" is gradual, being a matter of the "liveliness" or "force" with which they are experienced. Attaining the relevant level of "force" – *experiential vivacity* – entails crossing the dividing line between ideas and impressions. Secondary impressions have the requisite degrees of "force" whilst also, generally through the mediation of an idea, involving the representation of some object or state of affairs.

It is this structure that allows Hume to develop his componential analyses of the various types of motivation. The experience of pleasure in some context gives rise to an "idea" or thought of the cause, or perhaps some accompanying feature, of that pleasure, which in turn becomes coloured pleasurably. This simplest case is what Hume calls "desire" (T II, iii, 9). Enjoying the experience of drinking a cup of tea might, for instance, generate the pleasurable thought of, or desire for, a cup of tea. The admixture of degrees of subjective probability concerning the thought's object gives us the further "direct passions" of joy and hope. More complex motivational states, the "indirect passions" such as pride, involve several representations, which need to stand in a particular relationship to one another. But what enables them to move us is the same feature, namely a *hedonic colouring of some representation*.

As in the cases of joy or grief, hope or fear, the hedonically coloured representation can be accompanied by a *belief* as to the probability of its realisation. Indeed, thoughts of certain states of affairs may sometimes only be pleasurable if accompanied by some level of subjective probability. On the other hand, there is no necessity that Humean passions contain any reference to *future* pleasure. All that Hume requires is present hedonic colouring. Were this not to be the case, one could make no sense of Hume's famous pronouncement that it need not be irrational to prefer my own total ruin to the slightest unease of someone completely unknown to me (T II, iii, 3). Were expected pleasure to be what motivates us, this preference would be massively irrational. Hume's point here can only go through if he identifies a passion for something not with how one believes one will feel when it is realised, but with how one feels about it at the moment of passion.

One not inconsiderable problem with this interpretation of Hume's position is that it appears to contradict a somewhat mysterious passage from the section "Of the influencing motives of the will", according to which a passion "is an original existence ... and contains not any representative quality" (T II, iii, 3). In the course of her defence of Hume, a clearly exasperated Annette Baier (1991, 160–164) refers to this passage as a "very silly" and "unfortunate" paragraph. A little unfortunate it no doubt is. Nevertheless, I think it can be understood in a way that supports the interpretation I am canvassing. It does so if one understands the word "passion" here as, exceptionally, referring not to the *attitudinal compound* of which the affect is one component, but just to *the affect itself*.

Certainly, the affective component is not itself representational, but is, to use Hume's own expression, a "modification" of some such representation – what later became known as its "hedonic tone" (Broad 1985, 38). It is quite surely unthinkable that Hume, after spending a third of the *Treatise* detailing the representational qualities of the passions, should have suddenly forgotten everything he has been saying at precisely that moment at which he is explaining how they move us to act.

What has "no reference to any other object" is, then, the component in virtue of which a representation becomes a component in a passion.[13]

This reading would perfectly well provide Hume with the justification he obviously thinks he has for the claim that "passions" – taken either as the affective components or as the entire attitudinal compounds – cannot in any strict sense be "untrue". The point Hume is trying to make is one that since Frege is easily articulated (cf. Sect. 2.4.2). Just as linguistic representation need not involve assertion, so attitudinal "representation" need not involve belief. For it to do so, a "modal" component has to be involved. But there is also an everyday sense of the term "represent" which *does* entail the claim to veridicality.[14] That this is the sense in which Hume is using the term comes out in his argument that the contradiction to which passions are insusceptible "consists in the disagreement of ideas . . . with objects, which they represent". The passions do represent states of affairs; they simply don't represent them *as true*. Using post-Fregean terminology, one could say that for Hume *affect is the mode* of representation constitutive of the passions.

Hume's position is undoubtedly a candidate for the answer to Aristotle's question. Nevertheless, its disadvantages are not too difficult to see. I shall briefly name two. Part of the reason for the disunity in Aristotle's, and indeed Plato's theory of motivation derives from the recognition of the second.

The first point is the doubt as to whether such hedonic colouring of some representation is sufficient to move us towards what is thus represented. In fact, one might think that, if the mere thought of something triggers a pleasant sensation, this could easily lead to motivational inertia rather than to activity: if I feel good now, why move? It is this point that lends a certain plausibility to Locke's negative hedonic conception of desire as "an uneasiness of the Mind" (E II, xxi, §31).

Perhaps it was this that led Hume to put his position in slightly, yet decisively different terms where he is trying to explain the connection between "being affected" and "being moved": "Tis from the *prospect* of pain or pleasure", he now says, "that the aversion or propensity arises towards any object" (T II, iii, 3[15]).

[13] John Bricke, who, like Kenny (1963, 25, note), thinks Hume's official theory commits him to "denying the intentionality of desire", suggests that a proposal along the above lines would be a "modest revision" of Hume's official theory (Bricke 1996, 39ff.). Setiya argues that Hume was in fact articulating the very distinction between representing a content in a doxastic and in a "passionate" mode that Kenny and Bricke see him as rendering incomprehensible. Setiya's claim rests on the philological detail that in the eighteenth century, unlike today, commas could introduce restrictive relative clauses. According to Setiya, Hume's claim is incomplete until one adds that a passion "contains not any representative quality, which renders it a copy of any other existence or modification". What passions don't contain are thus particular sorts of representations, namely those that copy features of the world, not representations tout court (Setiya 2004, 374f.). This would indeed be a way of rendering this passage consistent with the main thrust of Hume's discussion of the passions. However, it seems to be less consistent with the way the quotation continues, as Hume goes on to assert that being angry – like being thirsty, sick or more than five feet high – involves "no reference" to any other object.

[14] There is also such a widespread philosophical usage. See Chapter 2, note 11.

[15] My emphasis.

Put this way, there appears to be an explanatory connection between affect and motivation that grounds in their bearer's rationality. However, the gain in rationality has been purchased at the price of a shift from a theory of present affect to a theory of hedonic expectation. It would make passions into *beliefs* and thus dissolve the distinction between motivational and non-motivational states Hume has been attempting to ground.

The second problem with Hume's theory of motivation is that viewing *prior qualitative experience* as the cause of action generates massive difficulties for the explanation of what we do where *no such experiential data* are given. Hume does offer a response to this fairly obvious problem. However, his response, rather than rescuing the theory, looks more like an admission of its incapacity to provide the general answer it was supposed to be providing.

Hume somewhat notoriously postulates "calm passions", which are "more known by their effects than by the immediate feeling or sensation" (II, iii, 3). This hardly squares well with the claim that "different degrees of force" are the "source of all the differences in the effects" of different kinds of mental state (I, iii, 10). At least this is so as long as the only concept of "force" Hume is working with is that of hedonic intensity or "vivacity". As ideas only become impressions through the attainment of a certain level of hedonic intensity, the notion of "calm passions" is an oxymoron.

Note, finally, that it is not only action according to certain kinds of moral judgement that falls outside the purview of causation by generic secondary impressions. Perhaps more important are the large amounts of everyday, instrumental and institutional actions for whose performance we form intentions without experiencing any kind of felt desire. Like Aristotelian "prohairesis" (EE 1225b24-26), the motivating states we call "intentions" require no feelings.

1.6 From Stevenson to Davidson: "Pro-Attitudes"

A last important move in the analysis of potentially motivating attitudes is bequeathed to us by the meta-ethics that came out of Logical Empiricism. The context is the attempt to make sense of value sentences against the background of the claim that such sentences are not verifiable and the assumption that non-verifiability threatens complete meaninglessness. The solution put forward by Charles L. Stevenson grounds in a psychological semantics, according to which meanings are mental states expressed by sentences. Of interest for our purposes is the claim that the mental states thus expressed divide into two semantically relevant types, those which can be true and those which can be satisfied. The first group Stevenson labels, traditionally, "beliefs", the second group he calls *attitudes*. Picking up R. B. Perry's definition of *interest*, Stevenson defines attitudes as ways of "being for or against something" (Stevenson 1948, 1f.; Perry 1967, 115).

Although both Stevenson and Perry are concerned not with the explanation of action, but with the foundation of a theory of value, their related conceptual moves are in a certain sense descended from the strategies of inner-psychological reduction pursued by Hobbes and Hume. Both "attitude" and "interest" are generic terms under which such divergent psychological phenomena as "liking", "desire", "will" and "seeking" (Perry), "love" and "approval" (Stevenson) can be grouped. However, the logical as opposed to explanatory aims of these authors allow them to remain agnostic as to the whether the states thus grouped together are united by some common mental trait realised – in whatever way – in the bearers of "attitudes" or "interests". It is conceivable that the "bias of the subject" thus named is describable in behavioural terms and that the states of the subjects which cause the relevant behavioural syndrome are themselves irreducibly plural. Indeed, Perry is explicit that "interest" does not simply label one kind of ontological category, but can refer to a "state, act, attitude or disposition", whereas for Stevenson all "attitudes" are dispositions.

The lack of clarity surrounding the status of the mental states classified as "attitudes" can be seen from the fact that Stevenson's meta-ethical theory ran under the infelicitous label emotivism (Carnap 1963, 1000). But there are certainly states of persons other than those which fall under the everyday concept of emotion that can involve them *being for* a proposition. The labels which have since become more usual for value theories of this kind, non-cognitivism and expressivism, reinforce the point. The first simply tells us what kind of states are *not* at issue. The second avoids the question as to what kinds of state we are dealing with altogether, focussing instead on the type of relation the theory claims holds between mental states – of whatever kind (except beliefs) – and evaluative language use.

What Stevenson (1948, 2) appears to offer as a criterion for "attitudes" – the susceptibility to "satisfaction" rather than truth – is in fact not intended to bear any systematic weight. He offers no analysis of the relevant concept of satisfaction. Indeed, in *Ethics and Language* (1944, 67), he argues that *both* "attitudes" and "beliefs" are dispositions to act and that the terms he is using to describe them "have only such clarity as is afforded by instances of their usage". In other words, like Plato arguing for the tripartite soul, Stevenson sees himself forced simply to rely on the recognisability of paradigmatic cases.

Stevenson's use of the word "attitude" is modelled on everyday uses, as in "a dismissive attitude" or "a positive attitude", although such uses often refer to a stable compound of different mental states with a variety of related objects over a period of time. Since Stevenson, "attitude" has become a central term in the philosophy of mind, generally standing for all mental states with contents. Where it is thought that the contents are necessarily propositions, talk of "propositional attitudes" has become commonplace. In such conceptions, not only liking and hoping, but also remembering and expecting are attitudes. Where this is the case, not only do Stevenson's criteria need sharpening; his terminology also needs revamping.

An influential terminological step was taken by Patrick Nowell-Smith in his *Ethics* (1957a), a step that has left traces in the present debates. Nowell-Smith picks out the non-epistemic – and non-imaginative – attitudes by the simple addition of a prefix to produce the notion of "pro-attitude". This terminological move raises the question explicitly as to what makes an attitude one in virtue of which someone is *for* the attitude's content. In other words, what precisely is the *pro*-component? Again, is this just a linguistic device by means of which a plural set of mental states can be collected or is there some unifying mental component picked out by the prefix?

Nowell-Smith specifies the meaning of "pro" by two means. Firstly, he provides a *list* of clear cases for both "pro-" and "con-attitudes", the former including "like", "want", "desire" and "pleasure". He argues, similarly to Stevenson, that we would have no difficulty knowing whether new cases should be assigned to either of these groups and, if so, to which. Such a remark is of course no more than a prima facie indication that there is some kind of unity. The *criterion* he then offers is that pro-attitudes are those states of an agent the mention of which provides a "logically complete explanation" of why she "chooses" to do something. He apparently sees this formulation as equivalent to the claim that mentioning the state provides a "logically impeccable reason" for their choice (1957a, 100).

From an action-theoretic perspective, it is striking that Nowell-Smith is not claiming that pro-attitudes explain action. Rather, he claims they provide explanations and reasons for "choosing" or "deciding". The "logical" relationship at work here is, on the one hand, supposed not to be analytic (1957a, 95): there is no implication from having a pro-attitude to choosing to do something – although there presumably is in the other direction. On the other hand, the connections between them entail that, under normal circumstances, a sentence of the form "X φ-d because of his PRO(φ)" makes further questions as to why X φ-d absurd or "logically odd". It is true that we can construct situations in which "because I enjoy it" would be an inadequate answer to the question "Why are you listening to that music?" But in the absence of such special constructions, the retort "I know that, but I still don't understand why you're listening to it" would be peculiar linguistic behaviour.

Leaving aside the justificatory issues which were Nowell-Smith's main target, his explanatory claim has at least a strong prima facie plausibility. For our purposes, however, his argument does little more than to sharpen the question: what is it that allows us to draw up such a list of mental states that all take on what he calls the "quasi-logical" role described above? Is the same factor responsible in each case?

The question appears particularly acute in the light of the fact that experiencing displeasure and developing the desire to be rid of it are often separate mental stages on the way to action, the latter being explained and justified by the former. But both "pleasure" (and "discomfort") and "desire" (and "aversion") appear on the lists of pro- and con-attitudes. Perhaps pro-attitudes can also both explain and justify each other, not just action choices. Moreover, one would like to know whether "choice" doesn't also belong on the pro-list. If "pleasure" can explain "desire", which in turn can explain "choice", then it seems natural to assume that we have three in some way significantly different types of state. Alternatively, perhaps we should not be taking "explain" literally and all three are only different manifestations of the same type of property. Or again, perhaps "choice" is itself to be thought of as a kind of action rather than as an attitude.

Nowell-Smith leaves us up in the air about these issues. His topic, the logic of moral language use, appears not to require, and his conception of philosophy appears unable to furnish answers. Nevertheless, it is striking that he goes out of his way to praise Hobbes' construction of a "single model" of action causation, seeing the distinction between pro- and con-attitudes as a descendant of the latter's "distinction between 'endeavour toward' and 'endeavour fromward'" (1957a, 98). Nowell-Smith is, however, also critical of Hobbes. He argues that the orientation to the paradigm of the appetites inaccurately assimilates all the pro-attitudes to one pattern of explanation, according to which motivation is necessarily motivation to rid oneself of prior unpleasant sensations: "the itch-scratch pattern". Were Hobbes to have suggested such an assimilation, this would be a valid criticism. But as I have argued, this is not the case. Hobbes is largely uninterested in the causal role of hedonic experience. Locke's identification of "desire" with "uneasiness" would have been a more appropriate target (cf. below Sect. 5.3.3). Actually, what Nowell-Smith seems primarily to be objecting to – and not unreasonably – is the fact that Hobbes' generic use of the terms "desire" and "appetite" can easily lead to the assimilation of the generic term to particular kinds of case.[16]

The term "pro-attitude" has since been given prominence within action theory through its use in Davidson's (1963) anti-Wittgensteinian reinstatement of the causal conception. Like Nowell-Smith, Davidson provides us with a *list* of what he sees as clear cases. Two differences from Nowell-Smith are worth remarking on. Firstly, Davidson's list contains no affective concepts, omissions that he does not feel need justifying. However, some justification would be in order. Hume was surely right to see affect as involving a pro- or con-component.

Alongside these omissions, Davidson also makes a number of additions, in particular of "moral views, aesthetic principles, economic prejudices, social conventions ..." (1963, 4). As principles and conventions are not necessarily, or at least not obviously, features of a person's mind, he adds "in as far as these can be interpreted as attitudes of an agent towards actions of a certain kind". Again, he offers no explanation of *why* these phenomena can be lumped together with the "desires", "urges" and "yens" he also mentions. What is clear is that the phrase "attitudes of an agent directed towards actions of a certain kind" can hardly count as informative.

The mental states that Davidson brought together under Nowell-Smith's term "pro-attitude" are in contemporary debates often dealt with under the very term that Nowell-Smith objected to because of its hedonic connotations, namely *desire*. The items that Bernard Williams (1980, 105), for instance, pulls together under the term are close to those on Davidson's list. And Harry Frankfurt (1971, 12f.), weighing up whether to use "want" or "desire" as his most general term, chooses the latter for reasons of style, because he feels that speaking in the singular of a person's "want" would be a linguistic "abomination"(!).

[16] Recall that Plato's use of "epithumia" faced a similar problem.

But perhaps the break with everyday understanding that this unusual use might produce is, for pragmatic rather than stylistic reasons, to be welcomed rather than rejected. Certainly, Nowell-Smith was right that, when choosing a term here, one has to be careful to avoid contingent connotations that its everyday use may unhelpfully transport. The systematic difficulty consists in clarifying just what is contingent and what belongs to the core of the concept. Only if this can be done will we know whether there is indeed some generic pro-component in our attitudinising that makes certain mental states apt to move their bearers to act.

Chapter 2
Motivational States

In what follows, I develop the first outline of an answer to Aristotle's question as to what it is in the soul that originates movement. Clearly, not all our "movements" result from the workings of our psychology. Rather, the relevant forms of behaviour are specifically those that, as Aristotle puts it, are "up to us" ("to eph'hêmin") (NE 1110a17-18; 1113b22), in other words, those that we intuitively think of as things that we do. It is these that figure as explananda in both empirical motivational psychology and what I am calling a philosophy of practical mind. The latter discipline also naturally asks whether there are such cases to be explained where we are dealing with certain non-human animals. It therefore also has some overlap with the empirical disciplines of comparative psychology and cognitive ethology.

In approaching Aristotle's question, I shall both draw on empirical data provided by the psychological disciplines mentioned and look at the way motivational psychologists circumscribe their field. I will be combining these glances across disciplinary fences with the revisiting of moves made by authors in the analytic tradition that have drifted out of the focus of current philosophy of mind and action. In particular, I will be taking leaves out of philosophical books that are usually not thought to be particularly compatible. The trajectory will move from a conception of behaviour, understood in a modified Dretskian fashion, to motivated behaviour, understood as dependent on representational states, where the understanding of the structure of those states owes much to work done in the 1950s and 1960s by Richard Hare and Anthony Kenny. At the end of the chapter, I arrive at the outline of a tripartite componential model of motivational states, before finally looking at empirical evidence for its instantiation in non-human animals. In partial agreement with Aristotle, I argue that some animals are indeed moved by motivational states. To see this, we need to see an important asymmetry between the "modal" components of motivational and doxastic attitudes. It is the modal feature of the former which will, in the course of the following chapters, turn out to be the key to understanding "wants", even in cases in which they don't motivate their bearers.

N. Roughley, *Wanting and Intending*, Philosophical Studies Series 123,
DOI 10.1007/978-94-017-7387-4_2

2.1 Starting Point: The Things We Do

Aristotle's question as to what it is in the (non-vegetative) soul that originates *movement* is posed on the basis of a presupposition that may not appear obvious: that it makes good scientific sense to isolate those movements of our bodies that are caused by properties of our "souls", distinguishing them from movements that have no such aetiology. A noteworthy rejection of this presupposition is provided by the motivational psychologist P.T. Young in his circumscription of the topic of his discipline. According to Young (1955, 193), motivation is best defined as "the determinants of behaviour". Consistent with this definition, he sees it as perfectly acceptable to ask of a billiard ball "What motivates it in its course across the table?" (1961, 18).

Young apparently speaks a different dialect of English to that of anyone I know. For most speakers of the language, "motivation" is, firstly, only applicable to certain kinds of subject. Billiard balls are normally not seen as being among them. Secondly, only certain "movements" of those subjects are generally understood to be "motivated". Talk of "motivation", then, usually marks a distinction within the causes of the behaviour of humans and possibly of some other animals. Were the term to mark no such distinction, it is unclear how the discipline of motivational psychology would be able to distinguish its topic from the questions as to the determinants of human behaviour posed by physiologists and macro-sociologists.

The boundary that is generally marked by distinguishing "motivated" from other forms of "movement" is closely related to that marked by Aristotle's distinction (Rhet 1368b33-34) between those forms of behaviour that are *due to* behavers and those which are not. The latter result, he says, from either chance or necessity, "necessity" being in turn divisible into necessitation by nature and compulsion. We do not move from *A* to *B* because we are motivated to do so if our movement is a result of us being carried there "by a wind or by men" (NE 1101a3). Similarly, if – to take an example that is not Aristotle's – a surgically installed quantum indeterministic device directly produces movements of a person's limbs, we would not say that she was "motivated" to produce those movements.

The latter example is appropriate to Aristotle's concerns, as the criterion for what he calls "voluntary" ("hekousion") movements, namely that their "origin" or "principle" be *in* the agent (NE 1111a23-24), does not, even in the case of animals, name sufficient conditions if "in the agent" is understood simply in terms of physical location within the agent's body. Moreover, the point of the distinction also cannot be to exclude the causal relevance of states and events outside the person's body. Otherwise, the properties of objects that make them "attractive" or "repellent" would render "approach and avoidance behaviour" as "necessitated" as bodily movements brought about by the wind (NE 1110b9-16). But, of course, the movements thus caused are prime examples of movements brought about by our "motivation".

The distinction Aristotle is concerned to reconstruct[1] is completely familiar to everyday agents today, although it starts to lose its contours once one attempts to give it criterial precision. This is as it should be if the question of practical mind is to warrant a sustained philosophical investigation. The familiar distinction is that between what we *do* – in some fairly strong sense – and all other events which involve us in one way or another. Put slightly differently, what we are after is a basic notion of *activity* or *active involvement in occurrences*. The relevant notion is basic in the sense that it has nothing to do either with special efforts to achieve something or with freedom of action, both of which are further specific properties of the basic form of doing we are after.

Like any other branch of philosophy attempting to make sense of human self-understanding, a philosophy of practical mind sensibly begins, as Aristotle recommended, with the relevant phenomena. These are in part the phenomena with which we are confronted in our *everyday lives*, in part the phenomena sedimented in our everyday *language* and in part the phenomena investigated under controlled experimental conditions by modern *psychology*. None of these sources provides infallible criteria for correct analyses, being themselves dependent on the subjective perspective of everyday agents, on contingencies in the development of natural languages and on the conceptual and methodological take of the experimenter. Nevertheless, they provide the necessary starting point for any investigation that is to avoid losing track of what it is investigating.

They also mark points of reference to which a philosophy of practical mind should be able to relate its results. Where, for instance, a large gap opens up between the claims of such a theory and the structure of everyday language, the theory ought to be able to give a plausible explanation for that gap. And that explanation ought to consist in more than the global claim that everyday language is simply confused. The same goes for the results of empirical psychological experiments. If a series of such results appear incompatible with a philosophical theory, then again a special explanation is called for. In the course of this investigation, I shall repeatedly draw on data from each of these three sources.

2.2 Behaviour

2.2.1 Two Repudiations

Frequently, when the topic of motivational psychology is circumscribed, it is done so by means of the term "behaviour" (Young 1961, 17; Wagner 1999, 2; Gollwitzer

[1] Everyday talk of whether bringing about some state of affairs is really "up to us" might be given a stronger interpretation, requiring perhaps indeterministic free will or at least the exercise of decision. This is not what Aristotle understands by the phrase. In spite of the efforts of Alexander of Aphrodisias to establish an identity between behaviour that is "to eph'hêmin" and actions deriving from "prohairesis", this is, as Susan Sauvé Meyer shows, clearly not Aristotle's position (Sauvé Meyer 1998, 227ff.).

et al. 2000, 191). This has the rhetorical advantage today of being a term used across the board in the sciences. As such, its use may appear to lay a solid foundation for conceptual clarity. However, precisely because of its recent terminological history, "behaviour" tends to be understood in a way that distorts the issue of practical mind. Nevertheless, if one explicitly rejects certain implications that the term is generally thought to have, then it can be usefully employed to circumscribe those events involving agents that are *candidates* for motivational occurrences.

In fact, the very reason why the use of the term "behaviour" appears to confer scientific respectability on the study of motivation is what renders it misleading if it is not clearly qualified. One can pick out the objects of study of the various sciences by means of the term: we can study the behaviour of electrons, of chemical compounds, of plants, of rats and of box-girder bridges. What are studied in all these cases are patterns of observable events in which the object of study is involved under specific conditions. Movements observable from a third-person perspective are seen across the board, and certainly with good justification, as the objects of scientific study.

The problem is that our motivated doings are an exception. They are not accurately seen as a sub-class of third-person-observable movements. There are two reasons for this. The first is that we can *do nothing* because we are motivated to do so. There are two sorts of motivated non-doings. Someone can have a good reason to stand completely still – if they are trying to avoid being noticed, for instance. Such cases are, up to a point, third-person observable in the way that the immobility of electrons caused by attractions on both sides may be observed. Moreover, they may well involve muscular movements carried out in order to suppress other movements the body might otherwise tend to bring forth. What causes trouble in principle for the subsumption of motivated doings under behaviour, understood as a sub-form of movement, are cases of omission or forebearance (Bentham PML VII, vii). In many cases, no amount of third-person observation is going to reveal that someone has refrained from doing something. A certain event not taking place at *t* is frequently[2] only an omission if some subject of a possible action is or has been the bearer of certain attitudinal events prior to or at *t*.

This brings us to the second reason why our motivated deeds are no sub-class of third-person-observable movements. Either actions or omissions may result from unobservable activity, such as consideration of the pros and cons of acting or forebearing. Deliberating on whether to do or refrain from doing something is obviously something an agent does actively. So also are mental arithmetic and voluntary fantasising. These are all things generally done because the person was motivated to do them, although they are internal (Bentham PML VII, xi), in the sense of being non-third person observable.

[2] Some non-movements are omissions in virtue of norms or expectations that specify relevant performances, independently of whether the agents in question have had thoughts concerning those performances. On the significance of this for intention, see Section 10.5.2.

There is a simple confusion here that may contribute to obscuring these problems. Where we bring about some bodily movement because we are motivated to do so, the "motivation" is not the movement itself, but what causes it. In such cases, we are, someone might say, "moved to move". But obviously, in such an expression, the verb "move" has two different meanings.[3] Thus, there is no conceptual reason why we should not be "moved" not to move, or be motivated to behave in a non-third-person observable manner.

If what is "motivated" is to be described as a sub-class of forms of human "behaviour", then "behaviour" must include non-third-person-observables of two kinds, namely omissions and mental occurrences. Although, in both cases there will no doubt be some kind of neurological or other movement *in* the agent, these do not amount to movements *of* the agent.[4]

If the reference to "movement" is dropped and it is clear that these two kinds of case are to be included, then we can continue talking of "behaviour". In doing so, however, we should be aware that the terminological continuity with the observational methods of the natural sciences (Watson 1913, 170) has been broken. This doesn't take the investigation outside the realm of science. It does mean, however, that adapting scientific methods to the phenomena involves accepting data that are not strictly subject to intersubjective verification. And it clearly involves abandoning the disjunction that Watson, in his programmatic statement of Behaviourism (1913, 163, 176), introduced between "behaviour" and "consciousness". If, then, we continue to say that what is motivated is necessarily "behaviour" – and there is certainly linguistic convenience in doing so – then we need to insist that doing so entails a repudiation of assumptions connected with the term that are prevalent even among non-behaviourists.

2.2.2 Behaving and Undergoing

The inclusion of both mental occurrences and forbearances widens our purview sufficiently to take in all the phenomena that are candidates for motivated behaviour. However, we also need some criterial restrictions on the kind of events involving humans or other animals in order not to end up with a hopelessly wide conception of behaviour. Where an animate body is moved by external force, thus excluding the movement from being "due to" the creature (cf. Sect. 2.1), we would not

[3] Hobbes explicitly denies this (cf. above Sect. 1.4). For similar reasons, Bentham (PML X, iii, note) suggests that in cases of forebearance, we should talk not of "motives", but of "determinatives".

[4] Although this is acknowledged by Dretske (1988, 29), he apparently thinks that the basic idea of defining "behaviour" as a kind of movement can be upheld. I fail to see how. A similar acknowledgement, followed by a similar dismissal of the problem is to be found in Davidson's action theory. Davidson claims (1971, 49) that, if the idea of bodily movement is interpreted "generously", it will include "mental acts", among which he numbers deciding. There is such a thing as misplaced generosity, of which this is surely an example.

consider it to have "behaved" in the relevant sense. Similarly, neither being crushed to smithereens nor becoming infected by disease are in themselves forms of behaviour.

"Behaviour" can be thought of as an intermediate category which, although it requires no "activity" on the part of its "subjects", nevertheless excludes the form of their involvement in an occurrence from qualifying them as the mere "objects" of that occurrence. When an entity is "behaving", it is, in a weak sense, the "subject" or "bearer" of the occurrence or process in question (cf. Seebass 1993, 10). Behavings and undergoings are thus mutually exclusive.[5] Behavings are forms of involvement of animals or humans in occurrences that exclude pure passivity. The further restrictions that will yield motivated behaviour should then qualify that involvement as active, that is, they should yield forms of behaviour that are "due to" their bearers.

Someone is behaving who is eating or climbing the academic hierarchy, but also if they are sweating, sneezing, stammering or sleeping. Similarly for non-third-person-observable events in which we are involved. We are not only behaving mentally when we deliberately picture scenery or try to solve philosophical problems in our heads; becoming angry, worrying, daydreaming and forgetting are also all forms of mental behaviour.[6]

The natural use of the terms "subject" and "object" in characterising behavings and undergoings may suggest that the distinctions we are after can be read off from the grammatical structures of natural languages. Someone might think that we are dealing with "behaviour" whenever the relevant events are denoted by verbs in the active voice, where the person or animal of which they are predicated is represented by an expression in the subject position. However, it is easily shown that this doesn't work (cf. Thalberg 1972, 49f.; Davidson 1971, 44). There are plenty of verbs in the active voice which describe a person's involvement in events of which she is the "object" rather than the "subject". Examples are "suffer", "undergo", "bear" and "contract" (a disease). We can also "acquire" properties as a result of "Cambridge changes", as when someone becomes an uncle perhaps without even aware of it. Even more obviously, we use other verbs such as "be", "have" and "possess" in the active voice to simply ascribe non-behavioural properties.

Note further that reference to the pro-verb "to do" is of no help here either.[7] On the one hand, as I remarked in 2.1, our intuitive understanding of "doing" is pretty much that of active behaviour. This is why, if someone replies to the question "What have you been doing over the past few years?" with "Losing hair", the reply is

[5] This is not to say that someone suffering an undergoing may not *also* be behaving, as when a victim of a robbery struggles against the robber (Thalberg 1972, 49). The point is simply that, in as far as some event involving a person is an undergoing of that person, it cannot also be a form of his behaviour.

[6] Here I agree with Galen Strawson (1994, 308–312).

[7] Equivalence with "doing" is suggested by G. Strawson (1994, 292) as one possible criterion for "behaviour".

likely to be meant as a joke. On the other hand, such a reply exploits a grammatical possibility provided by the English language. This is that "pro-predications" with "do" can unproblematically be used to stand for either doings or undergoings. The latter is illustrated by a sentence such as "Far more people contracted malaria than would have done under better sanitary conditions".

2.2.3 A Causal Criterion

The central criterion for behaviour can, I think, be usefully taken over from a proposal by Fred Dretske, in spite of Dretske's neo-behaviourist orientation. According to Dretske (1988, 1ff.), behaviour is a form of movement the primary cause of which is internal to the subject of that movement. Now that the behaviouristic restriction to third-person observable movements has been rejected, Dretske's causal criterion, when applied to events involving humans or other animals,[8] provides a plausible way of establishing a dividing line between motivational candidates and mere undergoings.

Drawing the dividing line in this way subsumes under "behaviour" not only those events we think of as our active deeds, but also bodily processes, such as haemorrhaging and trembling, bodily reactions such as blushing, and breakdowns in attempts to attain some goal, as when someone fumbles, stumbles or miscues (cf. Thalberg 1972, 55ff.). It also includes forms of mental behaviour which we do not bring about because of our motivation to do so, above all, our emotions.

The bodily movements of the man with a surgically implanted quantum indeterministic device would also count as forms of his behaviour, as would any thoughts of his that it also brought forth. In contrast, the external causes of bodily movements that for Aristotle exclude these from being "due to" us are, in as far as they bring about genuine undergoings – and not actions under duress (kinds of Aristotelian "mixed" cases (NE 1110a11)) – also excluded from the (reformed) category of behaviour.

Dretske's criterion provides us with a useful way of drawing a distinction on the way to understanding what it is to do something because one is motivated to do it. Moreover, it also marks a dividing line that is undoubtedly of importance for our self-understanding. We are not only concerned with our moments of active involvement in the ways of the world. Our instantiations of dynamic properties are of considerable *significance for the sort of people we are* whenever what causes those instantiations are features internal to us. The distinguishing features of fearful and empathic as well as asthmatic and clumsy people are all behavioural properties.

What is of particular importance for our purposes is that this aetiological definition of non-undergoings adumbrates the structure that can be filled out by the

[8] Dretske applies his criterion not only to humans and other animals, but also to plants and artefacts. I shall not consider cases of the latter kind.

theory of motivated behaviour. The way the latter, as it were, slots into the former reconstructs neatly, and I think plausibly, the sense of a broad set of occurrences and processes containing the particularly significant sub-set of motivated behaviour.

Talk of "internal causes" – of whatever kind and even where their effects are third-person observable movements – is itself a departure from classical Behaviourism, certainly as conceived by Skinner.[9] Classical psychological Behaviourism viewed the organism as a black box responding to external stimuli. The black box must presumably have been constructed in the right way, but once this is the case, the causes of the organism's behaviour are quite simply the inputs into the box, that is, features of its immediate environment.

Talk of "primary internal causes" picks out features inside "the box" whose causal function must in some sense be of greater significance for the relevant event than events outside the entity. What ought to be clear, however, is that the elevation of any event to the status of "primary" cause is going to be a relative matter. Whatever dispositional structures are installed in some entity, if no triggering events of the relevant kind take place, the causal conditions for the entity's reaction will not be completed. Nevertheless, we do appear to be justified in isolating cases in which internal dispositional structures are particularly important. The justification perhaps derives to some extent from the level of complexity of the internal structures (Dretske 1988, 11). It derives, however, above all from our *interest* in the kind of effects that result from their triggering (ibid., 24f.). Inevitably, therefore, a conception of behaviour along the modified Dretskean lines I have proposed is going to permit alternative classifications, depending on what is taken to be the primary cause. And it will certainly yield cases where thus identifying one particular factor will not be able to shake off the impression of arbitrariness.

Take the sneezing of a hay-fever sufferer. It is equally plausible to name either the disposition of the immune system or the increase in pollen concentration as *the* cause of individual sneezing bouts. Nevertheless, where the relevant internal structure is what we are primarily interested in, we can justifiably pick out its importance by describing it as the "primary cause". The use of the epithet "primary" should thus be understood as relative to the selective and evaluative activity of the observer or sufferer. This is, however, nothing special, as ascriptions of causal relations necessarily involve privileging certain features over others.[10] For this reason, talk of "primary" cause is preferable to the alternative formulation "proximal cause" (Strawson 1994, 292ff.), which misleadingly suggests a purely temporal or spatial qualification.

[9] Cf. Skinner (1953, 27ff., 167ff.) who argues that the "variables of which behavior is a function" are external to the organism and where both "psychic" and neurological inner causes are rejected. However, not all behaviourists repudiate all talk of the inner. See below, Section 3.3.1.

[10] On the relativity of causal ascriptions, see Feinberg (1968, 112ff).

2.3 Motivation and Representation

2.3.1 Representational Causes

The essential step now required in order to begin answering Aristotle's question is the isolation of the specific kinds of cause responsible for motivated behaviour. Reflexes are examples of behaviour a person thinks of as *not* being due to her in the relevant sense. The "internal" causal mechanism that results in a leg shooting forward after a tap on the knee essentially involves the triggering of a disposition. The explanatory structure is no different to that which constitutes a classical disposition such as solubility, where the behaviour results whenever specifiable conditions are satisfied.

It is against the background of such purely dispositional mechanisms that we can begin to grasp the characteristics of the particular kinds of internal causes required for us to talk of behaviour that is "due to" its bearers, or "motivated". Behaviour that is due to its bearers – who we can now also call its "agents" – simply cannot be explained in this way. Whatever external conditions we adduce as putative triggering conditions for some behavioural reaction, it is possible that the reaction either not take place where they are instantiated or else take place where they are not instantiated.

For instance, how someone reacts when faced with a large, ferocious dog may well depend on whether she thinks it is dangerous, whether she believes she should confront dangerous situations, what effect she believes standing still or running are likely to have, the impression she wants to make on her companions, etc. Conversely, someone can suddenly break into a run for all sorts of reasons that have nothing to do with the instantiation of any particular external conditions: he might have remembered that he has left the oven on; he could have developed the erroneous belief that he has left the oven on; he may have hallucinated the appearance of a large, ferocious dog or he might just feel like seeing how fast he can get from *A* to *B* (cf. Geach 1957, 8). In all these cases, the agent's behaviour is primarily caused by what we can, following widespread contemporary usage,[11] call *representations* of features of the world.

One reason we take it that representational factors intervene in human behaviour production is thus explanatory. The explanation and prediction of the behaviour of others would otherwise be impossible. Indeed, as Hume argued, the understanding that such representational features are responsible for much of the behaviour of our conspecifics is cornerstone of our everyday dealings with each other. In his

[11] The controversial details of the analysis of "representation", particularly in as far as it may be thought to facilitate a transition from the physical to the intentional (Tye 1985, 100f.; Dretske 1995, 48ff.), shall not be my concern here. As will become clear in Section 2.4 at the latest, my use of the term does not entail that what is represented is *represented as being the case*, an assumption often made by those theorists such as Dretske and Tye who are generally dubbed "representationalists". It is this narrower use that leads Dretske to deny that desires are representational states (1995, 127).

characteristically overstated phrasing (T II, iii,1), "'tis impossible to act or survive a moment without having recourse to it". This is a lesson that empirical and philosophical psychology both had to re-learn after psychological Behaviourism and its philosophical cousin, Logical Behaviourism. Indeed, a number of theorists influentially hold that the attitudes – representational states – are best understood as theoretical entities postulated in order to explain and predict behaviour (Fodor 1978, 505f.; 1987, 1ff.; Churchland 1981, 68–71).

We should note, however, that the representational character of the causes of much of our behaviour is not merely postulated to solve difficulties in third-person explanation. Although our own representations are normally transparent to us during much of our everyday behaviour, we can become introspectively aware of their mediating role. Greeting someone you suddenly realise you don't actually know or buying a cake it turns out you don't actually like are forms of behaviour explicable by *mis-representations* of the way things are. Noticing that something has gone wrong in our behaviour production can make clear to us what thought contents of ours were responsible for the strange or unpleasant turn of events. In such a context, we can think back to how we had viewed our action-to-be before the event and compare that perspective with how we now see things.

Moreover, bearing in mind the broad set of the types of behaviour that can be motivated (Sect. 2.1), it ought to be clear that there are forms of "mental movement" that are also to be explained by means of representations. Salient examples here are *practical deliberation* and *active imagining*, which appear to involve the production or manipulation of representations as a result of some representation of such processes.[12] Wanting to remember what *X* looks like might lead someone to conjure up an image of *X*; and wanting to decide what to do often leads us to think through the available options. Although we do have an explanatory relation here, it would hardly be plausible to claim that we postulate the relevant representational states for explanatory or predictive reasons.

In Chapter 3, I will argue in some detail that it is important to avoid seeing motivationally relevant representations as theoretical constructs with an exclusively explanatory rationale. Their explanatory importance is nevertheless a good place to begin. In contrast, the "internal cause" of reflex behaviour requires no such talk of representation – neither for explanatory purposes nor to account for related introspective experiences. An external stimulus is generally sufficient for the internal structure to do its work, independently of whether the subject is representing a movement of her leg, for instance expecting it to move or hoping that it will.

The fact that we are dealing here with a decisive difference is given further support by anatomical considerations. Whereas there is strong evidence for the fact that representation is subserved by the neocortex, the automatic withdrawal from

[12] Of course, this cannot mean that the active imagining of what an old acquaintance's face looked like need be preceded by the agent already imagining the face as represented by herself. The details of the imaginative representation may well be inaccessible to the agent prior to the mental action. A parallel point regarding practical deliberation will be of key importance in Section 8.6.

certain stimuli that also produce painful sensations, for instance from a hot oven plate, is caused by the excitation of alpha motor neurons in the spinal cord (Groves and Rebec 1988, 294). Such reactions are triggered before, and independently of, the arrival at the thalamus of the neural impulses transmitted along the A delta and C fibres, an event necessary for pain sensation (ibid, 262f.). This helps explain why we see the reflex withdrawal movement as non-motivated. Note, though, that non-motivated behaviour often feeds into motivated behaviour. Once you feel the pain, you will generally be strongly motivated to continue the movement that was inaugurated in a way that was not "due to you".

These points cohere with the fact that motivational psychologists (Wagner 1999, 2; Franken 1994, 1ff.) sometimes explicitly exclude reflexes from their field of research. On the other hand, the fact that the disciplinary boundary is often not drawn so tightly (Mook 1996, 164ff.) is no contradiction. Rather, as particularly the development of biological psychology has made clear, there are manifold interactions between the motivated and non-motivated components of human behaviour (Toates 1986, 34). Even the pain withdrawal reflex can be modified by inputs from the brain to the motor neurons. For this reason, hot objects that are expensive are dropped less easily than others (Melzack 1973, 165).

2.3.2 *Self-Representation*

There is one specific feature of the contents of motivating representations proximally responsible for motivated behaviour that needs commenting on. That this feature is required follows from the fact that an agent's motivation necessarily concerns *her own* action. The expression "*X* is motivated to φ" is elliptical, in that the infinitive form of the verb "to φ" conceals its subject. If *X* is motivated to φ, then the motivation to φ is necessarily *both* the motivation *of X* to φ and the motivation *that X* φ. An agent's being motivated thus entails her being the bearer of representations among which must be a representation of herself.

Importantly, as Castañeda and John Perry have shown, the self-representation in question must be of a particular kind. In fact, the way it picks out its referent raises doubts as to whether it represents in quite the same sense as do the other components of the contents of motivational states.[13] It neither involves the agent's representation of him- or herself as possessing particular attributes, nor can it fail to pick out its referent. For these reasons, its articulation in language is not a description. Instead, the relevant relation is expressed by the use of the first-person singular pronoun, which refers to its utterer both directly, i.e. without mediation of any predicate, and infallibly (Castañeda 1967, 86f.; 1975, 158f.). We can make no sense of the idea that someone might yearn to be near his beloved, but be mistaken as to who it is he wants to be near her.

[13] For the reasons, Kenny denies that "I" is a referring expression at all (Kenny 1989, 28).

The first-person singular pronoun articulates what appears to be a primitive form of self-reference without which no representations of behaviour-to-be could move their bearers to perform them.[14] Representing the head of department as needing to be at a meeting in 5 min is not going to contribute to my getting up and going if I have forgotten that *I*, unfortunately, am head of department at the moment. (Perry 1979, 3ff.). It seems plausible that this first-person form of reference might also be both ontogenetically and phylogenetically more primitive than the capacity to represent in terms of property possession. We should therefore not be astounded at the fact that we have developed a form of reference that has somehow "improved on" other forms by becoming infallible. Rather, it seems likely that such first-person reference may have come into being before the development of the full-blown capacity to represent by means of the representation of properties. The analytic recalcitrance of the phenomenon would then go hand in hand with the applicability of the concept of motivation to small children and non-human animals.

2.4 Representational Match and Representational Mode

2.4.1 Motivated Behaviour: Necessity of Representational Match

Motivated behaviour of an agent, then, is behaviour – mental or physical; a change or a continuation relative to the preceding moment – resulting from an occurrence in the agent with representational properties, among which is a self-representation of the unmediated first-person kind. These representational conditions on the aetiology of behaviour that can count as motivated are clearly only the starting point. There are forms of both bodily and mental behaviour that meet these conditions without being "due to" their bearers.

Take such phenomena as involuntary twitching, blushing, fist-clenching and increases in muscle tension. Bodily events or processes of these kinds can be caused by representations of either present of future events involving their bearer, for instance, of her sitting an exam or encountering a ferocious dog. Significant changes in the sexual organs can also be representationally induced. And where phenomena such as skin rashes, eczema, asthma and dysfunctions of the bowels and sexual organs are to be explained psychosomatically, what the patients are afflicted with are also the effects of certain of their representations.

In so far as psychosomatic phenomena are understood, they are generally thought of as resulting from psychological processes that centrally involve mental representations, namely from *emotions*. Fear tends to be considered the main culprit. Note, moreover, that the emoting itself also counts as representationally induced

[14] Richard Holton emphasizes the fact that Lewis's concept of de se reference works with an equally primitive notion of self-ascription (Holton 2015).

behaviour.[15] The thought of having to walk past a ferocious dog may be sufficient to instil in the thought's bearer an increased heart-rate, muscular tension and unpleasant feelings.

Clearly, psychosomatic and emotional reactions are in general not "due to" their bearers, although there are plenty of cases – watching horror films and bungee jumping, for example – in which people deliberately act so as to bring about the conditions conducive to such reactions. In spite of the susceptibility of many such reactions to mediated or manipulative production, they characteristically take a hold of their agents in ways that are experienced as purely passive. Psychosomatic *patients* and the bearers of *passions* are at least normally not playing host to the relevant phenomena because they are motivated to do so.

Why is this? One answer might appear to be that these are processes that their bearers can never have *under control*. But must we always have control over whether we do what we are motivated to do? It seems, on the contrary, that there are cases in which we are so strongly motivated to do something our motivation runs away with us. That such cases seem to occur is independent of whether we think that agents might always be able to do something to prevent such loss of control. Someone in the throes of binge drinking or eating may in an important sense have abandoned control of their behaviour, even if it is metaphysically possible for them to regain that control. And that behaviour isn't any less motivated for that.

Instead, the reason why emotions and psychosomatic occurrences do not count as motivated, in spite of being representationally induced, is that they are forms of behaviour no adequate description of which *matches* the representations that cause them, or appears to their bearer to pick out behaviour conducive to bringing about the contents of those representations. Jealousy, worry and contentment result causally from the interaction of representational states, but normally none of those states represent the experience of the emotion that they cause. If a person is motivated to avoid some event, but believes there is a certain probability that she will not be able to avoid it, she may well worry. At least in such standard cases, the worrying does not result from "motivation to worry".

2.4.2 Motivational States: Insufficiency of Representational Match

The importance of representational match was recognised by Hobbes in his specification of the claim that "Voluntary motions" are necessarily "first fancied

[15] To what extent the relevant behaviour is itself mental, and to what extent the relevant mental behaviour is representational, are disputed questions in the theory of emotions. For James (1890, 1058ff.), the essence of emotion is affect and the essence of affect is the representation of bodily changes in the bearer. For Broad (1954), emotions are affectively "toned" representations ("cognitions"), where affective tone is itself non-representational. Goldie has argued (2000, 56ff.) that feelings can themselves be representational ("feeling towards"), whereas Griffiths (1997, 77) claims that neither affective nor "cognitive" phenomena are essential to emotions.

in our minds": "because going, speaking and the like Voluntary motions, depend always upon a precedent thought of whither, which way and what; it is evident that the Imagination is the first internall beginning of all Voluntary Motion" (L VI). Hobbes is here using "imagination" and "thought" interchangeably to pick out a generic concept one of whose specifications is "desire" or "appetite". This terminology reflects the imagistic conception of thought and the genetic theory of its origin in decaying sense that Hobbes shares with Hume. Our distance from such a conception of representational content today is one reason why we are not tempted to talk of "imagination" here.

There is a second simple, but important reason, a reason that specifies why we require more than representational match between motivational states and the behaviour they motivate. The imagination, understood as the mere picturing of some perhaps non-existent particular or the entertaining of a proposition (cf. Scruton 1974, 87ff.), is clearly not the kind of mental state that is itself suited to harnessing the causal mechanisms that lead to the content's realization. Imagining some *p* can be a purely intellectual exercise, although there are cases in which a possible scenario someone pictures can, in being pictured, come to appear attractive to the picturer and thus generate the motivation to realize it. This, however, clearly involves a further mental step.[16]

Hobbes uses the terms "Imagination" and "fancy" to designate the representations of items in the world that have remained after the "decay" of their perceptual representations. His philosophy of matter in motion (Sect. 1.4) enables him to talk as if there is a continuity between the representational reaction to sense input, a reaction he construes as "Outward" "endeavour of the heart" (L I), and the motion "toward" some object constitutive of appetite, which he also calls "endeavour" (L VI). There is, however, something clearly askew in the claim that motivational states are an agent's *imaginative* action representations.

Merely entertaining a proposition is not a motivational attitude. Depending on how we understand the notion of "entertaining", the psychological framing of a proposition that makes it apt to motivate involves either an additional or a different feature than imagination. Clarity on this point is due to the generalisation of a well-known semantic argument of Frege's and the application of the generalised model to attitudinal structures.

Frege established that the assertion of a proposition requires an additional component beyond predication (1918–19, 355f.; 1979, 138f.). He extended this insight to questions, but refrained from applying to other linguistic performances, explicitly refusing to apply it to commands.[17] In the 1950s and 1960s, Richard

[16] Note that if attitudes are only postulated as theoretical constructs "in a mature science which aims at explaining behaviour" (Stich 1979, 27), imagination may drop out of the picture, or at least end up with a negligible role. I would suggest that this tells against seeing the explanatory project as providing the only rationale for granting the existence of the attitudes.

[17] What Geach called "the Frege point" (1965, 449), that the same content may appear asserted or unasserted within conditional assertions, was for Frege of restricted application. For this reason,

Hare and John Searle argued that the distinction is applicable across the board to speech acts. Semantic content can be distinguished from different – assertoric, interrogative, imperative, optative – contexts, within which it can retain its identity. Statements, questions, orders and wishes concerning someone's opening of the door are intuitively all about the same matter, differing only in their "force" or "mood" (Hare 1952, 17ff.; Searle 1969, 29ff.; Dummett 1972, 449f.; Kenny 1975, 37–40).

Whereas Frege's Platonism about "thoughts" led him to retract formulations in his earlier work that admit of a psychological reading (Frege 1879, 111f.; 1979, 198), Kenny and Searle showed that the analysis of speech acts into two dimensions has an exact parallel in psychological attitudes: they can be "about" different matters or the same matter[18] and their matter can be framed in different psychological "modes" (Kenny 1963, 206ff.; 1975, 40ff.; Searle 1983, 6f., 29ff.). Coming to believe that you have some attribute *X* involves framing the thought that you possess *X* in a decisively different manner to imagining your possession of *X*.[19] And, to return to our topic, so does being motivated to bring about your possession of *X*.

Being motivated to φ plausibly involves being the bearer of a representation of your φ-ing in a specific *mode*. I shall come back to this in detail in Chapter 4. For the moment it will suffice that there is an intuitive plausibility to the claim that a specific way of psychologically framing some content is necessary for the bearer of that attitude to be motivated to bring the about the relevant content. Indeed, it is plausibly the presence of this feature that enables us to round up attitudes referred to by a large set of everyday linguistic items and to subsume them under the term "motivational state". In his *Introduction to Motivation* (1964, 4f.), John Atkinson lists as everyday terms for "motives"[20]: "wish", "want", "desire", "long for", "crave", "yearn" and "hanker". All the attitudes thus designated appear to have in common a specific way of representing some action-to-be of their bearer, a representational mode whose presence is decisive for the constitution of the "primary cause" of motivated behaviour.

Peter Hanks says that Frege only advanced "an attenuated form of the content-force distinction" (Hanks 2007, 143).

[18] Castañeda denies that it is possible to think about the same thing in different attitudinal modes. Indeed, according to Castañeda (1975, 158ff.), it is not only impossible for me to have a want and a belief with the same content. It is also impossible for me to want to do the same thing as you want to do. Following Castañeda, Pendlebury has argued that difficulties at the seams between content and force should persuade us to abandon the distinction in the theory of speech acts (Pendlebury 1986, 362ff.; similarly Hanks 2007, 144ff.).

[19] The standard post-Fregean view, represented by Geach and Scruton, is that belief involves more than imagination or supposition, which is the mode-less representation of a proposition. For doubts about this, see the end of Section 4.1.1, especially footnote 4.

[20] It should be noted that Atkinson's generic use of the term "motive" here – as synonymous with what I have been calling "motivational states" – is unusual among psychologists. In social and motivational psychology, the term "motive" generally refers to enduring dispositional structures that explain the formation of specific preferences, independently of whether the bearer of those preferences is aware of why she has them (cf. McClelland et al. 1976, 76–81; Heckhausen 1991, 8). The "achievement motive", the "affiliation motive" and the "power motive" are the three classical species of the genus thus construed.

2.4.3 The Ideo-Motor Theory

It is the failure to distinguish between content and mode that robs the so-called ideo-motor theory of the plausibility that it otherwise might have as a theory of *motivated* behaviour. This theory, which is currently enjoying popularity among empirical psychologists, originates with William James. James claimed that the mere "thought" of some bodily movement can be sufficient to inaugurate that movement, or at least an incipient variant of the movement, which then can only be prevented from coming about by some "conflicting notion".

The attraction of the ideo-motor theory is that it seems to account for the flow of much of our everyday action that is not punctuated by introspectably accessible episodes of conscious "resolve", "effort" or an exertion of "will power". In such cases, there is, as the motivational psychologist Wolfgang Prinz has put it, a "phenomenal continuity between percepts and acts" (Prinz 1990, 171). And certainly, an attempt to understand the specific modal character of motivating states should not seek it in such feelings of effort.

Certain remarks of James do seem to suggest that he sees a mere *perception* as sufficient to inaugurate an action that follows it. An example that may appear to support this reading is the following: "Whilst talking I become conscious of a pin on the floor or some dust on my sleeve. Without interrupting the conversation I brush away the dust or pick up the pin" (James 1890, 1130ff.). And James has indeed been read in this way (Prinz 1990, 171). Clearly, however, although the perception of a speck of dust has a content, its content contains no reference either to the perceiver himself or to the action of brushing it away that follows. There is thus no match between this representation and the action it is conceived as bringing about.

Actually, James seems to believe that the move from perception to action is mediated by an evanescent, but nevertheless introspectible *further* "thought": "I make no express resolve, but the mere perception of the object and *the fleeting notion of the act* seem of themselves to bring the latter about" (James 1890, 1131[21]). But once it is admitted that there is a further intervening representation of the act itself, that raises the question of whether *any* such representation will do. The terms "thought" and "notion", as used here by James, seem to be catch-all expressions for any conscious occurrence with content. James', like Hobbes' talk of "thought" conceals the question of the mode of representation. Perhaps this lack of differentiation appeared justified by the assumption that the only alternative to a "mere" representation is its accompaniment by feelings of effort or by explicit decision. There are, however, no good grounds for postulating such an exhaustive disjunction.

Now, there are a considerable number of interesting behavioural phenomena for which some version of the ideo-motor theory may provide explanations. For instance, empirical evidence indicates that involuntary muscular tension and related

[21] My emphasis.

minimal body movements may result from watching, listening to or reading about the actions of other real or fictional humans or even animals. Lotze (James 1890, 1133) mentions such movements in spectators of fencing, Prinz (1987, 80) in the viewers of slapstick films and Goldman (1976, 79) the movements of the throat musculature apparently caused by mere reading, an activity that is for this reason sometimes forbidden for patients recovering from throat operations. What, however, ought to be clear from the discussion so far is that the causal role of the perceptual states in these examples does not qualify them as motivational states. This is because the behaviour they bring about neither matches the content of the representations nor is the sort of behaviour we understand as being due to its subjects. What is up for explanatory grabs in such examples is *unmotivated* behaviour.[22]

Note, finally, that the explanations of the two spectator cases mentioned by Lotze and Prinz *are* likely to involve mentioning the fact that the spectators "want" something specific – for instance, that certain mishaps not befall the film protagonist. Clearly, though, the contents of these attitudes neither need match what is being perceived nor the ways in which the spectator, or the spectator's body, reacts. I shall argue in the next chapter that spectator cases are of particular importance for an understanding of the nature of wanting.

2.5 The Two Dimensions of Motivation

2.5.1 Motivational States and Motivational Force

I have thus far argued that *motivational states* involve a content represented in a specific mode, a content-modally-represented that can be picked out by a whole set of everyday terms. *Motivated behaviour* is behaviour that only occurs either as a result of its representation in the content of a motivational state or as a result of appearing to be required by, or conducive to the behaviour represented in a motivational state.[23]

However, there is a sense of the term "motivation" which may seem to require no reference to representation. In everyday language, someone can be "unmotivated" in

[22] Prinz (1987, 50) states explicitly that the phenomena plausibly explained by some "ideomotor" mechanism "usually arise unintentionally, frequently even counterintentionally". Their non-motivated character entails non-intentionality, although their being unintentional need not make them unmotivated. I make the case for nonintentional, but motivated behaviour in Section 5.1.

[23] This characterisation offers merely necessary conditions. The reason for the second disjunct is the fact that we can do things as parts of motivated behavioural complexes without having to represent those parts individually (cf. Sect. 9.5.3). But clearly, not everything we do in the course of doing something we are motivated to do is itself something we are motivated to do. Moreover, the concept of motivated behaviour shares the problem of all aetiological concepts – such as hail damage or sunburn – which need not be satisfied if the relevant causal route is particularly unusual, or "deviant".

the sense that they lack what we think of as a kind of *mental energy* requisite for *any* kind of action. This may, so we assume, be the result either of physical exhaustion or of some psychological conditions such as general lethargy or depression. Lack of motivation can thus, in everyday understanding, be a *general* state of a person. When we talk in this way, no reference to any kind of mental representation on the part of the agent appears to be necessary. How, then, are the two uses of the word connected?

A first point is this: where we say that someone is unqualifiedly "unmotivated", what we may mean is that, whatever action predicate we insert in the structure "motivated to φ", it will not be applicable to the agent. She is simply the bearer of no effectively motivating state. Debora, in the grip of a deep depression, who says she has no interest in anything whatsoever anymore and who correspondingly remains in bed for days on end, appears to fit this description.

Lack of motivation, however, can also be specific. This can be so in one of two ways. Little Christopher illustrates the straightforward way: he has been told he has to write thank-you-letters to the relatives who sent him Christmas presents, but is not in the slightest motivated to do so.[24] He is decidedly not the bearer of a motivational state with the relevant letter writing as its content.

The less straightforward case is illustrated by Corinne: at six o'clock in the evening she is in a motivational state the content of which specifies her correcting her students' essays. However, end of term is approaching and she is very, very tired. By half past ten she feels completely "drained" and, as a result, no longer has the energy to complete her project. After a good night's sleep, things will feel different and she will be able to finish what she started the night before. Note that her problem is not that some competing motivation, for instance to watch television, has supplanted her motivation to do the corrections. On the contrary, her lack of competing motivation is made clear by the fact that she ends up sitting at her desk staring into space for a considerable length of time, before coming round and deciding that the only rational course of action in such a state is to get off to bed. However, it isn't the motivation to go to bed that interferes with the correction project. Rather, her "energy" had just "drained away". She has fallen into a state of *motivational inertia*.

There is a clear sense in which, at 22.30, Corinne is no longer motivated to correct the rest of the essays. On the other hand, there is also a sense in which she has not ceased to be the bearer of the representational state that contributed decisively to her correcting essays for the preceding four and a half hours. It seems plausible to say that she will carry on playing host to that state while she is asleep and to explain her speedy completion of the task the next morning by the "regeneration" resulting from her sleep, as a result of which she has the "energy" required to do it. One way of capturing the distinction required here is to say that *motivational states* – the kind of states to which Corinne continues to play host even at 10.30 – are not necessarily

[24] Cf. also the case of little Pia (Sect. 5.2.2).

motivating states. Motivational states are the kind of representational states capable of channelling the "energy" required in order to give rise to motivated behaviour.

In order to see more clearly that agents can be the bearers of *potentially* motivating states that are at least for the moment motivationally inert, consider the way Corinne's capacity to carry out her task gradually seeps away. As the evening wears on, she finds it more and more difficult to keep at her task. Even if the idea of a *completely inert* motivational state sounds like a square circle,[25] surely there can be no doubt that we are dealing with separable dimensions of her motivation: her representation of what is to be done as to be done and the extent to which she is moved by that representation to do what is represented. The idea that the energy available for the task can gradually seep away provides a coherent *limit concept* of a motivational state – perhaps temporarily – devoid of all motivational force.

The distinction between these two dimensions of motivation is familiar within social and motivational psychology. They are often labelled the "directional" and "energising" components (Duffy 1941, 191ff.; 1951, 30ff.; Hebb 1955, 244; Young 1961, 24; Stagner 1977, 103ff.; Gollwitzer et al. 2000, 198). The distinction is also given some support by some classic empirical findings.

In one experiment, cats whose cerebral cortex had been removed reacted to prodding by the experimenter with the usual aggression; however, the resultant aggressive behaviour latched onto random objects such as a table leg rather than the source of the prodding (Duffy 1951, 31). One interpretation of such experiments is that the cats had been deprived of the apparatus necessary to give direction to the "motivational force" aroused by the experimenter's prodding. On the other hand, it may seem more appropriate to explain the cat's behaviour by the simple triggering of what Tinbergen (1989, 42), following Lorenz, called an "innate releasing mechanism". Even if this is all that is left to move the scientifically disabled cat under such conditions, the key question is how the triggering of such mechanisms dovetails causally with the representational states that are plausibly at work when the feline cortex is intact.

It is also worth mentioning a second group of experiments that provide some support for the conceptual bifurcation between motivational states and the physiological mechanisms that can power them. Although these experiments were designed to test hypotheses concerning the concept of emotion, they are, I suggest, better seen as providing support for the claim at issue here.

In the first set of experiments, conducted in the 1930s, subjects injected with adrenalin reported feelings of "unrest", "unease" and "a desire for action", together with a certain frustration at those feelings having "no definite object" or "no content" (Cantril and Hunt 1932, 303–5). In later, more well-known experiments, Schachter and Singer (1962) disguised the content of the adrenalin injection. This enabled them to demonstrate that both the behaviour and the subjective experience of subjects thus physiologically aroused are strongly influenced by their perception of environmental factors. The behaviour of "stooges" was shown

[25] Al Mele has argued that it isn't (Mele 2003a, 26f.).

to be a significant factor in determining how the injectees interpreted their own physiological arousal. Where a stooge behaved euphorically, the subjects tended to interpret their own arousal as euphoria; where the stooge pointed out reasons to be angry, the subjects tended to interpret their own arousal as anger (Schachter and Singer 1962, 383ff.).

Schachter and Singer – implausibly – saw the experiments as supporting the claim that emotions are physiological arousal plus "cognition", where the relevant "cognition" is an attempted explanation by the person of their own physiological arousal. What I think the experiments more plausibly show is that physiological arousal of the kind produced by adrenalin injections constitutes an *unspecific form of behavioural readiness* that is open for representational channelling in such a way as to give rise to specific actions. Such "channelling" or "directing" may, but need not take place on the basis of hypotheses about the cause of one's arousal. It certainly need not constitute an emotion. What is relevant for our purposes is that the supplementary representational factor is both *necessary for motivation* and *separable* – conceptually and empirically – from the physiological phenomenon of "arousal".

2.5.2 Motivational Force and Arousal

These lines of thought may lead one to suspect that the notion of "arousal" might be able to contribute to an understanding of the notion of motivational "force" or "strength" we apparently need to complement the idea of motivational states. Clearly, there is *some* connection between the two phenomena, most obviously between being unspecifically "aroused" and being unspecifically "motivated". As the example of Corinne illustrates, the motivational "energy" to engage in or complete a specific task co-varies, at least to a certain extent, with the positioning of the organism on a general scale that extends from alertness to deep sleep. The persistence, vigour and speed with which a task is approached or accomplished can vary considerably in a manner which may be systematically related to the location on such a scale.

However, there seems at the moment to be little chance of saying anything more precise than this. One reason for this is the lack of a unitary concept of "arousal" (Mook 1996, 241ff.). There are a number of phenomena that can be measured with considerable precision and that are seen as standard indicators that their bearer is in some sense "aroused". These include lower amplitude and higher frequency of brain waves measured by the EEG, decreased galvanic skin resistance, increased skeletal muscle tension, adrenalin level and blood pressure. However, the characteristic correlation between these measures is by no means general. Electrocortical activity and autonomic activity can diverge significantly, components of the latter kind, such as heart rate and blood pressure, can inhibit processes of the former kind (Lacey 1967, 25). Even the phenomena taken as indicators of unitary autonomic arousal can drastically fail to correlate with one another (Lacey 1967, 21ff.).

The disunity of the phenomena summarised under the term "arousal" concerns not only the lack of correlation between the various measures. It also characterises their relations to their bearers' proclivities to behave in ways that are "due to them". Although being strongly motivated may register significantly on the various arousal scales, so does intense *perceptual* and *cognitive* stimulation, as do general states of internal *disorganisation* that are unconducive to action. Further, the presence of the motivational "force" required for most of our actions hardly need be of the dramatic levels brought about by "the four S's" – stimulant drugs, intense sensory input, stress and surprise – often seen as the paradigmatic causes of significant "arousal" level (Mook 1996, 241). On the contrary, as we spend most of our waking hours performing more or less humdrum actions, some relevant physiological processes must be going on most of the time, presumably at levels that are fairly unobtrusive as far as the standard "arousal" measures are concerned.

"Arousal" appears to stand for the activation of a fairly large variety of physiological processes, where the activation exceeds certain levels. The breadth of the phenomena in which it can be involved makes it considerably less specific than the notion of motivational force. Indeed, compared with "arousal", the notion of motivational force has clearer psychological contours – in spite of the fact that we have no precise idea how to measure it physiologically. What justifies the subsumption of what are no doubt highly diverse physiological processes under one concept is a *functional* characteristic: their proclivity to combine with representations to cause behaviour of their bearer aimed at realising the representation's contents. Whatever those processes may precisely consist in – and their investigation is undoubtedly an important task for the behavioural sciences – a philosophy of practical mind can content itself with the claim that there are such processes (cf. Woodfield 1981; 1982, 75).

Note, finally, that an understanding of motivating states as analysable into three components – representational content and attitudinal mode plus motivational force – involves no commitment as to how these components are generated. It certainly need not be taken to suggest that anything like the Schachter-Singer structure – the "channelling" of prior "arousal" by some supplementary representation – is responsible for all our action.[26] The analysis is, on the contrary, open for understandings of the interaction between motivational states and motivational force as two-way. No doubt there are times at which we seek ways to "channel" motivational "energy", manifested perhaps in an abstract urge to "just do something". But the phenomenology of our experience indicates equally that such "energising" phenomena can be *consequent* rather than *antecedent* to the formation of a motivational state. Perceiving an opportunity for an action you had

[26] A radical view of this kind is offered by Nietzsche (FW §360), who pictures the relationship between "force" and representational states by means of the metaphor of steam and helmsman. Nietzsche goes on to suggest what appears to be a form of epiphenomenalism, according to which the idea of a helmsman is actually an illusion. In motivational psychology, drive theory suggested metaphors similar to Nietzsche's, for instance Hebb's picture (1955, 244) of engine and steering wheel.

not previously considered may give rise to a sudden desire which blows away a prior general state of lethargy. And it is surely a part of our everyday experience that decisions can "muster", rather than merely "channel" the "energy" necessary for action (cf. Sect. 8.2.3).

2.5.3 Motivational States and Motivated Behaviour

I began this chapter by recalling Aristotle's question as to what it is "in the soul" that originates "movement". I followed Aristotle in narrowing down the scope of the explanandum to those forms of "behaviour" – understood in a suitably reformed sense – that are "due to" their bearers, that is, which are intuitively seen as their active products. These forms of behaviour can be designated "motivated behaviour" and their psychological causes "motivational states".

This move establishes a conceptual link between the two terms of the explanation. Following the definition of "behaviour" in terms of causes internal to the behaving entity, motivated behaviour is defined in terms of specific types of internal cause. Moves of this kind, which are characteristic of the causal theory of action, have been repeatedly felt to be problematic. According to the once popular "Logical Connection Argument" (Melden 1960, 482; Taylor 1966, 52; von Wright 1971, 110ff.; Alvarez[27] 2007, 117ff.), the conceptual connection thus generated excludes motivational states and motivated behaviour being causally related.

The worries thus articulated are, however, only of any substance if the two terms of the conceptual relation are *entirely* co-dependent, i.e. if the only notion we have of motivated behaviour is of behaviour caused by motivational states *and* if the only notion we have of motivational states is of states that cause motivated behaviour. But this is not the case.

Certainly the term "motivated behaviour" establishes a conceptual tie between particular kinds of events we can be involved in and their aetiology. Those kinds of events are, however, already by and large phenomenologically identifiable as those forms of our behaviour in whose performance we take ourselves to be actively involved. "Motivated behaviour" is a term introduced to clarify the idea of activity at work in our everyday understanding. We can, of course, give the forms of behaviour a different label, for instance calling them "actions".

Still, the definition of the forms of behaviour intuitively identified as active does rely on the aetiological category of causation by states of a certain kind. The simple reason why this is not circular is that the logical connection running from "motivational states" to "motivational behaviour" does not run in the opposite direction. "Motivational states" are not defined in terms of their action-causing properties. Three important considerations are indicative of the conceptual gap here.

[27] Alvarez claims that the argument should be understood to concern not a logical, but a "conceptual" connection.

First, it is conceivable that motivational states, where they do not occur in combination with some level of motivational force, do not cause motivated behaviour. Their specific mode of representation is not reducible to their action-causing propensity. A first example that shows this is that of Corinne.

Second, the assumption of the advocates of the Logical Connection Argument that motivated behaviour is the *only* criterion for the existence of the kind of states apt to cause it, is quite simply false. There is, as I shall argue in the next chapter, a whole *syndrome of effects* characteristic of representations in the relevant mode. This becomes particularly clear once the representational states and their – variable – motivational properties are seen as conceptually separable.

Finally, although the Logical Behaviourists were right to reject the Cartesian claim that we have access to our own mental states through the medium of an inner sense (Kenny 1989, 8; 89f.), the denial that we have any immediate, i.e. non-inferential access to the kinds of states that lead us to act was a massive overreaction (cf. Moran 2001, 5). This is perhaps an uncomfortable datum for a philosophy of mind that aims at scientific respectability. Nevertheless, it is something a theory that is adequate to the phenomena will have to takes pains to account for. As long as we at least sometimes have such immediate access to our motivational states and these very states can move us to act, there can be no question of a logical connection here excluding in principle an explanatory model of the kind I am canvassing.[28]

The following chapters aim to clarify the concept of the kind of states apt to motivate behaviour, that is, to mobilise motivational force, whilst recognising that there is no conceptual necessity that states of the relevant kind be effectively motivating.

2.6 Excursus: Motivating Representations in Non-Human Animals

Only creatures whose behaviour results from their representing that behaviour in a mode distinguishable from belief can be appropriately described as behaving in the way they do because they are motivated to do so. The much-discussed question as to whether the behaviour of at least some non-human animals is best explained by the same mechanisms as human action – i.e. whether Aristotle was right to see (some) behaviour of animals as up to them (NE 1111b8-9) – can therefore be approached by asking two questions: first, should the creatures in question be seen as bearers of mental representations? Second, do we have grounds for attributing to them the capacity to represent contents in different attitudinal modes? The strongest positive evidence for an affirmative answer to the former question turns out to be the best evidence for an affirmative answer to the latter.

[28] I return to worries about logical connections in detail in Section 4.4.

Much of the literature on these questions has focussed on sceptical arguments advanced by Davidson and Stich (Davidson 1975; 1982a; 1997; Stich 1979; Lockery and Stich 1989). The first of these grounds in the referential opacity of intentional state ascriptions, the second in the holism of the mental. A third argument of Davidson's concerns a feature he sees as constitutive of believing, namely a basic form of normativity. I think Davidson is right about this feature, but wrong about its consequences. I will say something about this at the end of this subchapter, but will only be able to clarify its consequences in Section 4.5.3, after I have explained the sense in which both believing and wanting are normative states. I therefore begin by focussing on the first two sceptical arguments. Davidson has argued that the problems picked out by both – in fact by all three – arguments result in turn from the fact that, or at least in as far as, non-human animals are not language users.

According to the first argument, ascribing mental states to non-linguistic animals is problematic because we have no way of knowing by means of what features animals represent entities in the world. According to the second argument, representations require concepts, and concepts only have content in virtue of being part of a web of concepts that mutually inform each other. The susceptibility of a representation to accurate ascription through higher-order representation by an ascriber depends, so the linguistic version of the argument goes, on there being some publicly accessible route to the representation-to-be-represented, a route that can only be laid out in language. Similarly, the content-conferring embedding of a concept in a web of interrelated concepts has also been claimed to require language.

However, it is anything but clear why either line of argument should be thought compelling. Certainly, we have no way of knowing how a wild animal thinks of its prey or how a domesticated animal thinks of its owner. Moreover, were a dog to be able to think of its master as owning it, it would require concepts such as property and legality, which it obviously can't possess. Nevertheless, the first problem is, on the face of it, epistemic in nature, whilst the second may simply be a matter of quantity and complexity.

We indeed have no way of fully respecting the constraints of referential opacity in ascribing intentional states to non-linguistic animals. But that doesn't entail that animals have no way of representing items in the world and their relations.[29] Moreover, there seems to be no reason of principle why some non-linguistic animals might not be furnished with fairly small and simple representational webs. A chimpanzee's representation "red colobus monkey" is not going to involve zoological classification, but might involve such contents as being a self-mover, edibility or tastiness, perhaps of tasting and smelling a certain way and of tree-climbing. Again, the chimpanzee thought that identifies something as a tree climber won't involve our concept of tree, but perhaps something climbable in a certain way, suitable for nest-building, etc.

[29] Allen and Bekoff emphasize the epistemic character of the sceptical argument in the hands of Stich (Allen and Beckoff 1997, 80f.).

If these sceptical arguments fail, the question is whether we have good positive reasons to think of at least some animals as behaving as they do because of representing their behaviour-to-be in a relevant psychological mode. We do, as can be seen by attention to prominent work in what is somewhat misleadingly called "animal cognition".[30] Broadly, the reason is that at least some of the behaviour of apes, corvids, dolphins and domestic dogs is *flexible* in such a way that its best – systematic, non-ad-hoc – explanation is in terms of representations. In the relevant cases, there are categorically different mechanisms at work to those controlling such activities as the stereotypical nest-investigating of the digger wasp, activities whose sequence is completely fixed once a stimulus has set the ball rolling. Beyond such cases of what Lorenz called Fixed Action Patterns, animal behaviour appears explicable by mechanisms with varying degrees and dimensions of flexibility. The burden of this subchapter is that there is a point on a scale of increasing flexibility after which inference to the best explanation unequivocally picks out representational causes. There is plausibly some later point on the same scale after which the satisfaction of further conditions justifies ascribing the animals concepts. Now, there are no knock-down arguments as to when talk of concepts becomes justified. Various authors have put forward different proposals, but I suspect the differences are primarily a matter of nomenclature. In what follows, I will sketch a series of cases in which the animals concerned display increasing degrees of flexibility and who thus have claims of increasing strength to be seen as possessors of concepts. For someone with highly demanding conditions on concepts, they may only be in play by the last of my examples[31]; others will see little reason not to talk of them in earlier cases. However the concept of a concept is precisely analysed, an analysis compatible with the empirical data requires that there be groups of cases in which some, but not all the conditions are satisfied.[32] As my interest here is in the question of whether the animals are bearers of motivating representations and as I assume that these may be in place without qualifying as conceptual, I can happily avoid the question of the conditions sufficient for attributing concepts.

(1) A first form of flexibility is codified in a weak version of Evans' generality constraint, a constraint normally taken to pick out concepts. According to this requirement, we should only think of a mental episode as involving representations if it has a detachable subject-predicate structure *Fa*, the components of which are re-combinable with at least one further subject component and at least one further

[30] For "cognition", read "representation". The relevant "cognitive" states include motivational states.

[31] For Davidson, and others who require language, even the chimpanzees in the last example must fall short.

[32] Talk of "proto-concepts" is a somewhat helpless attempt to deal with this fact (cf. Dummett 1993, 125). Hanjo Glock reacts to the problem by drawing strict lines around the use of concepts – unlike Dummett, within animal behaviour – construing forms of pre-conceptual thought as "holodoxastic" (Glock 1999, 181; 2010, 19ff.). (Presumably non-conceptual motivation should be "holo-orectic".) In my claim that we should reckon with less-than-full forms of whatever it is we draw the line around, I am in agreement with Susan Hurley (Hurley 2006, 152f.).

predicate component, viz. *Fb* and *Ga*.[33] The constraint need not be seen as setting enormously high standards, as it simply requires that attitudes come in at least minimal constellations: a chimpanzee can only believe that a red colobus monkey has gone up a tree if it can also believe, for instance, that a bush baby has gone up a tree and that a red colobus has escaped.

Whether a creature possesses the capacity to represent, and recombine representations of, particulars and ways they can be may appear to be inferable from patterns of constant reaction to varying situations containing the relevant particulars. This is, however, only going to be plausible where we see ourselves as justified in ascribing motivational states with which representations of the way things appear can combine to produce further forms of motivation. The plausible ascription of flexibility of representations is thus bound to flexibility in the combination of those representations with others in a different psychological mode. That is, our only access to something like a subject-predicate structure in non-linguistic animals is via the ascription of attitudes in differing psychological modes. Flexibility in the recombination of representations of particulars and ways particulars can be is tied to flexibility in the adoption of means to ends.[34]

Cases of flexibility of the adaption of means to end appear to be provided by the findings of Tetsuro Matsuzawa's research team concerning the cultural variation between the chimpanzees in the geographically close sites of Bossou, Seringbara and Yealé. Although the neighbouring chimpanzee communities live under more or less identical conditions, they have developed and passed on significantly varying kinds of feeding, tool-constructing and tool-using behaviour (Humle 2003, 147ff., 213ff.). The learning of means-ends connections that could have been different, as they are in the next community, seems to presuppose the representation of the form of behaviour adopted independently of a hard-wired determination to work toward the end-state. For whatever reasons, members of one chimpanzee community (Bossou) appear to represent pestle-pounding the fruit of the oil-palm tree as a means of getting food, whereas no such representation has been developed at the other two sites, the inhabitants of one of which (Yealé) gain food from the tree in various other ways, whereas the chimpanzees at the third site (Seringbara) don't see the tree as the source of food in any form at all. The varying ways of dealing with the need for food only seem explicable by cultural processes, the explanation of which in turn appears to require the development and transmission of representations of the tree or other features of the environment. Such a case provides no evidence of spontaneous insight on the part of individuals. Nevertheless, it is hard to see how

[33] On the distinction between strong and weak versions of the constraint, see Carruthers (2009, 96ff.).

[34] Eric Saidel and Elisabeth Camp both argue that key evidence for the mentality of non-linguistic animals is provided by evidence for their ability to flexibly combine and recombine means and ends (Saidel 2009, 39ff.; Camp 2009, 292ff.). Hume argues in the *Treatise* that "adapting means to ends" is the obvious criterion (T I, iii, 16). I am basically agreeing, whilst arguing that we need to be extremely circumspect about what "adapting" might mean in this context.

the adoption of such different means to the end of feeding could be explained if not via the development of something like beliefs, which in turn generate motivation to adopt specific means.

Certainly, it is not impossible to see such cases as resulting from "associative learning". But it is far from easy to imagine an empirically plausible story of how chimpanzees could be waving around rocks or branches that happen to connect in the right way with fruit so as to produce an edible result without the relevant movements being purposeful. Some such thing is not inconceivable, but it certainly doesn't look like the best explanation to which we should be inferring. Moreover, the cultural transmission of specific ways of extracting food from the tree requires learning by observation, a process that itself poses a serious problem to explanation in associative terms. As Papineau and Heyes point out for an analogous case, conditioning could only explain why an observer associates seeing a conspecific moving the stone with seeing that conspecific receiving the food "reward" (Papineau and Heyes 2006, 190). What it seems unable to explain is how the observer can come to associate similar behaviour on its own part with *its* receiving the food reward itself.[35] Thus it looks very much as though both the development and cultural transmission of locally specific solutions to problems universally faced by species members require the mental features constitutive of motivation.

(2) Where novel forms of behaviour are developed by an *individual* in response to the requirements imposed by a new situation, the case for explanation via the generation of new beliefs is still stronger. The paradigmatic case here is the fabrication of tools. Where the tool is constructed spontaneously in order to solve a one-off problem, an inference to any other explanation would appear outlandish. The New Caledonian crow, Betty, who in Alex Kacelnik's laboratory used various techniques to bend wire, a material not available in her natural habitat, thus creating a hook with which she could lift a bucket containing food out of a vertical tube, is surely acting so as to attain a goal that is hers (Weir et al. 2002).

Now, Weir and Kacelnik caution against concluding too quickly that the behaviour of New Caledonian crows like Betty is to be explained by "cognitive" representations, since it is as yet unclear to what extent these corvids are fashioning appropriate means because of their *understanding* of the relevant causal relationships or whether their behaviour is the result of "within-task trial-and-error learning" (Weir and Kacelnik 2006, 318). However, the ability to represent some form of behaviour as a means-to-be-adopted in order to achieve some end doesn't require sophisticated causal understanding. We humans frequently adopt means to ends because they have worked in the past, although we know little

[35] Papineau and Heyes suggest that a mechanism facilitating "sensitivity to demonstrator reward" is conceivable in the case of quails observing the feeding of conspecifics whilst themselves feeding – a naturally given situation, as quails tend to feed together. In such situations, the observer might come to associate its own food reward with seeing a conspecific feed. However, an analogous association-based explanation of learning by the Nimba mountain chimpanzees wouldn't even get off the ground. Pestle-pounding is not something chimpanzees are hard-wired to do in the company of conspecifics – or alone.

about the mechanisms supporting the connection. Certainly, the more a creature understands causally, i.e. the better it is able to represent causal connections as obtaining across situations, the more flexible it can be in seeking ways of attaining its ends. Conversely, the greater the flexibility of a creature's behaviour when faced with novel obstacles and furnished with unfamiliar materials, the stronger the case for representational explanations of the resulting behaviour. Clearly, though, the ability to represent in terms of nomological concepts is an advanced conceptual ability. Whether or not Betty possesses it, she is able to work towards her goal by developing varying techniques in dealing with materials with unfamiliar properties (the stable pliability of wire) to reliably produce a tool she uses immediately to get the food, but which she has no tendency to fabricate in the absence of any such hindrances. Betty has the goal of obtaining the food and is therefore motivated to seek and produce means of doing so.

(3) A farther-reaching form of flexibility characteristic of the human pursuit of goals requires the ability to represent states of affairs as obtaining outside the situation in which the representation takes place. Again, the evidence for such trans-situational representations is clearest in the case of motivating states. If a creature is planning in situation s_1 to carry out some action in s_2, then it is clearly in the business of representing states of affairs with a significant degree of stimulus-independence. For behaviour in s_1 to count as a preparatory step in a plan to be realised in s_2, it cannot be purely instinctive or the result of operant conditioning. Rather, it must have been chosen in s_1 because the animal takes it to be conducive to the realisation of its end in s_2. The scrub-jays studied by Nicola Clayton are the clearest examples of this ability in non-linguistic creatures. They have shown themselves able to differentially cache food for situations in which they have been led to expect that they would otherwise have none and to develop appropriately distinctive forms of caching behaviour for perishable and non-perishable foods depending on whether they had come to expect that the preferred, but perishable food would degrade by the time of recovery (Clayton et al. 2005; 2006; Raby et al. 2007).

The capacity for trans-situational planning seems also to be available to primates. In experiments conducted by Mulcahy and Call, bonobos and orangutans selected tools appropriate to a food recovery task, kept them whilst away from the apparatus and returned with them to recover the food after a 14-hour delay (Mulcahy and Call 2006). Because this behaviour involves not only the flexible choice of an appropriate behavioural means to some end, but also the adoption of behavioural means in s_1 to ensure that the behavioural means chosen for s_2 is indeed available in s_2, there are no explanatory models that can seriously compete with a motivational explanation of the apes' behaviour.

Return now to Davidson's sceptical challenge: Davidson wouldn't accept any of these data as evidence for animal mentality. This follows from his positive criterion for the ascription of beliefs, together with the assumption that the propositional attitudes – at least beliefs, desires and intentions – come together in a package (Davidson 1982a, 96). Whatever mechanisms may be at work in the differential control of non-linguistic animals' behaviour, these cannot count as mental, he assumes, because they cannot be employed in the light of a notion of objectivity, that is, of the distinction between appearance and reality. He sees this contrast and the associated normative concept of a mistake as the key to belief and, as beliefs only crop up in the company of desires, as an essential condition of any mentality at all. The concept of objectivity, he thinks, in turn requires the coordination between agents of ways of interacting with elements of the environment, a form of coordination that is stabilised in a way that only language can achieve (Davidson 1982a, 105; 1997, 130). He offers no argument as to why it has to be language that fulfils this function, in particular as to why the pre-linguistic "triangulation" he himself describes (1997, 128ff.) is insufficient.[36]

I want to close this chapter with a word about why pre-linguistic triangulation is a more subtle instrument than Davidson recognises. Pace Davidson, I think it can provide clear evidence of mentality. However, as it can come in grades, it may only be sufficient for the stirring of mentality in ways that don't yet involve the full Davidsonian package. My claim is that certain empirically observed constellations among chimpanzees are sufficient for motivational states, although they aren't sufficient for full-blown beliefs. This, as I shall argue in Section 4.5.3, is because, in spite of the fact that both believing and wanting involve normative dimensions, they do so in decisively different ways.

Triangulation involves the orientation of (at least) two agents to each other's orientation to features of the environment, to which they are themselves orientated. Thus described, triangulation satisfies a strong reciprocity condition – Davidson requires "mutual react[tion]" to the environment (Davidson 1997, 130). In an experiment conducted by Brian Hare, this strong condition is not fulfilled: a chimpanzee reacts differentially in the presence of a dominant conspecific to the availability of food, depending on whether, as is visible from the standpoint of the subordinate, the dominant can or cannot see – or has or has not seen – the food (Hare et al. 2000; 2001). The subordinate chimpanzee clearly orientates his behaviour in the light of the information available to his conspecific relative to the environment to which he is himself reacting. This modification of behaviour in the light of information about information available to a competitor – call this *one-way triangulation* – is sufficient for the chimpanzee to be the bearer of states representing a goal: we have no other explanation of why his behaviour changes are contingent on the perspective available to his conspecific apart from in terms of his own representations. As Call and Tomasello remark, "the only reasonable conclusion" from such experiments is that the subordinate chimp represents what the dominant represents or doesn't represent from his perspective, thus generating differentially modified motivational states (Tomasello and Call 2006, 376; Call and Tomasello 2008, 190).

One-way triangulation thus seems to be sufficient for motivating representations. Nevertheless, the lack of genuine reciprocity, that is, of mutual adjustment, is

[36] Glock claims that Davidson's "lingualism" is a priori (Glock 2010, 28). It is unclear to me whether Glock is right. The kind of argument Davidson says he needs here, but doesn't provide, could well be empirical.

cause for doubt as to whether the subordinate's representation of the dominant's representation should count as a belief. What the subordinate needs to distinguish in order to appropriately modify his striving and thus count as the bearer of genuine goals are representations from two different perspectives. What he doesn't require is an objective standard against which his competitor's and his own representations can be measured and in the light of which they can count as correct or incorrect. In Chapter 4, I shall be arguing that belief does indeed require the conception of such a standard, as Davidson thinks. For this reason, attributing genuine beliefs to non-linguistic creatures requires stricter conditions than are satisfied in Hare's experiment. Being the bearer of a motivating state is, however, less demanding. Such states are plausibly at work in this last case – as in the cases of the planning and tool-using chimpanzees and corvids described previously. As I shall argue explicitly at the end of Chapter 4, this asymmetry between belief and motivation grounds in the fact that, although there is a sense in which both kinds of state involve standards, the way they involve them is very different.

My conclusion here is that some non-human animal behaviour is motivated. The challenges for research on animal mind concern the degrees, dimensions and extensions of the representational capacities at work, not whether they are the exclusive province of humans.

Chapter 3
Wanting* and Its Symptoms

Before, in Chapter 4, I offer a constructive proposal as to how the modal component of motivational states is to be understood, there are a number of preparatory steps to be taken. These should bring clarity on the requirements such a proposal has to meet. The steps are three in number.

Firstly, the modal component needs to be isolated from the features that naturally accompany it and which are correspondingly also denoted by the everyday terms by means of which it is generally picked out. A look at some everyday linguistic data and particular kinds of examples (Sect. 3.1) offers substantial support for the claim that both the doxastic and the motivational strength components generally referred to by the everyday expressions "motivated to", "want" and "desire" are variable, supplementary features that can be analytically peeled off from a modal core.

It follows that the modal core conception of wanting entertains no *necessary* connection to the disposition to act so as to realise its content. Rather, the core component of wanting that *p* is a representational state that *characteristically* mobilises processes that lead to actions to bring about *p*. However, it also characteristically gives rise to a whole syndrome of *other* effects, both agential and non-agential (Sect. 3.2). Understanding wanting involves understanding what attitudinal feature typically brings about this constellation of effects, although it is not constituted by any of them.

In a third step, this conception is contrasted with the model proposed by conceptual functionalism and its relatives, which define "desire" simply as whatever it is that has these typical effects. In Section 3.3, I discuss the functionalist conception, arguing that it obscures central features of the phenomenon of wanting, in particular its essentially first-person and practical character. It is *both* these features *and* the attitude's causal role that have to be accounted for by a theory of wanting (Sect. 3.4).

N. Roughley, *Wanting and Intending*, Philosophical Studies Series 123,
DOI 10.1007/978-94-017-7387-4_3

3.1 Wanting*: Factoring Out Believing and Fuelling

In Chapter 4, I shall advance a constructive proposal as to how we should understand the attitudinal mode constitutive of motivational states. Clearing the way for that proposal requires as a first step clarity that the modal element of motivational states can be present where it would be inappropriate to talk of "motivation" and that, conversely, everyday talk of "motivation" implies, or perhaps implicates the presence of factors that go beyond the presence of that key attitudinal feature. One such factor, which the example of Corinne (Sect. 2.5.1) has already highlighted, is the presence of *fuelling* or *energising* mechanisms. Another is that a person characterised as "motivated to φ" will normally be thought to satisfy certain *doxastic* conditions. Our analytic purposes require us to prise these three factors apart.

*3.1.1 Belief, Wanting and Wanting**

I begin with the question of doxastic conditions. In everyday language, there is something askew in the sentence "He is motivated to φ although he believes that φ-ing is impossible for him". Things can be easily righted by transforming the conjunctive sentence into a conditional one: "He would be motivated to φ, if he didn't believe that φ-ing was impossible for him" is fine. But, obviously, this sentence ascribes not actual, but counterfactual motivation.

It is not easy to say *what precisely* is conveyed doxastically by the use of the expression "motivated to". Must the person of whom the feature is predicated have beliefs as to the feasibility for him of the relevant action or is it sufficient for him *not* to have beliefs as to its *in*feasibility? One reason why the question is not easily answered is that the verb "to motivate" in everyday parlance is an import from more technical contexts. However, it seems that we might coherently describe someone as "motivated" to perform some action the question of whose feasibility he has never considered and about which he has thus formed no beliefs. If this is correct, then the doxastic condition in play here is purely negative, specifying merely that an agent motivated to φ cannot be the bearer of a belief that he is unable to φ.

Note that the everyday terms that Atkinson (Sect. 2.4.2) sees as the non-scientific placeholders for the concept of motivation have differing doxastic implications. To say that someone "craves" or "yearns for" some *X* seems neither to entail nor to implicate anything as to the person's subjective probabilities of attaining *X*. Things appear to change once one uses the everyday verb "to want". There is frequently something logically odd about a sentence that provides the verb "to want" with a linguistic object which represents something whose realisation the attitude's bearer explicitly believes to be impossible. This is immediately obvious where the content is indexed at a time prior to the time of the attitude's being held. No adult English language user with a normal belief set would say "I want to have seen the match yesterday", just as they would not assert "I want to run faster than the speed of light". In cases like these, the verb we would normally choose is "to wish", which is generally distinguished from “to want” by its sceptical character relative to the realisability of its content. Normally, the content of the attitude is expressed in the subjunctive or the conditional.

Nevertheless, an agent’s belief that it will be impossible for her to φ in some specific context doesn’t necessarily seem to preclude the acceptability of saying that she wants to φ. “I want to go out tonight, but I can’t because I’m looking after the children” appears to be more or less in order.[1] Thus the doxastic condition on wanting to φ cannot be the absence of *any* belief that one’s φ-ing is impossible. A belief in empirical impossibility in the circumstances appears compatible as long as the action is one that wouldn’t be impossible for the particular person under other circumstances. It looks like there is a limit to how far removed these circumstances might be. Assuming it is possible that I could learn Swahili, it is empirically possible that I could have a conversation in that language. Nevertheless, it would be strange for me to say that I want to talk to a speaker of the language in her mother tongue. Even “I want to be able to have a conversation in Swahili” appears odd if I have no plans to acquire the necessary ability. There appears to be some room here for some divergence in linguistic intuitions. However, what is surely incompatible is a belief in the logical or physical impossibility of the deed for beings with more or less my capacities.

However the lines are precisely drawn here, wanting in an everyday sense appears to involve a specific framing of a representational content in conjunction with a negative doxastic condition. There are two possible explanations for this. The *first* would be that there is a feature of wanting that for reasons of *consistency requires* the absence of the relevant beliefs in the impossibility of the content’s realisation. The *second* explanation would be quite simply that everyday English picks out a *frequent conjunction* of some generic representational mode plus negative doxastic conditions. Were the first explanation to be the correct one, wanting something and craving or yearning for it would, because of differences in their doxastic implications, have to be thought of as unrelated mental states. But surely this is implausible. Yearning for something seems to be more or less the same as wanting it, minus – or perhaps with significantly weakened – feasibility beliefs and conjoined with unpleasant feelings at the want’s non-realisation.

Where someone takes some proposition to be logically or, under the circumstances, physically or even psychologically unrealisable for her, she may yearn for its realisation. Less dramatically, she may say she “wishes” it were, or “would like” it to be the case. Someone might “wish” he could travel backwards in time or “wish” he could leave his wife. He might say he “would like to” be able to run faster than an Olympic champion or to become a famous pop star. The latter idiom seems, interestingly, to be fairly resistant to doxastic variation. People say that they “would like” to do something they believe impossible, unlikely or fairly probable.

[1] There actually seems to me to an element of tension here, an element that disappears if the sentence is transposed to the third person. This suggests that the tension is pragmatic, rather than semantic.

It is surely natural to think of the attitudiniser in all these cases as having a certain kind of attitude with the relevant content, an attitude that is further qualified in doxastic or hedonic terms. Certainly, it seems unlikely that the attitude in question ever exists in a pure, unqualified form (cf. Sect. 4.2). Nevertheless, this core concept is what is instantiated by an attitudiniser who either “wants” or “desires” or “longs for” or “craves” or “wishes for” or “would like” ... some thing or other. It is designated by a term that would fill in the blank in sentences such as “Wanting to φ is ——ing to φ whilst not believing that φ-ing is logically or physically impossible for oneself...”; “yearning to φ is ——ing to φ whilst experiencing disagreeable feelings as a result of one’s not φ-ing ...”; “wishing to φ is ——ing to φ and believing that one’s φ-ing is at least fairly unlikely ...”.[2] Alternatively, the attitude’s extension could be disjunctively specified as “wanting to φ or longing to φ or wishing for ones φ-ing or yearning to φ ...” (cf. Kenny 1963, 215). As a placeholder, I shall use the technical term *want**. Alongside the asterisk, perhaps what Frankfurt (Sect. 1.6) called the “abomination” of using the word as a singular noun will keep in mind the fact that it designates the attitudinal core of the various compound states I have been discussing.

I am thus claiming that the second of the possible explanations for the doxastic conditions on everyday wanting is the correct one. This claim is supported by two further sorts of empirical data. Firstly, it is fairly clear that the development of children’s use of the verb “to want” involves the learning of doxastic restrictions: a young child might say it “wants” to eat all the chocolate in the world or to fly on a magic carpet. The difference between adults’ and children’s criteria for the use of the word “want” may plausibly be thought to have one or both of the following explanations: on the one hand, the child may not as yet have learnt to distinguish precisely between what he can bring about and what is outside the sphere of his influence. On the other hand, such knowledge as he has acquired may not yet have been brought into a systematic relationship with the kind of representational contents he is picking out by his use of “want”.[3]

A second source of support for the claim that there is a unitary core component of wanting* seems to be provided by linguistic data from languages other than English. The English language works with a lexical series that moves from “wish” via “want” to “will”, a series that, among other things, involves an increase in subjective probabilities. German, in contrast, makes do with the two terms “wünschen” and “wollen”, where some of the tasks of the English “want” are fulfilled by the former term, some by the latter. The German language certainly makes no clear lexical cut at the point where English speakers would cease using “wish” and begin using

[2] The precise details of the explications here offered exemplarily are not important for my central claim.

[3] The developmental psychologists Bartsch and Wellman (1995, 96) report that children in their second year of life begin to attribute “desires” at least 7 months before they start ascribing beliefs. Further, they found (1995, 70–72) that in their sample 97 % of all references to “desires” were accomplished by means of the word “want”, which was also employed where the children took some object to be unobtainable.

"want". This speaks for the claim that the lexical severance points in English are a matter of convenience and not of any deeper significance for the structure of the relevant psychological kinds.

Once the plausibility of this move is accepted, this opens up the way for the insight that wanting*, the core of what I have been calling "motivational states", can also be present in cases that do not essentially involve motivation. In what follows, I shall embed the technical term "want*" in everyday English sentences. At this stage, the term is merely a placeholder. Should this appear to provide difficulties of interpretation, the reader may replace it with a disjunction of the terms of the sort grouped together by Atkinson (Sect. 2.4.2). The artificial term is preferable to the widespread, but philosophically disingenuous use of "desire", which suggests, contrary to fact, that we are dealing with an everyday notion (cf. Sect. 4.2). What philosophers subsume under the term tends to be far broader than the everyday sense (Sect. 1.5). The question as to what makes such generalisation legitimate is therefore paramount.

3.1.2 Wanting and Motivation*

A first clue as to the divergence between wanting* and motivation lies in the fact that we don't only "desire" or "long to" *do* things, "have an eye to" or "set our hearts on" doing them. We also desire them or long for them *to be*, or fancy them being the case. Although the latter kinds of state are missing the special form of first-person reference required by states that become effectively motivating (Sect. 2.3.2), it is surely no coincidence that many of the same terms are used to pick out mental states of both kinds. Unsurprisingly, people sometimes long to bring about some *p* because they long for *p* to be the case.

Someone who yearns for an alleviation of the living conditions of the poor may well for that reason herself take steps to contribute to the alleviation of the poor's living conditions. Strictly speaking, an attitude whose content has no first-person index should not, I suggest, be seen as qualifying as a motivational state, as the harnessing of motivational force requires channelling by means of first-person reference. Nevertheless, there are obvious reasons why the agent's yearning in this case is sometimes characterised as a "motivating reason" (Smith 1987; 1994, 92ff.). Her taking steps *to* alleviate the plight of the poor makes sense in the light of her yearning *for* such an improvement. The plausibility of characterising someone's wanting* some impersonal content as a motivating reason depends on the fact that the want's* causal role is unproblematically mediated by a first-person indexed attitude.[4]

[4] What kinds of entities we *should* see as motivating reasons is a further question. The answer depends on the correct understanding of the concept of a reason, in particular whether it is unitary notion (Dancy 2000, 2) or two different categories covered by one term (Smith 1994, 96f.). For the record: I don't think that motivational states themselves are reasons, as I take reasons to be necessarily propositional in form. That an agent yearns for something may be a reason, as may

"Wanting", "desiring", "craving for" or "fancying" the instantiation of some non-first-person indexed state of affairs is thus at least one step removed from being in a motivational state. Nevertheless, it seems clear that we are dealing in both cases with an attitudinal framing of the same sort. If this is correct, we can want* something without being actually motivated to realise it, although cases of the kind just discussed seem obviously to involve a disposition to the relevant behaviour once the minimal, and generally natural step of putting oneself in the frame has been taken.

The gap between wanting* and being motivated widens if we turn our attention to wants* concerning states of affairs taken by their bearers to be *outside the sphere of their influence*. Because of the importance of such examples in understanding the phenomenon of wanting*, it will be convenient to pick them out terminologically. I shall refer to them as *OSI cases*. In such cases, the everyday English "want" remains appropriate as long as the speaker does not take the realisation of the attitude's content to be logically or physically impossible. Thus, we say, we "want" it to rain or a certain team to win a football match.[5] At least under many conditions, neither want* is likely to generate wants* representing instrumental actions of its bearer and is thus not going to harness the body's energy resources in such a way as to count as motivational. Of course, wants* with either content *can* do so. People perform rain dances or bribe referees as a result of forming such wants*. However, the formation of such instrumental action wants* is no necessary part of wanting* the relevant states of affairs to come about. The belief that there is nothing one can do to bring it about simply leaves no rational room for instrumental wanting*.

Moreover, other conditions may also exclude the generation of instrumental wants*. For instance, wanting* a team to win a match, for at least some people, involves wanting* a win of that team to result from events on the pitch that conform to the rules of the game. At least for those of us that watch the match on television, that tends to exclude wanting* to influence the events ourselves. No doubt, chanting fans at a football ground are generally acting in part instrumentally, i.e. attempting to "motivate" their team. But is the same true of the armchair spectator shouting "Shoot!" at his television? That perhaps depends on the amount of alcohol he has consumed. Surely, though, many such vocalisations are simply *expressive* of what the person wants* (cf. Sect. 3.2). Certainly, the conceptual gap ought to

non-psychological facts, for instance, concerning people's living conditions. My claim here is that such reasons will only be able to do motivational work if they are framed as contents of motivational states.

[5] Aristotle (NE 1111b22-24) introduces the paradigm of spectator sport cases to distinguish between "boulēsis" and "prohairesis". It is, admittedly, difficult to see how the "wish" that a particular athlete win a race can qualify as the specifically "rational" attitude "boulēsis" is supposed to be (Sect. 1.3). More importantly, it is also unclear in what sense the desire that something happen over which one has no influence can itself be a form of "striving" or "orexis".

be clear: wanting* a team to win does not imply wanting* to contribute to their winning, although, of course, the first state can generate the second. Where this has not taken place, a spectator's wanting* some *p* cannot mobilise his motivational resources relative to *p*'s realisation.

Similarly, people are sometimes bearers of wants* with contents that explicitly exclude themselves contributing to the genesis of the state of affairs they represent. A parent may very much want* her son to keep a promise without him needing to be reminded to do so – where the last clause belongs to the representational content. Should the parent drop subtle hints or be tempted to do so, that would indicate that the want* had to compete with, and was perhaps overridden by, the want* that the promise be kept under any conditions. There is, however, no reason why this need be the case and there could surely be educators with no motivation to intervene in such cases.

There are thus two kinds of case in which wants* are not motivation mobilisers. The first can be termed *pure spectator cases*. The second group involves wants* whose contents contain an *intervention exclusion clause*.

Wants*, then, are mental states that represent contents in a specific mode. Where these states are not combined with beliefs in the logical or physical unrealisability of their contents, we often refer to them in everyday language by means of the verb "to want". Where such beliefs are given, i.e. in *OSI* cases, we normally use other lexical items to pick out the state.

3.2 Symptoms of Wanting*

Being motivated to φ, or to bring about *p*, is certainly a characteristic accompaniment or effect of wanting* to φ, or that *p*. But, as has been pointed out (Brandt and Kim 1963, 427; Audi 1973a, 39), wanting* also has a whole spectrum of further effects, some of which, because of their non-agential character, fall more readily under the broader concept of arousal. It is a plausible assumption that there is a *continuity* between the two types of effects and that at least the more prominent processes functionally identified as motivational in nature are a sub-set of arousal processes (Sect. 2.5.2).

3.2.1 Agential Symptoms

The most obvious motivational symptom of an agent's wanting* that *p* is her φ-ing where she takes φ-ing to *be* bringing about *p*. There seem to be examples where the form of "taking" at issue is implausibly conceived as a belief. If I feel like stretching and therefore do so, it would surely be incorrect to say that I do the particular stretch I end up performing because I believe that that is a way of stretching.[6]

Such cases, however, are undoubtedly limit cases. Most actions are involved in hierarchical structures of want* derivation, where an agent brings about some wanted* *q in order to* bring about some wanted* *p*, which she believes she can bring about if she brings about *q*. I shall call wants* generated in this way with a view to the in-order-to relation *subordinate* and the wants* in order to satisfy which subordinate wants* are generated *superordinate*. The distinction is more abstract than a number of related distinctions.

It deviates, first, from the talk of extrinsic and intrinsic motives. Whereas intrinsic motives are generally thought to have contents the desire for which requires no further justification, talk of superordinate wants* requires no such assumption. The distinction between subordinate and superordinate wanting* is local, leaving the question open as to whether the superordinate want* is not itself subordinate to some further want*. All subordinate wants* are extrinsic, although not all superordinate wants* are intrinsic.[7]

Second, although all instrumental wants* are subordinate wants*, again the converse is not true. An instrumental want* concerns a means to some end, where means are antecedent causal conditions deemed necessary, sufficient or conducive to satisfying some want*. It is the causal structure of means-ends relations that makes their representation by certain non-linguistic animals plausible. That structure also enables observers to infer from behaviour patterns that are flexibly sensitive to causal information that the behaviour is explained by such representations (Sect. 2.6). In contrast, what we think of as *ways* of bringing about goals can presumably not be adopted as want* contents by non-linguistic creatures. Ways of φ-ing are frequently specified by constitutive rules, according to which ψ-ing counts as φ-ing in some context: in order for a cyclist to indicate that he's going to turn right, he needs to stick out his right hand. Ways of φ-ing may also be determined by geographical or historical facts: the way of visiting the place where the first homo neanderthalensis was found is to go to the Neander Valley in Germany; it's also the way to go to the valley east of Düsseldorf through which the River Düssel flows. Agents can thus develop subordinate wants* as a result of conventional, historical or geographical as well as causal beliefs relevant to the realisation of superordinate wants*.[8]

[6] Note that this is not because stretching is a "basic action". Stretching is something I do by doing something else, namely moving both my arms in a particular way.

[7] This has the advantage of avoiding difficulties in providing a positive explication of the notion of intrinsic wanting*. If a child's wish to play is a paradigmatic intrinsic want*, is intrinsic wanting* really distinguishable from wanting* something for the sake of *pleasure*? Or if an agent's desire for global justice is an intrinsic want*, how are we to reconcile this with the idea that the agent is likely to justify this desire by referring to his intrinsic *valuing* of the want's* content? (cf. Roughley 2010).

[8] The locus classicus for these distinctions is Alvin Goldman's *A Theory of Human Action* (Goldman 1970, 20ff.). The claim that our closest primate relatives are unable to make sense of

These last distinctions are normally made in contexts in which the “by” locution is natural for everyday language: A cyclist signals that he is going to turn right by extending his right arm and a chimpanzee breaks a nut open by hitting it with a stone. “Coarse-grained” action individuators take the view that the “by” locution connects two descriptions of the same action, whereas for “fine-grained” conceptions the descriptions pick out two different actions. Nevertheless, even a theorist of the latter ilk insists that the two actions fulfil a “nonsubsequence requirement”, according to which the “two actions” must be performed during the same time interval (Goldman 1970, 21f.).[9] For my purposes, the important point is that the relationship of subordination between action-controlling wants* doesn’t require the satisfaction of any such nonsubsequence requirement by the actions thus motivated. This is particularly clear from the way we talk about adopting courses of action as means to bringing about some further state of affairs. Means-directed and ends-directed action descriptions can be connected by the “by” locution. However, an agent’s ψ-ing can also be a means to his φ-ing if his ψ-ing only brings about the conditions necessary for him to go on to φ. Boiling the kettle is a means to making tea, although no-one is likely to identify the one with the other. Independently of whether the word “means” would always seem appropriate in everyday language, the term “subordinate wanting*” also covers the motivation to bring about conditions preparatory to the realisation of some further want*. This is because preparatory conditions are brought about in order to bring about whatever it is we are thus preparing for.

The in-order-to relation can, finally, often be picked out by talk of “trying”: where someone ψs in the belief that her ψ-ing will, or might lead to, or count as her performing some action φ that she wants* to perform, it will frequently be appropriate to say that she is trying to φ. The appropriateness of talking of trying or attempting depends on a number of factors, some of which need have nothing to do with the action, its explanation or the psychology of the agent. These factors may be a matter of the agent’s subjective probabilities relative to her ψ-ing’s leading to her φ-ing, but can also be a matter of how difficult an ascriber expects ψ-ing to be for the agent or of the fact that, as the ascriber knows, the agent has actually failed to φ (Grice 1967/87, 7; O’Shaughnessy 1980II, 42f.; McCann 1975, 96; Hornsby 1980, 34f.). Talk of trying implies[10] that, from the perspective of the

constitutive rules is argued for by Rakoczy and Tomasello (for instance in Rakoczy and Tomasello 2007, 125ff.).

[9] This doesn’t mean that, according to the fine-grained theorist, one of the actions cannot finish before the other. The point is, rather, that the performance of neither can begin after that of the other. Even if the person shot by A doesn’t die until several days later, it would be wrong to say that A first shot her “and then” killed her.

[10] Following Grice, a number of authors took the perspective-relative doubt about the success of an action characterised as a trying to be a mere pragmatic implicature (Grice 1967/87, 43). However, I agree with Severin Schroeder that this conclusion presupposes an overly narrow conception of what can belong to the meaning of a term. Schroeder argues plausibly that perspective-relative doubt is a component of the meaning of “try” (Schroeder 2001, 219ff.). That this is a semantic presupposition

agent or the speaker, there appears to be something not so straightforward about an agent's doing something. In the cases at issue here, there appears to be something not so straightforward about an agent's doing something by, or as a result of doing something else.[11] The basic structure generally picked out is nevertheless that of action subordination. If apes are, as I have argued, indeed bearers of motivational states, then a chimpanzee in the Taï National Park pounding a nut placed on a rock with a carefully selected branch (Boesch and Boesch 1984) can be said to be trying to get at the nut.

Alongside cases of these kinds which, because of their rational structure, are the standard fare of philosophical action theory, there are other cases in which agential effects of wanting* come about through psychological mechanisms less amenable to rational reconstruction. In order to get a conceptual grip on these cases, I shall distinguish between *primary* and *secondary* wants*. Where an agent's wanting* to ψ is explained by her wanting to φ, her want* to ψ is a secondary want* and her want* to φ a primary want*. Like the distinction between super- and subordinate wants*, this is a local distinction: what is primary relative to one want* may be secondary relative to another. The distinction comprehends that between super- and subordinate wanting*: subordinate wants* are secondary wants*, but not all the explanations of secondary wants* work with the in-order-to relation. It is to those that are explained in other ways that I now turn.

Actions motivated by non-subordinate secondary wants* do not contribute to the realisation of the primary want*, except by chance. This is particularly clear in *OSI* cases, although the phenomena are not restricted to these. We can usefully distinguish three types, which I shall label *expressive*, *epistemic* and *imaginative*. For examples of these, take a football fan, who would dearly love her team to win their next match. She is not, let us imagine, in the slightest motivated to contribute to the satisfaction of her want*. Nevertheless, wanting* the team to win has three effects on her that we would describe as motivational. She has the tendency to talk incessantly about the approaching match to anyone who will listen. She attempts to acquire information about which members of the team will be playing and, after the match, tries to discover as soon as possible if her want* has been fulfilled. And sometimes she switches off from work and pictures vividly the way she would like them to outplay the opposing team.

Note that there are subordinate, usually instrumental variants of each of these. We talk to people about what we desire in order to receive advice or to persuade

and not a mere pragmatic implicature is shown by the incoherent character of Moore-paradoxical utterances such as "Tracy is trying to sit down, although there is no psychological or physical factor preventing her from sitting down and she doesn't believe that anything might prevent her from sitting down".

[11] Notoriously difficult cases of trying are those in which there are problems specifying *what* it is an agent does in order to φ, for instance, in trying to get up in the morning, to concentrate or, somewhat less usually, to move his arm when it is paralysed or anaesthetised (James 1890, 1101ff.; Hornsby 1980, 40ff.). There are grounds for doubting whether in such cases what an agent does counts as an action and a fortiori whether the structure of action subordination is realised.

them to help us achieve it. We seek information about things we want in order to be better placed to bring them about. And we sometimes imagine scenarios in which a goal could be realised in order to think through which method would be most conducive to achieving that result.

Instrumentally motivated imaginative actions are frequently subordinate epistemic actions. In these cases, an agent imagines *p* in order to discover features relevant for his prospective realisation of *p*. In contrast, non-subordinate imaginative action will generally be performed for its own sake or for the sake of the pleasure it is expected to bring. There are also hybrid cases, in which people fantasize about some wanted* *p* in order to gain the expected positive hedonic effects for instrumental reasons: someone might imagine how great it would feel to achieve some aim, because they believe that will help them muster more motivational force for the task of trying to achieve it.

The explicability of secondary actions in terms of desires for the pleasure of their realisation is less obvious for secondary expressive actions. It is presumably a fairly safe bet that our football fan does feel good when talking about the match. Nevertheless, it is no part of the phenomenology of such actions that she must have attended to that feeling and have chosen to behave discursively in order to experience it. Such deductively structured hedonic explanations are, moreover, particularly implausible for secondary epistemic actions, which may well be performed with equal frequency when the agent expects the primary want* to be frustrated.

Agents' wanting* is most clearly manifest for a third-person observer in their instrumental behaviour. It is, however, a commonplace that there are other characteristic patterns of nonrational, associative generation of secondary wants*, wants* that in turn lead to characteristic forms of behaviour. Such arational effects are among the symptoms of wanting* that tell against the Logical Connection Argument (Sect. 2.5.3). There are more of them, as we see when we turn to wanting's* non-agential symptoms.

3.2.2 Non-agential Symptoms

The genesis and causal efficacy of secondary wants indicates that the kind of physiological processes we tie together under the term "motivational force" can be triggered by a person wanting* *p*, in spite of not being motivated to bring about *p*. It is reasonable to assume that similar processes are at work in the production of non-agential symptoms of wanting*, some of which appear to be involuntary variants of secondary actions. They can be divided into two groups: firstly, effects of coming to believe that some *p* one wants* has, or has not, been realised and secondly, the symptoms at *t* of wanting* *p* at *t*. Whereas the latter are primarily effects on what the person *feels*, the former are effects not only on what he feels, but also on his *imagination*, *thought* and *perception*.

Among these latter effects are involuntary equivalents of both secondary imaginative and epistemic actions. Just as people sometimes deliberately fantasize about things they want*, they are also prey to involuntary daydreaming, dwelling on, or the repeated conscious tokening of beliefs about states of affairs they want*, or want* not to come about.

These are all ways in which thoughts about something a person wants* or to which she is averse*[12] can occupy their consciousness – more or less completely and more or less focally. Another kind of effect on conscious thought concerns involuntary attention, a phenomenon that shades into active attending.[13] Jenny for instance tends to "switch off" when her mother is expounding the details of her shopping expeditions, until she mentions having seen Jimmy, at which point Jenny is "all ears". Such attention-related phenomena offer particularly strong evidence for continuity between want*-dependent agential and non-agential processes.[14]

A third group of involuntary phenomena that belong here has been investigated in considerable detail in the psychology of perception. Unlike the non-agential symptoms mentioned thus far, neither the existence of the relevant perceptual effects nor their causal relation to the bearer's wants* are generally transparent to their bearers. The phenomena in question are examples of perceptual salience, a particularly high susceptibility of certain features of a subject's environment to being perceived by her. Empirical psychology tells us that perceptual salience can result from a number of different kinds of cause – from features of the perceived objects, such as brightness, from features of a particular perceptual context, such as a contrasting background or from features of the perceiver herself (Taylor and Fiske 1978; McArthur 1981).[15] Among the properties of persons that demonstrably have these effects are, firstly, the "activation" in the person of a semantic field by supraliminal or subliminal priming and, secondly, their "desires" or "values". For our purposes, a word about the second connection is in order.

What a person wants* has been shown to have a significant influence both on what he perceives and on how he perceives it. Everyday self-understanding and motivational psychology agree that caring about, wanting or worrying about

[12] I will sometimes use "to be averse* to p" and "to be averse* to φ-ing" as alternative formulations for "to want* $\neg p$" and "to want* not to φ".

[13] T.M. Scanlon distinguishes what he labels "the idea of desire in the directed attention sense" (Scanlon 1988, 39). This is odd, as what he is picking out is surely a symptom rather than a particular sense of "desire". Here, I agree with Dancy (2000, 88).

[14] Perhaps involuntary attending should be thought of as a form of subintentional mental action (cf. Sect. 5.1.3): although Jenny has not "switched on" intentionally, it seems that she nevertheless finds her attention suddenly focused because this facilitates something she wants* to do, viz. think about Jimmy.

[15] E.T. Higgins (1996, 156) has suggested that "salience" be only applied to entities easily perceived as a result of their objective properties. As there is no general acceptance of this suggestion within cognitive and social psychology and as it is useful to have a term for the perceptual phenomenon that is independent of its cause, I shall stick to the more general concept of salience specified in the text.

something makes the want's* bearer particularly susceptible to perceiving features of the environment related to what is wanted*, particularly if that feature represents an opportunity for its satisfaction. If you are hungry, you are when walking down the street more likely to notice a restaurant than under other circumstances (Brandt and Kim 1963, 435).

Studies carried out in the 1940s and 1950s by representatives of the "New Look" psychology of perception demonstrated correlations between motivational strength and various parameters of perceptual facility. Although, unsurprisingly, no precise one-to-one generalisations were forthcoming (Bruner and Postman 1948, 207), the studies convincingly showed that the strengths of what was fairly indiscriminately termed "desires", "needs" and "values" can be significantly correlated with the perceived size (Bruner and Goodman 1947, 49ff.; Bruner and Postman 1948, 296f.), the speed of recognition (Bruner and Postman 1947–1948, 75f.; Postman et al. 1948, 150ff.) and other parameters of perceived "vividness" (McClelland and Atkinson 1948, 218ff.) of objects.

In contrast to wanting's* effects on perceptual salience, the symptomatic character of much of our *hedonic* experience tends itself to be experientially salient. Wanting* something can, as Plato emphasised in the *Philebus* (35e-36c), be accompanied by anticipatory pleasure or by displeasure at the want's* present non-realisation. More salient still are the characteristic hedonic effects of the realisation of want* contents – both when their realisation has been striven for and when it was merely the object of hope or fear. We are also aware that there tends to some sort of correlation between the motivational strength of the relevant want* and the intensity of the hedonic experience that follows its satisfaction or frustration.

The English language provides us with a whole set of terms to designate the various compound states of wanting*, belief in the want's* (non-)realisation and resultant affect. These range from "frustration" and "disappointment" on the negative side to being "pleased", "contented", "satisfied", "glad", "happy" and "joyful" on the positive side. I take it that these terms designate overlapping and imprecisely distinguished affective-attitudinal compounds.

A note of caution here: it would be a mistake to see all uses of the terms "joy" and "happy" as entailing the presence of satisfied wants*.[16] It seems, on the contrary, that there may be experiences, particularly aesthetic in character, which trigger the relevant positive affect in their bearer in spite of his having had no prior want* with the relevant occurrence in its content. Think of experiences of nature or of music. Of course, in such cases, we tend to form wants* to repeat the experience, but that is clearly a distinguishable, additional step.[17]

[16] The analysis of "enjoyment" advanced by Davis (1982, 249) entails such a conceptual dependence of the relevant hedonic phenomena on satisfied desires. Tim Schroeder argues that this is true of all pleasure (Schroeder 2004, 88ff.).

[17] This has also been denied. Cf. Brandt (1979) 40f; Brandt and Kim (1963) 429, where "pleasure" is defined as whatever experience causes us to want its continuation. On conceptions of this kind, see Roughley (1999 and unpublished a).

3.3 Symptomatic Definition

3.3.1 Functionalism and Behaviourism

It has been suggested that adducing the symptoms of "wanting" provides a simple way of clarifying the concept. According to suggestions of this ilk, to "want" something is to be the bearer of *whatever kind of state it is* that generally has these effects (Brandt and Kim 1963, 427; Audi 1973a, 36ff.). This suggestion is part and parcel of a general doctrine in the philosophy of mind, according to which the essence of a mental state is its having characteristic effects and characteristic causes (Lewis 1966, 19ff.; 1972, 250; 1978, 124; Armstrong 1968, 82f.; 1977, 20f.), a doctrine that was appropriately baptised "conceptual functionalism" (Block 1994, 325[18]). Interestingly, the analyses of wanting* that have been advanced within this paradigm have tended to focus on what are seen as its characteristic effects, avoiding any reference to typical causes.[19]

The general doctrine is an attempt to maintain what are taken to be strengths of behaviourism, whilst discarding its weaknesses (Lewis 1966, 20ff.; 1972, 257; Armstrong 1968, 85ff., 129). The main inheritance of behaviourism at work here is the rejection of the claim that conscious experience could furnish conceptual criteria. Like the behaviourist, the conceptual functionalist insists that an understanding of human agents has to be in terms of third-person observable "inputs" and "outputs", in both cases an insistence grounded in the concern to satisfy standards of strict scientific methodology. Moreover, both approaches are equally concerned to develop an understanding of the mental that is compatible with the substance of a natural scientific world view, often labelled "materialism" or "physicalism". Behaviourists tended simply to expect some sort of convergence between the results of their – macroscopic or 'molar" – studies and those of the developing – microscopic or "molecular" – discipline of neurophysiology (Hull 1943a, 19).[20] In contrast, the functionalist's primary goal is to demonstrate the specific form that compatibility between these levels of description must be seen to have, assuming

[18] Block (cf. 1980b, 271) distinguishes the position thus labelled, and which he also calls "a priori functionalism", from the "empirical functionalism" or "psychofunctionalism" of Putnam and Fodor, for whom functional analyses represent empirical hypotheses. It is only with the former that I shall be concerned here. Brandt/Kim and Audi do not characterise their proposals as "functionalist", but refer to them as "theoretical construct" analyses (Brandt and Kim 1963, 427; Audi 1973a, 36; but cf. Kim's later remarks in his (1993), 191ff.). There is nevertheless a sense in which their holistic approaches are more strictly functionalist than Armstrong's single factor, neo-behaviourist analysis of "purpose" and "wishing".

[19] The fact that this may seem a natural way to approach "desires", whereas beliefs appear more easily definable by their causes, has itself been elevated within this perspective to the status of the defining difference between the two kinds of state. See the discussion of Smith's interpretation of the direction of fit metaphor in Section 4.5.1.

[20] Skinner's scepticism about the capacities of even future neurophysiological research to explain behaviour (1953, 27ff.) is untypical.

that the type of entity that occupies the causal roles it specifies will in humans turn out to be neurophysiological. Indeed, functionalism is first and foremost an attempt to solve the “mind-body problem”.

Conceptual functionalism sees itself as taking three important steps beyond behaviourism (Lewis 1966, 21ff.; Kim 1996, 77ff.). These concern the theoretical status of mental terms, that is, the ontological status of their referents, the criterial role of third-person observable input and output and the general conception of definition seen as appropriate for the mental sphere.

In the *first* place, functionalists take mental talk to be referring to *real* states internal to the entities to which those states are ascribed. This is contrasted with the view of behaviourists, according to whom everyday mental state terms refer to nothing more than behaviour patterns. According to Watson and Skinner for example, the ascription of an emotion such as anger is the claim that a person is either observably behaving in a certain way in response to certain conditions or else would manifest the relevant behaviour were those conditions to arise (Watson 1924, 145ff.; Skinner 1953, 162, 166).[21] Whether emotion talk is ascribing an occurrent or a dispositional property, it does not, according to this view, pick out any property internal to the relevant entity that could be *causally* responsible for observable “output”. In contrast, functionalism is resolutely realist, conceiving the referents of mental terms as real states of the bearer that contribute causally to what she does.

The *second* way in which functionalists see themselves as improving on behaviourism grounds in the recognition that any strict behaviourist attempt to provide constructive definitions of mental terms would be faced by an insuperable difficulty: that of defining mental terms with reference to purely bodily movements. This was recognised by proponents of the Logical Connection Argument (Sect. 2.5.3), who however, like their behaviourist cousins, drew the phenomenally implausible conclusion that motivational states therefore forfeit any claim to reality as states within the motivated person. Functionalists on the other hand, assuming that mental states are indeed internal to the person characterised by the corresponding predicate, drop the behaviourist premise that definition would have to proceed with exclusive reference to behaviour observable from a third-person perspective. Instead, consonant with the fact that most of the symptoms of wanting* – emotions, subordinate or secondary wants*, imaginings, perceptual salience – are themselves mental in character, they include *other mental states* among the criterially relevant behaviour.

A detailed discussion of the relationship between functionalism and behaviourism would make clear that neither of these two claims – of the reality of the attitudes and of the criterial significance of other mental states – were entirely

[21] Skinner and Watson differ as to whether they accept non-overt physiological reactions as part of the criteria for emotions. Watson, who does (1924, 165), is in this respect close to Logical Empiricists such as Carnap (1935, 89f.) and Hempel (1935, 17f.), who, interpreting psychological statements on the basis of the verification theory of meaning, saw both overt behaviour and experimentally observable physiological changes as constitutive of the referents of such statements.

foreign to behaviourism.[22] The theories of many second-generation behaviourist psychologists, for instance of Tolman and Hull, are built on the conviction that much observable behaviour of organisms can only be explained by postulating the interaction of various unobservable factors. These are given labels such as “perception”, “expectation”, “cognition”, “hypothesis”, “need” and “psychological force” (Hull 1943b, 278). Hull calls such unobservable entities “intervening variables” or “symbolic constructs”. They are postulated as entities that together entertain “quantitative functional relationships” to inputs and outputs (Hull 1943b, 278, 1943, 22). There is no question here of a simple one-to-one relationship between behavioural output and any such intervening variable.

Functionalism’s picture of behaviourism is closer to the truth where the latter is portrayed as non-realist about the ontological status of the “intervening” entities. There are programmatic statements by Tolman (1938, 9) and Hull (1943a, 22, 1943b, 281) to the effect that these “entities” are actually no more than the mathematical product of other functions, themselves either directly measurable or calculable on the basis of measurements. Where this position is consistently maintained, the employment of the relevant terms is no more than a matter of convenience or economy, enabling easily manageable reference to particular parts of equations linking input and output (Berlyne 1965, 13). However, this official operational or instrumentalist conception is incompatible with a number of developments within behaviourism. One central example is the idea that an organism’s acquisition of “knowledge” involves it becoming fitted out with “internal stimulus components” so that “a functional parallel” to elements of the external world becomes “a part of the organism” (Hull 1930, 512ff.; cf. Berlyne 1965, 96ff.). This conception of the mental as a kind of internal stimulus pattern only makes sense if conceived realistically and the “functional” relationships between the “intervening variables” are to be understood in causal terms.[23] Hull’s comparison with subatomic particles (1943a, 21; 1943b, 277; cf. Berlyne 1965, 13f.), an analogy later to be invoked by functionalists (Audi 1973a, 36), also supports this reading. Finally, Hull’s remarks that the behavioural mechanisms he describes ground in neural processes, which in the 1940s were inaccessible to observation (e.g. 1943a, 117f.), are strong indications that his intervening variables are by no means divorced from realistic assumptions.

Thus, although functionalism can legitimately distance itself from the official behaviourist doctrine of intervening variables, a number of “unofficial” moves within psychological behaviourism are much closer to functionalism than the official functionalist perspective would have us believe. Indeed, there is no clear

[22] This point is due to Gottfried Seebass. For a discussion of the motivation and cogency of so-called “mediation theory” in second-generation behaviourism, see his (1981/82).

[23] MacCorquodale and Meehl (1948) argue for a clear distinction in behaviour theory between those “intervening variables” that are to be understood instrumentally and “hypothetical constructs”, whose postulation entails genuine ontological commitments. Cf. Seebass (1981) 87ff.

dividing line between the two families of theories.[24] The most salient difference actually concerns the *genesis* of those internal factors thought to be relevant for behavioural output. Where the behaviourists conceived such factors as the effects of a history of exposure to certain kinds of stimulus (i.e. operant conditioning), functionalism is committed to no such genetic story. This reflects the differing aims of the two theories. Behaviourism is an attempt to explain observable animal and human behaviour; in contrast, functionalism assumes that there is already a more or less accurate explanatory theory of at least human behaviour contained in the "platitudes" of everyday psychology. Where the definitions of behaviourists are generated in the course of attempting to explain behaviour, functionalists use what they take to be a sufficiently accurate explanatory everyday "theory" in order to define the terms it employs.

This brings us to the *final* point in which functionalism sees itself as going beyond behaviourism. This is the abandonment of the idea that mental states might, even in principle, be definable in terms of necessary and sufficient conditions. This traditional idea is replaced by a looser mode of determining a concept, according to which none of the criterially relevant symptoms of the state in question are individually necessary, but a significant – not exactly specifiable – number of them are conjunctively sufficient (cf. Brandt and Kim 1963, 426f.; Lewis 1966, 22). As Lewis rightly says (1978, 122), it is always possible that certain typical symptoms of an attitude don't occur, or that, when they do occur, the relevant mental state is not their cause. He lists the perfect actor as illustrating the latter point, the total paralytic and a certain kind of "madman" as examples of the former. For this reason, he argues, we should understand the concept of a mental state as that of a state with *typical* or *characteristic* causes and effects. This can be rendered formally by the stipulation that a state's causal role be specified by a disjunction of conjunctions of *most* of the "everyday platitudes" concerning that particular state (1966, 22, 1972, 256, 1978, 127). This solves the – behaviouristically insoluble – problem of accounting for unavoidable counterexamples. These are classed as untypical cases and are taken to be covered by one or other of the less prominent disjoined conjuncts.

The upshot is that functionalism conceives mental vocabulary as terms in a somewhat vague, more or less explicit commonsense theory, which are nevertheless analogous to the theoretical terms of subatomic physics in as far as these are thought to pick out unobservable, but nevertheless existent entities. According to this perspective, everyday psychological talk is the *application of a theory* of the way certain events in us cause others, which, at the end of longer or shorter causal chains, issue in overt behaviour. Psychological terms such as "want", "concern", "care", "longing" mark nodes in causal networks, individuated by what we "folk" take to be their nomological properties (Lewis 1972, 256f.; Brandt and Kim 1963,

[24] Cf. Lewis (1978, 124), where the Lewis-Armstrong philosophy of mind is straightforwardly characterised as "behaviorist *or* functionalist" (my emphasis).

427f.; Audi 1973a, 36f.). According to Lewis, the terms for mental states function as if they had been invented in order to explain input-output regularities among agents who had until then got along without psychological vocabulary.[25]

3.3.2 Two General Objections

The idea, then, is that the symptoms of wanting*, at least in as far as they are grasped by everyday language users – presumably some of those adduced in Section 3.2 will not count (Brandt and Kim 1963, 428f.) – define the types of entity we postulate in order to explain these effects. Before I come in a moment to look at the specific claims made by David Armstrong about the practical attitudes, I want to voice two objections to the general programme.

The first point is that functionalism's move from behavioural to mental criteria sacrifices the *operationalisability* that behaviourism at least held out hopes for. Behaviourism rejected the results of introspection as intersubjectively unverifiable, insisting instead that third-person observable events – bodily movements – be seen as the only valid criteria for the ascription of mental terms. The replacement criteria offered by conceptual functionalism are no longer third-person observable. Thus, when attempting to identify an individual attitude, it becomes difficult to see where the greater precision vis-à-vis introspection is supposed to lie. It also ought to be obvious that the identification of the other attitudes on which "desiring" is thought to depend is itself only possible where introspection is employed. If we really do exclude this possibility, then it is unclear how we ever come to know that we "desire" anything. For the functionalist, the formation of secondary and subordinate wants* and the occurrence of relevantly salient perceptions are themselves no more than events which generate further nodes in nomological networks, which are in turn only identifiable by their effects, for instance on other of our "desires".

Functionalists have argued that the circularity at work here only serves to secure the claim that mental states only come in packages (Brandt and Kim 1963, 429; Lewis 1966, 21; Armstrong 1977, 24), a feature emphasized by Nick Zangwill's term "network functionalism" (Zangwill 1998, 185ff.). However, this claim sits uncomfortably with the idea that there might be creatures fitted out with smaller arrays of attitudes. Human wants* are closely bound up with our emotional life. Would we have a knock-down argument against the claim that invertebrates have beliefs and wants* if it could be shown that the creatures in question, for instance bees, don't have any emotions? Is a figure such as the Star Trek android Data an incoherent thought experiment if he is supposed to have wants* and beliefs, but no emotions? The high functioning autist, Temple Grandin, a fan of Data, has no access to the complex social emotions with which normal human wants* are interwoven (Sachs 1995, 262, 248). It would, however, be absurd to refuse to ascribe

[25] This is the myth suggested by Sellars (1956, 178).

her the wants* she obviously has just because she doesn't have the whole package of attitudes everyday psychology assumes are possessed by persons (cf. Carruthers 2009, 101f.).

A requirement that wants* be necessarily incorporated within the network of causal connections normal for humans is thus inacceptable. A weaker version of the holism of the mental with a stronger degree of plausibility is Davidson's claim that beliefs and "desires" form a closed package or "matched set" (Davidson 1981, 96). In Section 2.6, I argued that this is, strictly speaking, incorrect (cf. Sect. 4.5.3). More importantly for our context, there is no hope that such a thin set of attitudes could, together with the behaviour of their bearers, furnish criteria that are sufficiently differentiated to enable accurate identification of states of "desire".

My second point concerns the assumption that the language of everyday psychology is best understood as an *explanatory-predictive theory*. In spite of its popularity, the idea is anything but self-evident. Of course, as Hume pointed out (Sect. 2.3.1), it is important that we can make successful assumptions about what people are going to do on the basis of what we know about their motives. But that doesn't make prediction and explanation of behaviour the only significant function of our psychological idioms.

The language of our concerns, wants, aims and emotions is certainly not exclusively, and probably not primarily, either predictive or theoretical. On such a view, two highly significant uses of the relevant linguistic forms are marginalised and become difficult to make sense of: firstly their use to *express* our *first-person relations* to the world and, secondly, to facilitate and represent forms of *practical deliberation*.[26] We don't normally think about our concerns, clarify and mull over what is bothering us and relate these things to our overarching aims and our reasons for having them in order to discover what we are going to do[27] or to explain why we have done what we have done. Reflective episodes in which we take on such a distanced theoretical perspective toward ourselves are the exception and require

[26] Richard Moran believes, I think correctly, that there is an intimate connection between an agent's first-person perspective and the perspective she takes on in deliberation. However, he conceives that relationship as *too* intimate, seeing the first-person perspective as constituted by the normative reasons an agent has for deliberatively forming attitudes with particular contents (Moran 2001, 65; 113ff.). We can express our wants* outside deliberative contexts. Moreover, we need not take those wants* to be backed by sufficient reasons for them to be expressive of our practical point of view. I return to this topic in Section 8.5.

[27] David Velleman denies this. According to Velleman (1989, 90–101, 1996, 188ff.), practical reasoning turns out to be a peculiar, self-referential form of theoretical reasoning. Wondering what to do, the question that initiates a practical reasoning episode, is, he claims, a matter of wondering what one is going to do as a result of the belief that one will acquire in answering the question. Any plausibility that Velleman's view may be thought to have results from his distinguishing this self-referential, predictive kind of thought from the explanatory enterprise normally seen as constitutive of the theoretical. Velleman's "theoretical" take on practical reasoning thus offers no support to the claim that the attitudes are entities postulated in the service of a fledgling explanatory science of human behaviour. I return to Velleman's doxastic conception of intention in Section 6.3.1 and to his particular take on practical reasoning in Section 8.5.1.

specific explanations. What we are generally doing in such episodes of reflection in terms of our wants* is wondering *what to do*. As the folk know, this is not at all the same as wondering what you are going to do.

For this reason, there is a central use of the language of wants, desires, concerns and intentions that is first and foremost *practical*. The essentially first-person forms of reference to the relevant psychological states *articulate* movements of the minds of their bearers in ways that are, at least by and large, transparent to them. It is a striking fact that we generally know what we want* and, moreover, do so without having to observe ourselves to see whether we are playing host to any of wanting's* characteristic effects (Goldman 1970, 95f.; Moran 2001, xxix; Finkelstein 2003, 1). For this reason, we don't have to postulate the relevant psychological entities when we are reasoning practically in the first person. Where we do develop beliefs about our wants*, we tend to do so on the basis of the fact that taking the relevant kind of attitudinal stand is frequently a conscious movement of the mind, on which we only need to focus doxastically.

3.3.3 "Purposes"

If we turn now to David Armstrong's explicit application of conceptual functionalism to the question of practical mind, we are presented with illustrations of the way the doctrine obscures both the first-personal dimension and the practical features of wanting*. Armstrong attempts to provide an explanation of the transparency of mental states to their bearers within a functionalist framework. In a curious deviation from the holism which would seem to require awareness of a whole set of causal relationships (cf. Brandt and Kim 1963, 427; Audi 1973a, 39), he claims that we have a direct form of introspective access to our "purposes". This involves being aware of "the presence of factors that drive in a certain direction" (1977, 25) or of something in us apt for bringing about a certain effect (1968, 135).[28]

Armstrong's departure from the holistic premise of functionalist analysis is understandable. Compare Brandt and Kim (1963, 429), who remain faithful to their programmatic methodological holism here, offering as a paradigmatic case of knowing what you want the inference from noticing your joy on hearing a piece of news. Their – perfectly appropriate – concern to argue that we are *sometimes* unaware of what we want (cf. Sect. 5.1.4) leads them to present a phenomenally implausible picture of self-knowledge as *entirely* derived from accurate self-observation.

In contrast, Armstrong, by sacrificing the holistic premise, is able to offer a prima facie more plausible account of our immediate access to our wants*. We

[28] What someone with a purpose is aware of according to Armstrong has notable similarities with what Hull calls "the purpose mechanism", namely "a persisting core of sameness in the [internal] stimulus complexes throughout successive phases of the reaction sequence" (1930, 519).

certainly don't generally have to wait and see how we react under certain conditions to find out what we want*. Nevertheless, the feature of which Armstrong's agent is aware has nothing specifically first-personal about it. Someone possessed of an Armstrongian "purpose" observes goings on in herself in the same way she might try to work out whether certain of her sensations indicate the imminent triggering of reverse peristalsis. The psyche appears as a machine observed by "the mind's eye".[29] But conscious wanting* is not an *epistemic relation* to one's wants*.[30] Rather, in consciously wanting* something, a person represents the relevant content in a non-epistemic mode in a way that is unmediatedly present to its bearer.

Now, there are drawn-out arguments as to the plausibility of the higher-order perception ("inner sense") theory of consciousness that Armstrong advocates. These go way beyond our concerns here. However, one problem that results from the conjunction of his concept of purpose with the higher-order perception theory should be mentioned. In contrast to the higher-order thought theory (Rosenthal 1993, 204ff.), the inner sense conception has no problems attributing consciousness to non-linguistic animals, as perception plausibly works non-conceptually. However, it is unable to provide a coherent construal of what it is to *perceive* some goings-on in ourselves as "driving in a certain direction". This may seem plausible if you are hungry and the food is right before your eyes and there is a literal sense of "direction" available. However, for most other purposes, talk of "direction" is a metaphor for the attitude's intentional content, as is Hobbes' talk of "motion towards" an appetite's object (Sect. 1.4). Once one replaces the idea of direction with that of intentional content, there is no way that non-conceptual perception could make sense of what it is perceiving.

Note that these criticisms of Armstrong's phenomenology of "purpose" have nothing to do with functionalism's inability to account for qualia. Indeed, the "awareness" of features in us "driving" somewhere seems to require qualitative experience. Functionalists have been perfectly right to insist that wanting* need not involve any *qualitative* experience (Brandt and Kim 1963, 425f.; Smith 1994, 108, 112f.) – although many states referred to by terms such as "desire", or the ancient Greek "epithumia" (Sect. 1.2-3), certainly do.

A final objection to Armstrong's proposal is that, alongside the misconstrual of the first-person perspective on one's wants*, it renders incomprehensible the fact that an agent's wanting* some content at least sometimes provides its bearer with a reason to bring about the corresponding state of affairs. The perception of

[29] Compare Skinner (1953, 262): "'I was on the point of going home' may be regarded as the equivalent of 'I observed events in myself which characteristically precede or accompany my going home'". In spite of Skinner's insistence on the causal inefficacy of the relevant events, the epistemic relation he sees an agent as entertaining to those events is precisely that assumed by Armstrong.

[30] I thus second Finkelstein's Wittgensteinian critique of "detectionism" (Finkelstein 2003, 9ff.). Bizarrely, Moran's insistence that the first-person perspective is constituted by the deliberative weighing of reasons leaves him with a phenomenology of reason-insensitive wants* that is very close to that of the behaviourists: "A brute desire", he thinks, " is a bit of reality for the agent to accommodate, like a sensation, or a broken leg, or an obstacle in one's path" (Moran 2001, 115).

something as pushing us in a certain direction is hardly a candidate for this role. At most, it is going to give us grounds for predicting our behaviour or for seeking ways to resist. As theories that see reasons as grounding in the agent's wanting* are frequently characterised as "Humean" (Smith 1987; Schroeder 2007), it is worth remembering that an understanding of the desiderative generation of reasons that worked with Hume's own conception of "desire" would be considerably more plausible. If desires are essentially hedonic states, then we have an answer to the question as to what is so "pro-" about this sort of attitude. If, on the other hand, they are merely "hydraulic" forces,[31] it is understandable how they might explain behaviour. However, the play of such forces provides no anchor for even the weak kinds of normative claims we make when we say that wanting* some *X* makes it *intelligible* why the want's* bearer acted as she did: they would give us no clue as to "what was to be said for" the action from the agent's point of view.[32] Hydraulically understood "desires" could at most furnish explanatory, but certainly not motivating reasons.

*3.3.4 "Wishes" and OSI Wants**

One obvious objection to any theory that analyses "desire" neobehaviouristically in terms of – felt or unfelt – "drives" to bring about the desire's object is provided by *OSI* cases (Sect. 3.1.2). Armstrong is well aware of this and offers a proposal of how to understand them. "To want" or "to wish" in *OSI* cases, he suggests, is to be in a state of which a certain kind of counterfactual claim is true. This states that, if the bearer were to believe in the realisability of the attitude's content by her action, then she would act to realise that content, or attempt to do so, or at least tend to attempt to do so (1968, 155[33]). If wants* are dispositions, the occurrence of the behaviour that defines them is dependent on the instantiation of the relevant triggering conditions. "Mere wishes" would then simply be wants* whose triggering conditions appear impossible.

Consider the two kinds of *OSI* cases I discussed in Section 3.1.2, namely pure spectator cases and wants* with contents containing intervention exclusion clauses. In the first kind of case, we can want* things to happen whilst not wanting* to

[31] For critical uses of the metaphor of hydraulic forces, see McDowell (1981, 212f.), Gilbert (1989, 419f.), Dancy (1993, 13, 2000, 11) and Wallace (1999, 630ff.).

[32] This is Davidson's notion of "rationalization" (Davidson 1963, 8f.), a relationship between a person's motives and her actions that, although it doesn't guarantee justification, nevertheless involves more than mere causation. See on this point Mele (2003a, 75f.; 2007, 113). The argument that "desires" understood as mere functional states could not "rationalise" actions is advanced by Warren Quinn (1993, 246). What Quinn takes to be missing in such a conception is an axiological dimension (see below Sect. 4.3).

[33] Compare the similar suggestion made more recently by J. Dancy (2000, 87f.) in a different context.

intervene. In the second kind of case, the non-intervention of the attitude's bearer is an internal condition on the want's* content being realised. In both sorts of case, the wants'* contents exclude not only actual, but also counterfactual motivation. Think back to the mother who wants her son to keep his promise without intervention on her part. Her wanting this could certainly be related to her longing to go back in time and change the way she brought him up.[34] Nevertheless, this latter want* has a different content to the want* with the intervention exclusion clause. For one thing, it depends on her belief that her son is in fact motivated not to keep his promise. More importantly, it depends on a particular causal conception of how people come to be promise keepers or promise breakers, a conception that may well be foreign to the mother. Should she have beliefs of these two sorts, then that may well lead to her developing wishes concerning her earlier parenting behaviour. But such a development would be a further attitudinal step. It seems, then, that she can – knowingly – play host to such a want* in spite of not believing anything that would trigger behaviour that the functionalist deems criterially relevant.

If the arguments from *OSI* cases work, then Armstrong's dispositional analysis of mere wishing fails. Perhaps, however, it works at least for all action wants* that would not be classified as purposes. Armstrong's own example of a wish concerns an action want* the opportunity for whose realisation is past, namely wishing one had attended the first performance of *Twelfth Night*. Note, though, even if it were true of Toby now that, were he to be teletransported back to England on February 2nd 1602, he would attempt to see the play,[35] that would normally not be thought a sufficient condition for ascribing him the wish now. Someone disposed to act in a certain way under specific conditions, but who has as yet not considered the matter, does not at the moment want* that particular content. She is, rather, merely disposed to acquire an attitude she might never end up having. If *X* is disposed to fall in love with *Y*, should they meet, he is hardly already in love with her. A precondition for ascribing an agent the want* that *p* is that the agent be playing or have played host to some representation of *p* – in the normal human case a conceptually structured representation. A mere dispositional analysis ignores this essential feature.

This point is related to Armstrong's inability to reconstruct appropriately the first-person perspective of the bearer of a wish. Unlike the agent with a "purpose", he would not necessarily require introspection in order to know what he wants*. What he would need to know is the truth of relevant counterfactuals. Here again, then, the immediate knowledge we generally, although not infallibly, have of what we want* becomes inconceivable. Those dispositional features of agents we pull together under the heading "motivational force" may be only accessible to their bearers, if at all, through counterfactual thought processes (or self-observation). This is not true of wants* themselves.

[34] This was suggested by an anonymous reviewer.

[35] Assuming that the correct theory of personal identity allows us to make sense of this counterfactual.

In fact, it seems that even action wants* need not at the time of being tokened necessarily be backed by a level of motivational force which under other doxastic conditions would lead the agent to act so as to realise their content. Return to the example of Corinne (Sect. 2.5.1), who is so exhausted that she cannot muster the motivation to do something she genuinely wants* to do. Corinne might well also believe that she is motivationally unable to realise her want's* content. Nevertheless, what explains her not making any such attempt will normally not be her belief. It is at least conceivable that her inaction is explained by the fact that the content of her belief is true, i.e. by the fact that she is insufficiently motivated. Indeed, the explanation in terms of her insufficient motivation may even be true if she has not developed any such belief at all.

Finally, is it not conceivable that someone might form a spontaneous *whim* that something be the case, but simply not think it worth the bother to embark on realising it – although he believes he easily could? Now, people sometimes say that when someone only has a "whim" to do something, she doesn't "really want" to do it. What they tend to mean by that is something that may appear to provide support for Armstrong's counterfactual proposal. At least on one use of the expression, we deny that someone "really wants" something if they are not prepared to adopt the means they take to be necessary for its attainment. In other words, the use of the adverb "really" imposes the requirement that the want* denoted be backed by a contextually sufficient level of motivational force. As "really" is generally an "emphasiser", that is, an adverb that emphasises the truth of what is asserted (Quirk et al. 1985, 583), it may look as if there is enshrined in the English language the implicit claim that motivation is the essence of wanting.

Two further linguistic facts relativize any such impression. Firstly, "really" is one of the emphasisers which, when used with a gradable verb, can also have a scaling or "boosting" effect (Quirk et al. 1985, 586). That is, it can be used simply to denote a high level of *strength on some scale*. The fact that the adverb can be used to place an instance of wanting* on a motivational scale doesn't of course say anything about whether or not the scale begins at zero. Secondly, there are other uses of "to really want" which locate the attitude high up on *different* scales, in particular that of *hedonic* strength. For instance, we also sometimes deny that someone "really wanted" some proposition if they experience no pleasure on its instantiation, in spite of perhaps having been sufficiently motivated to bring it about.

It is a significant fact that there are *two* such prominent notions of "really wanting", each of which may turn out to be applicable where the other is not.[36] This fact might suggest that we actually have two completely different concepts that ordinary language keeps insufficiently separate: the most prominent symptoms of *OSI* cases are hedonic, whereas action wants* typically have observable motivational effects. The varying salience of these two sorts of phenomena have led to two families of competing theories of wanting. Hobbes' and Hume's conceptions can be seen as representatives of these two families.

[36] It turns out, as I shall argue in Section 5.2, that there are actually four such notions.

In contrast, I have been arguing that we are indeed dealing with a generic mental state, a sort of state that readily enters into compounds both with motivational and hedonic or affective phenomena. On the one hand, the prominence and independent variability of these two features at least suggests that neither can be exclusively responsible for the phenomena at issue. On the other hand, the way they tend to cluster around what appear to be identical contents suggests their equal dependence on an independent feature of our mentality. We can feel pleasure at the realisation of some state of affairs independently of our action. But such realisation-dependent pleasure can equally come about in cases in which that very state of affairs was aimed at by our motivated behaviour. It is surely no coincidence that frequently those states of affairs whose realisation causes us pleasure are those states of affairs whose realisation we were motivated to bring about. Were the dispositions to feel pleasure at some *p* and the motivation to bring *p* about unconnected dispositions, then we would have such a coincidence.

3.4 A Theory of Wanting*: Key Questions and Sketch of Some Answers

The discussion of functionalist conceptions of "desire" brings into focus a key feature of wanting* that such a conception is unable to account for. Wants* are in general immediately accessible to their bearers, where the relevant form of accessibility is not primarily epistemic, grounding agents' beliefs about their own wants*. Rather, wanting* is primarily the optative framing of certain contents, an essentially first-personal state of mind. You can form higher-order beliefs that you are in that state of mind yourself, as you can about the wants* of others. When you do so, what you are forming beliefs about is a specific first-person perspective on the contents in question. Taking on that perspective tends to have motivational and frequently hedonic consequences or accompaniments. There are, however, as I have argued, contexts in which no such consequences are necessary.

What functionalism correctly points out is the importance of the characteristic syndrome of wanting's* symptoms. Although these are not definitive of what it is to want*, there are, however, plausibly important cases in which our access to what we want* is mediated by observation of such effects in ourselves.

A theory of wanting* should be able to tell us why we *generally and reliably* have direct access to our attitudes, *without necessarily* doing so. Phenomena of perceptual salience may alert a person to the fact that she has wanted* something for some time without having been aware of doing so – although she need not become aware of the way her perceptions and thoughts are being influenced by what she wants*. The theory ought to tell us how this can be the case, whilst at the same time explaining the priority of those cases in which the relevant mental states are directly accessed. A theory that fails to do this misses the essential structure of the phenomena it aims to reconstruct. Correlatively, this peculiar structure makes clear demands on the theory.

Whatever wanting* is, its explication should enable us to explain why it characteristically leads to the symptoms listed under Section 3.2. The resulting theory should explain why someone who wants* something typically knows that she does. And it should also explain why uncharacteristic cases can also occur, in which a want's* bearer is unaware of wanting* what she wants*. Characteristic accessibility, but potential non-access is plausibly a general feature of all the attitudes, whilst the explanatory relation to a particular set of characteristic symptoms is specific to wanting*.

In the next two chapters of this study, I shall suggest that these requirements can be met by a position along the following lines: firstly, attitudinising is being the bearer of a representational structure the only adequate rendering of which is in terms of its *expressive articulation* in a public medium. Such articulation of an attitude, to be distinguished from its description, necessarily involves an irreducible first-person framing of the proposition that is the attitude's content. The fact that there is no such thing as an attitude without a bearer is not merely external to the structure of the attitudes, although the exclusive focus on belief may easily conceal this.

Secondly, attitudinising typically, but by no means necessarily involves having a *conscious* thought – in linguistic, or proto-linguistic form. Where no such conscious phenomenology is present, we may have good reason to assume that the same kinds of representational processes are at work below the level of consciousness. Contents can drift in and out of awareness and can, at least up to a point, produce an attitude's characteristic symptoms independently of whether that awareness is given.

Thirdly, the appropriate expression of wanting* is by means of sentences of the form "Let it be the case that *p*". In such utterances, the propositional content is represented in what can be labelled the *optative* mode, which is to be contrasted with the *assertoric* mode, expressed by "It is the case that *p*". The essence of the attitudinal framing constitutive of the two basic attitudes is the "taking of a stand" by the attitude's bearer on the question of the proposition's realisation. This standpoint-relativity is made explicit in the expression of the optative mode. The mental posture thus expressed, whether consciously or unconsciously taken on or maintained, is an excellent candidate for a mobiliser of those fuel resources that power primary or secondary action and which plausibly also give rise to the involuntary perceptual, emotional and imaginative effects of wanting*.[37] From here on I shall dub the collected characteristic symptoms of wanting *the optative syndrome*.

Finally, the optative attitudes that we label with the terms of everyday psychology consist of the conjunction of the optative stand with identifiable constellations of further factors. These are at least three in number: firstly, an assertoric attitude with a more or less specified content – concerning such factors as the realisability of the want's* content or further features of its relation to the attitude's bearer. The second

[37] I thus agree with Nick Zangwill's claim that "the essence of a mental state" – at least in the basic cases of wanting* and believing – "explains its causal powers; it is not constituted by them" (Zangwill 1998, 179).

component is motivational force of some more or less clearly specifiable strength. The third is the felt dimension of hedonic experience. Everyday uses of the term “desire” tend to pick out an optative attitude conjoined with all three additional factors.

The rest of Part I of this study will be taken up with arguing for the claims made in the last four paragraphs. I will do this in two broad steps. In the next chapter, I argue for an expressive explication of the attitudes, in particular for the optative conception of wanting*. Chapter 5 then deals with the relationship of optative attitudinising to hedonic experience and to consciousness.

Chapter 4
Expressive Explication and the Optative Mode

In this chapter I shall defend the claim that the essence of wanting* is being the bearer of a representation in the generic optative mode expressed by utterances of the form "Let it be the case that *p*". I will do so in five steps. The first (Sect. 4.1) explains the idea of an expressive explication of the attitudes. The idea can be helpfully approached by getting clear on the phenomenon of "Moore-paradoxicality". An understanding of what is going on in Moorean sentences reveals beliefs to be essentially assertoric attitudes, expressed by simply tokening a linguistic analogon of the proposition believed. In a second step (Sect. 4.2), I discuss the possibility that we can construct parallel sentences for wanting*. Although it turns out that no strictly parallel sentences can be built in everyday English, because of the lack of a precise term for the ascription of "pure" optative attitudes, Moore-paradoxical sentences for wants* are certainly constructible. The possibility of doing so, along with the specific moves necessary to this end, confirms the analysis of wants* as essentially representations of contents in the optative mode, characteristically conjoined with additional motivational, affective or doxastic components.

The claim that motivational states are at core optative is controversial. In the third and fourth sections of the chapter, I consider two competing construals. According to the first (Sect. 4.3), "wants" or "desires" are to be understood as attitudes which, in strict parallel to the relation of beliefs to truth, "track the good". According to the second construal (Sect. 4.4), it is, in at least some, possibly in all cases, a mistake to understand "wants" as playing any substantive role in the production of our motivated behaviour. Instead, so the claim goes, to say that someone behaved in some way because she wanted to is simply to state that her behaviour was motivated or intentional. This statement is taken to be compatible with the action being explained entirely by the agent's belief that she was required to act in that way. In the course of rejecting both claims, I attempt to clarify further what is meant by, and what follows from the optative analysis.

Finally, in an appendix to the chapter, I relate the optative proposal to the notion of "direction of fit". The latter is best understood as a metaphor for the

N. Roughley, *Wanting and Intending*, Philosophical Studies Series 123,
DOI 10.1007/978-94-017-7387-4_4

perhaps puzzling conclusion of the expressive explication: that wanting* is at core a "subjectively normative" matter. I show that two prominent attempts to avoid this conclusion fail and argue that the less demanding character of the subjective standards set in wanting* may explain the possibility of wants* without beliefs, a possibility plausibly realised in both young children and some primates.

4.1 Moore's Paradox and the Idea of Expressive Explication

4.1.1 Expressive Articulation and Assertoric Attitudinising

I begin by arguing that an understanding of the peculiar kind of incompatibility of the two conjuncts of "Moorean sentences" gives us a grip on the core structure of belief. There is, as I shall claim, such a thing as Moore-paradoxicality because language enables the production of publicly accessible analogues of our attitudes. Where we do this, our utterances express, rather than ascribe corresponding mental states.

Moorean "paradoxicality" is given where someone utters a sentence of either the form "*p*, but I don't believe *p*" or "*p*, but I believe that non-*p*". As has often been remarked (Williams 1970, 137), both sentences are logically strange, in spite of not being self-contradictory. The relation between the conjuncts in these sentences is peculiar in two ways that require explanation: firstly, we are dealing with a form of incompatibility that is somehow less than contradiction. Secondly, the relevant form of incompatibility is only given under certain grammatical restrictions on both the tense and the person of the second conjunct. There is no problem with these sentences if the second conjunct is put into either the past or the future. There is also no problem if its verb is not in the first, but in the third or second person.

These grammatical restrictions are indicative of the fact that the problem is somehow grounded in the standpoint of the utterer at the time of utterance. Obviously there is no incompatibility between *p* being the case at *t* and someone not believing *p* at *t*. In what follows, I will concentrate first on sentences of the form "*p*, but I don't believe that *p*", for reasons that will become apparent. In doing so, I shall – contrary to the practice of most everyday English speakers – read this sentence literally, that is, ignore the phenomenon of what linguists call "transferred negation".[1] I will refer to these as *type 1 Moorean sentences*.

Understanding the source of the non-contradictory logical oddity of such sentences involves appreciating two points that will be of use to us in the analysis of

[1] Transferred negation involves a transfer of the negative particle from a subordinate clause, to which the speaker means it to apply, to the matrix clause in which it is nested (Quirk et al. 1985, 1033). In everyday English, for instance, "I don't think that's a good idea" means "I think (that is not a good idea)".

wanting*. The first is the distinction between the assertion and the expression of an attitude. The second is the essentially assertoric nature of belief.

To take the first point first: it is a striking fact that we do not normally communicate our beliefs by means of sentences that employ the verb "to believe" or any of its cognates. Rather, utterances of the form "I believe that *p*" are generally used by a speaker to "distance himself" from the content of his belief. Before giving some substance to this metaphor, let me illustrate it with two typical uses of the sentence. In the first, a person adverts explicitly to the doxastic framing of a proposition when she wants to indicate her uncertainty as to its truth: paradoxically "I believe that *p*" generally means "I'm not sure that *p*" (cf. Hare 1952, 6). A second use of "I believe that *p*" is to ascribe oneself a dispositional doxastic state, making what Williams (1970, 138) calls "an autobiographical remark". Imagine someone writing a brief self-characterisation that includes descriptions of her physical characteristics, hobbies and beliefs. Such a characterisation could just as easily be made from a third-person perspective, simply replacing "I" with "she".

In both these cases, the belief's bearer, in making an assertion *about* himself, takes on an "external" perspective towards the bearer of the relevant attitude. In as far as the speaker is indeed making a genuine assertion about himself, he can be mistaken: the claim to truth and the possibility of falsehood are essential to an utterance being an assertion. Someone might have temporarily forgotten the details of some belief of theirs and thus be led to ascribe themselves a belief that they don't have. Note that, although consciously believing that one dispositionally believes *p* will often lead a person to acquire the belief that *p*,[2] this need not (pace Baumann 2000, 97f.) be the case. There is a mental step involved in moving from one to the other. And the believer may have reasons for not taking that step.

The difference is evident from the fact that the reasons someone might have for believing occurrently that they believe *p* dispositionally are different in kind to the reasons she might have for believing *p* occurrently. Whereas reasons of the second kind concern the way the world is, reasons of the first kind concern the way the

[2] For a case of a mistaken conscious belief about a dispositional belief, consider Ada. Ada has seen an advert for Cleen and come to believe that Cleen is the only washing powder that will remove the stains that have proved resistant to all those powders she has so far tried. However, on standing before a shelf stocked with various washing powders, she realises that she is unsure which powder she believes to be the best stain-remover. She knows she has a belief about some washing powder, but which one is it? Standing before a shelf stacked with Bright, Brite, Bryte and Kleen, she forms the belief that Kleen is the powder which she believes to be the best stain-remover. Moreover, this second-order belief naturally leads her to form the first-order belief that Kleen possesses the best stain-removing properties. As a result of this latter belief, she buys a packet of Kleen, but on the way home suddenly realises that she was mistaken, having confused two similar representations. She had forgotten that her belief was predicated of Cleen and not Kleen. Note that this does not mean that she had ceased to have the belief about Cleen. She in fact remained disposed to have the relevant thought about Cleen – which is the very fact about her that explains how she came to buy Kleen, rather than, say, Bryte. I have profited from discussions with Peter Baumann on these matters.

attitudiniser is. She might, for instance, conclude from certain recurrent forms of her own behaviour that she dispositionally believes some proposition, although she doesn't at that moment consciously accept it. The logical gap is most obvious in psychoanalytic cases. Whatever may be the truth of the various explanatory theories on offer in the field of psychotherapy, the plausibility of *some* such explanatory claim rests on the existence of the logical gap opened up by the possibility of "distanced" or "external" assertions about one's own doxastic states. Unconscious beliefs are beliefs a person can *only* ascribe to herself on the basis of behavioural evidence in exactly the same way in which others might ascribe the belief. It is in this respect that unconscious beliefs are missing a feature essential to standard cases of believing.

That feature is manifested in the fact that a conscious believer will generally be prepared to offer reasons for the belief's content, whereas a person ascribing himself an unconscious belief will only be prepared to justify the belief's self-ascription (cf. Collins 1969, 671). I can ascribe to myself the unconscious belief that spiders are dangerous whilst agreeing that all the evidence supports the contrary conclusion; this is impossible for a conscious belief, at least where the believer is minimally rational. In standard cases of believing that *p*, a believer will be prepared to support his belief with evidence that *p* and convey his belief not by means of an assertion about himself, but by simply uttering "*p*". It is this relation between the belief that *p* and the utterance "*p*" that I shall be designating with the term "expression", or more precisely "expressive articulation".

The lack of "distance" or the "internal" character of the relation between the belief that *p* and the utterance "*p*" can be read off from the fact that the belief and the utterance have the same truth conditions – different truth conditions from the utterance "I believe that *p*" (Rosenthal 1989, 315ff.; 1993, 201). Someone who utters "*p*" is, so it seems, transposing an attitudinal structure into a publicly accessible medium. He is *articulating* or *expressing* that attitudinal structure, rather than predicating it of himself. Of course, uttering "*p*" does not *entail* that one believes that *p*. This only follows if the utterance is sincere.[3]

[3] Moore claimed that uttering "*p*" "implies" that one believes "*p*" (1942, 540ff.; 1944, 204). Clearly, the sense of "imply" at issue is not that of logical implication. Suggestions have been advanced (e.g. Copp 2001, 9ff.), according to which "imply" should be replaced here with "conversationally" or "conventionally implicates" in a broadly Gricean sense. Certainly, uttering "*p*" implicates in one way or another *that* the speaker is the bearer of the belief that *p*. Note, though, that what is thus implicated is a higher-order proposition *about* the speaker. This relation is to be distinguished from expression, as I am understanding it here, which consists in the *articulation*, rather than the – explicit or implicit – self-ascription, of the relevant attitude. A plausible *explanation* for the conventional implicature of the higher-order proposition is that an assertion is a linguistic articulation or expression of the relevant belief. The implication is given because we generally assume that the expression is sincere, an assumption that can, however, be cancelled by any number of contextual or additional linguistic devices. Searle (1969, 65, 1979, 4f.) makes what is basically the same point in describing the relation between the belief that *p* and the assertion of *p* as an utterance-internal "sincerity condition". I am grateful to David Copp for discussion of these issues.

We are now in a position to pin down the oddity of type 1 Moorean sentences. It derives from the fact that a belief with a specific content is expressed in conjunction with an assertion by the believer that she does not possess that belief. The contradiction that would result from the conjunction of two assertions in "I believe *p*, but I don't believe *p*" doesn't result because the first conjunct expresses a belief of its bearer rather than asserting of her that she is the bearer of the belief.

Of course, the grounds for such a contradictory assertion might be absent if the belief expression were to be insincere. But then the utterance itself would be peculiarly *pointless*: what could be the point of someone insincerely expressing some belief that she denies she possesses in the very same breath? Under these conditions, the utterance would be communicatively self-defeating.

The second point of importance for my purposes can now also be made. This concerns the reason why the assertion "*p*" can be taken as an expression of the belief that *p*. The reason, quite simply, is that believing is essentially *assertoric attitudinising*. In asserting that *p*, we transpose the attitudinal structure of belief into an analogous linguistic form. This is why the bare linguistic representation of a proposition is particularly apt to fulfil the two functions of assertion and belief expression simultaneously.

Now, there are contexts in which "*p*" does not express a belief, but rather a *supposition* or the imaginative *entertaining* of the proposition (Sect. 2.4.2). Where this is the case, we generally understand it is so because some linguistic marker or contextual factor indicates the *explicit withholding* of the usual assertoric framework. Although the standard post-Fregean line here is that imagination and supposition are attitudinally less complex than belief, as they involve the mode-less representation of some proposition (cf. Scruton 1974, 88f.), we need more complex linguistic means to refer to them. This may indicate that there are actually *more* operations involved in imagining than in believing. Imagination or supposition seem to involve what one might think of as a kind of *subscriptive restraint*, that is a mental cancellation of the subscription to the truth of the content that goes with assertion. If this is correct, the capacity to refrain from believing a content one represents as truth-apt is presumably a capacity that has to be acquired through fairly demanding learning processes.[4]

[4] Geach (1965, 457) suggests that the ability to think purely unasserted thoughts may be ontogenetically secondary, but thinks of that ability as one to simply hold the unasserted thought before one's mind's eye. Hare proposes a model for the greater complexity of subscriptive restraint. He claims that assertion itself consists of both the assignation of a content to the assertoric mood and the subscription to the assertorically framed content (Hare 1970, 20ff.). If we apply this distinction to assertoric attitudinising, imagination might then be thought of as an attitude that assigns a content to the assertoric mode, but cancels the possibility of subscription. This would account for the fact that we imagine something being true, not as to-be-realised.

David Velleman distinguishes imagination from belief by means of the "purposes for which" an attitudiniser regards the attitude's content as "true". Whereas in belief one regards some proposition as true with the aim of only doing so if it is true, in imagination one does so "for recreational or motivational purposes" (Velleman 1996, 183f.). Even leaving aside the idea of evolution designing agents' capacities with an eye to their recreational activities (cf. Velleman

The utterance of "*p*", then, is taken, by default, to be the expression of a belief, that is, to articulate an attitude characterised by the assertoric mode, although that mode is not made explicit.[5] Making the assertoric mode explicit involves the kind of addition we usually only make when the belief has in some way been questioned. Then we bring the assertoric mode out into the open with linguistic means such as "It *is* so that *p*", "It is the case that *p*" or "*p* is true".[6] In doing so, we make explicit the point made by Bernard Williams' (1970, 136) justifiably much-quoted slogan, according to which beliefs essentially "aim at truth".

4.1.2 Type 2 Cases: Assertoric Incoherence and Irrationality

I have so far focussed on the distinction between assertion and expressive articulation. This seems to be the best explanation available for the non-contradictory incompatibility of the conjuncts of type 1 Moorean sentences.[7] Before applying the idea of expressive explication to the case of wanting*, it will be useful to examine briefly the case in which there *is* a contradiction between two conjoined self-ascriptions of belief. By the end of the next section, I hope to show that the idea of expressive explication of the attitudes furnishes an explanation for the differences in both our usual and our rational reactions to the recognition that we have contradictory beliefs on the one hand and wants* with incompatible contents on the other.

The sentences that self-ascribe contradictory beliefs correspond not to type 1 Moorean sentences (B(p) & ¬B(p)), but to their type 2 relatives (B(p) & B(¬p)). The self-ascription of both the beliefs conveyed in type 1 sentences delivers a contradiction: the speaker both attributes to herself a property and denies that she possesses it. In contrast, similarly transformed type 2 cases, in which she attributes to herself a belief with contradictory content to another belief she also self-attributes, provide no such contradictory description (cf. Hare 1968, 51). Rather, they provide a description of someone with contradictory assertoric attitudes. Ascribing someone

1992a, 113, note 24), the idea of tying imagination down to particular "internal" aims seems wildly implausible. One of the reasons why the imagination has repeatedly been seen as peculiarly liberating surely lies in the fact that the "mere entertaining" of a proposition can be enlisted in the service of whatever – aesthetic, moral or cognitive – cause an agent may adopt.

[5] Moran makes this point in terms of the "transparency" of the question of whether I believe *p* to the question of whether it is the case that *p* (Moran 2001, 61). I have been arguing that such transparency is only given when the agent is not entertaining an external epistemic relation to his belief.

[6] This is not to say that these linguistic devices *guarantee* the assertoric character of an utterance. It remains possible that *p* being the case or being true is only a supposition.

[7] Here I am agreeing with David Rosenthal's claim (1989, 316; 1993, 201) that we can infer to the distinction between expressing and reporting thoughts as the best explanation for the peculiar nature of Moorean sentences.

some property, be it mental or otherwise, and in the same breath denying that the person possesses it is as blatant a contravention of the law of contradiction as you're going to get. Ascribing someone two contradictory beliefs, on the other hand, is coherent, although, if true, it is likely to spell trouble for their bearer.

Where someone ascribes contradictory beliefs *to herself*, she will normally have a level of awareness that will make that attitudinal combination highly *unstable*. Still, that doesn't make it logically incoherent. Rather than the self-ascription of two beliefs with contradictory contents, what is more likely in everyday language is that a person might actually utter a type 2 Moorean sentence. Someone could well say "I'm over forty, but I don't believe I'm over forty". In doing so, she expresses in the first conjunct a belief whose content she negates in the second[8] – perhaps on the basis of inference from the way she behaves and feels. Of course, when people say such things, they may not mean the second conjunct literally. The important point is that there is no incoherence in the claim that someone might be the bearer of the two states thus denoted.

If someone genuinely believes both that she is over forty and that she is not over forty, then one might suspect that some form of psychopathology is at work. No doubt, someone who had whole sets of such contradictory beliefs about herself would be well on the way to developing a split personality. Under normal conditions, it is highly unlikely that someone who is conscious of having both beliefs will not see a problem in this. Anyone thinking clearly will make the *mental move* from the conjunction of two such beliefs to the formation of a conjunctive belief with the contradictory contents.[9] But this leaves her with an attitudinal content, namely (p&$\neg p$), which is incoherent, i.e. it furnishes her with a self-dissolving belief (cf. Seebass 1993, 284). Because something being the case is equivalent to its negation not being the case, we can make no sense of the idea of something both being the case and not being the case and therefore cannot believe any such thing.

Finally, for my comparative purposes here we should note that the move of "agglomeration", as it has become known since Williams (1965, 180ff.), i.e. from B(p)&B(q) to B(p&q), is, where p and q are contradictory, a *rational* requirement on believing. But why should we make that move? Can there be anything rational about forming a contradictory belief? The answer, I think, is that there indeed can be if doing so makes it clear to the belief's bearer that this what her present doxastic attitudes rationally commit her to. The requirement to agglomerate beliefs with contradictory contents seems to be an instance of a more general requirement that we agglomerate those beliefs that are relevant for each other. It is certainly not

[8] Here, the sentence should be interpreted as it would be in everyday practice, i.e. as an example of transferred negation.

[9] Searle (2001, 249ff.) seems to suggest that beliefs, unlike "desires", necessarily "agglomerate". In order to explain the difference in the relationships of believing and desiring to consistency, he argues that W(p)&W($\neg p$) does not entail W(p&$\neg p$). It is, as I will argue in a moment, correct that there is no such entailment. However, unlike what Searle appears to suggest, there is no such entailment in the case of beliefs either. The differences in the relationship to consistency between the two attitudes cannot be explained in this way.

rational to agglomerate all our beliefs. Obviously, it is hard enough work keeping track of a significant number of our beliefs without permanently attempting to pack them into one big attitude – an aim we could never attain even if we tried. Which beliefs are to count as "relevant" may be difficult to say, but there can be no doubt that, if any beliefs qualify, then those with contradictory contents do.

4.2 Optative and Assertoric Expression

The fact that "*p*" or "*p* is the case" are linguistic analoga of the structure of believing is brought out by the logical oddity of type 1 Moore sentences, which conjoin the expression of a belief with the denial that one is the bearer of the attitude thus articulated. If "*p*" didn't express a belief, there would be nothing strange about conjoining it with the negation of the claim that one believes that *p*.

This suggests, I have been arguing, that the medium of our attitudinising is sufficiently analogous to that of language for a direct transposition from the former to the latter to provide a publicly accessible analogue to the relevant attitudinal structure.[10] And if this true of belief, we would expect it to be true of wanting*. Now such an extension means that the criterion of a sufficient structural analogy between attitudes and utterance cannot lie in the identity of *truth* conditions. This criterion is only available in the case of beliefs because of their assertoric character, that is, because it is of their essence to "aim at truth". Rather, the relevant identity in the case of wanting* has to be more generally conceived as what Searle has dubbed the "conditions of satisfaction" of attitude and utterance (Searle 1983, 20ff.). This in turn can only be guaranteed by identity not only in propositional content, but also in "force" or mode.[11] As I now want to argue, this criterion is met by utterances of the

[10] Conceptions of this kind have been advanced by Geach (1957) 75ff.; Kenny (1963) 209ff.; Vendler (1972) 6ff.; Woodfield (1981/82) 82; (1982) 259ff., Searle (1983) 4ff. and Rosenthal (1989) 311ff. It is no part of the analogy theory put forward here that the analogy is to be *explained* by a particular phylogenetic process. There is a variant of the theory that sees a specific genesis, namely the development of mental structures through the internalisation of overt linguistic processes, as the central argument for the theory. This position, inspired by Sellars (1956) and given its most extended exposition in Aune (1967) 180ff., is neo-behaviouristic in motivation: it attempts to show how the "intervening variables" that are our mental states came into being as a result of the continuation of quasi-linguistic processes, where their third-person observable features had, as it were, withered away. Thus conceived, the analogy grounds in a one-way explanatory story from "outside" to "inside", employing the same Sellarsian myth that functionalism sees as illustrating the genesis of mental talk as an explanatory theory (cf. Sect. 3.3.1). Such a conception has the serious problem of explaining linguistic processes without any reference to mental states.

[11] Note that the utterances sometimes taken in metaethics as paradigmatic expressions of emotions, namely interjections such as "Boo!" and "Hurrah!" (Blackburn 1984, 193ff.), do not qualify as expressions in the restrictive sense of articulations or structural analogues of mental states. As such interjections are *doxastically unmediated linguistic evincings* of mental states, they qualify as expressions in a wider sense. However, as they have no internal semantic structure, no sense can be given to the claim that they are structurally analogous to mental states. A semantically

form “Let *p* be the case”. A want*, I am claiming, is simply the attitude articulated by this linguistic structure. “Let it be the case” expresses the *optative mode*, the converse attitudinal framing to that provided by the assertoric mode.

In order to show this, we need to return to the central idea of Section 3.1: optative attitudes are, in a way that has no precise, or at least only a limited, parallel on the assertoric side, subject to *compounding* with further attitudinal and non-attitudinal factors. These are three in number. In Chapter 3, I began arguing for the existence of a generic concept of wanting* by showing its analytic separability from two such further factors that are normally part of the package picked out by everyday terms such as “want” or “desire”: from beliefs and motivational force. The third factor, which came into play in Sections 3.2.2 and 3.3.4, is a hedonic qualification of some kind, a component often referred to as “affect”. The central claim of part I of this study is that the phenomena of practical mind are analysable as compounds of these factors. This goes together with the claim, which I shall not try to substantiate in any great detail, that the everyday terms of natural languages such as English pick out particularly salient or important compounds.[12]

4.2.1 Optative Attitudinising

If it is correct that optative attitudinising is generally conjoined with at least one of these three additional factors, then one might expect there to be difficulties in finding sentences for wants* that are precisely analogous to Moorean sentences for beliefs. And indeed, because we don’t have any exact equivalent to the term “want*” in everyday parlance, there is no precisely parallel structure. Nevertheless, we can get pretty close. A sentence that just about gets there is “Let it be the case that *p*, but I have no desire for *p*”. “Desire” is preferable to “want” here because it can be used nominally in everyday language; and a nominal formulation is preferable to a verbal one, because the former is, unlike the latter, not subject to complication by the phenomenon of transferred negation.

However, there are still a number of factors at work which prevent strict parallelism with the assertoric case. The first point is merely idiomatic, but is

unarticulated utterance such as “Ouch!” might appear particularly appropriate to evince pain in as far as being in pain is itself semantically unstructured. (On what is right and what is wrong about this claim, see Roughley unpublished b.) This is certainly not true of emotions.

[12] Alongside the terms that appear in Atkinson’s list of “motives” – “wish”, “want”, “desire”, “long for”, “crave”, “yearn” and “hanker” (Sect. 2.4.2) – many emotions either have an optative component or are closely bound up with optative stands. The core of hoping that *p* seems to be taking an optative stand to the effect that *p* be the case. However, perhaps hope is an untypical emotion. Fearing some *x* in the absence of an attitude appropriately expressed by the utterance “Let me get away from *x*” seems inconceivable (Searle 1983, 31). Nevertheless, there are arguments for the claim that such aversive reactions are consequences, rather than features of the emotion (Prinz 2004, 68f.).

perhaps worth mentioning because it also prevents the fully-fledged mobilisation of the kind of intuitions triggered by the doxastic equivalents. This is that the somewhat archaic linguistic structure in the first conjunct is, in contrast to the expression of belief, one that is not used with any frequency by contemporary speakers of English.

More importantly, the use of "desire" tends to be understood as having hedonic implications: it is possible to want* some state of affairs, say a prospectively unpleasant visit to the dentist and, precisely because of the discomfort one expects to accompany its realisation, not to "desire" it. The Moorean logical oddity only arises if we subtract the idea that the realisation of what is "desired" is expected to bring with it hedonic gain.

Now, there is in fact a use of the term "desire" which transports no such implication, although this use is also somewhat old-fashioned or formal.[13] Take the case of an aristocrat on a diet. Asked by his butler "Do you desire to dine, sir?", he might reply sincerely in the negative, in spite of longing to get his teeth into a joint of prize venison.[14] "Desire"-without-hedonic-implications is sometimes complemented by a that-clause, as in "The King desires that you attend the ceremony", a use in which, significantly, "desire" can often be replaced by "request".

Strictly speaking, though, we have no individual terms in the English language by means of which we could reliably refer to what Carnap (1963, 1001) calls "pure optatives" without risking misunderstanding.[15] In our Moorean optative sentences, we will therefore have to make do with the use of "desire", whilst bearing in mind the abstraction from affect. This is, of course, unfortunate in as far as neat linguistic parallels tend to exert more direct influence on intuitions. However, the contingencies of the precise divisions our language makes cannot be criterially decisive for the conceptual question as to what mental structures are thus being picked out. Even less elegant, but no less correct would be a construction of the second conjunct out of a disjunction of all the lexical items that can stand for optative attitudes, giving us an utterance such as "Let it be the case that *p*, but I neither want *p*, nor yearn for *p*, nor fancy that *p*, nor fear that $\neg p$, nor ..." (cf. Sect. 3.1.1).

Actually, the expression of the optative mode itself, even ignoring its somewhat archaic tone, also transports assumptions that are extrinsic to its modality. This is true whether it is rendered by "Let *p* be the case" (Kenny 1963, 218; 1975, 32;

[13] As Robert Audi (1973b, 59ff.) has pointed out, there are also some uses of "to want", as in the question "What do you want?", which convey no such hedonic implications.

[14] The example is taken from Davis (1986), 67. Davis has the linguistic intuition that the nominal and verbal uses of "desire" are criterially separable, an intuition I (unimportantly) don't share. An important, philosophical disagreement with Davis is with his systematic claim (1986, 63ff.) that hedonically qualified and non-hedonically qualified "desires" are two completely different kinds of state. See below, Section 5.4.

[15] For this reason, Carnap suggests the use of the Latin "utinam".

1989, 41; Hare 1963, 55; Goldman 1970, 101f.[16]; Seebass 1993, 71; 2006, 114) or "Would that *p*" (Carnap 1963, 1004; Kenny 1989, 40). Both at least conversationally implicate doxastic or motivational features of the kind that were factored out in Section 3.1. In the former case, the implication is that the person believes that *p is not the case* (cf. Sect. 4.3.3). The same assertoric implication is also transported by the latter expression, along with the belief of the utterer that she is unable or insufficiently motivated to bring about its content.

A number of authors have use the term "optative" to refer to precisely this latter compound state, which is thus by definition deprived of any relevance for action (Peters 1961/62, 127; Hare 1968, 51; Donagan 1987, 142; cf. Anscombe 1957, 67). This terminological step is generally bound up with the claim that "wants" are best expressed by commands (Hare 1952, 34; Woodfield 1981/82, 77; Donagan 1987, 142). From this perspective, "Let it be the case that *p*" is categorised as an imperative.

Although Kenny, who first developed the expressive explication systematically,[17] freely interchanges the designations "optative" and "imperative", doing so is easily misleading (cf. Carnap 1963, 1001; Dummett 1972, 449; Seebass 1993, 266, 277). The use of a verb in the imperative mood generally implies that it has a grammatical subject which has been omitted, usually a second-person but possibly a third-person subject or other description, the latter variant being not uncommon in commands with "let" (Quirk et al. 1985, 828ff.). However, as I am using it, the formulation "Let it be the case that ..." is to be understood as having no implications as to whether it has an *addressee*. One terminological proposal advanced to deal with this is that we divide imperatives into "directives" and "fiats", where the former "[include] an indication of the agent who is to carry it out" and the latter don't (Hofstadter and McKinsey 1939, 446).

What is at stake here is, of course, not the term we should be using, but the precise features of the attitude in question. The unaddressed "Let it be the case ..." is apt to capture the generic state of wanting* because it articulates the mode of non-action wants* as accurately as that of action wants*. There are obviously no potential addressees where *OSI* wants* concern states of affairs for which non-agential processes would have to be causally responsible. And clearly, wanting* does not presuppose the belief in any transcendent agency. Moreover, it follows from the arguments advanced in Sections 2.3.2 and 3.1.2 that, even when we turn to wants* whose contents are states of affairs realisable by the attitude's bearer, wanting* such a state of affairs is not equivalent to wanting* to bring it about. Although wanting* to bring about *p* entails wanting* *p*, the converse is not the true. Wanting* *p* has no implications as to how *p* is to come into being. Specifications of agency, if there

[16] In contrast to the broadly Fregean proposal that we distinguish an optative mode by means of which a proposition is framed in wanting*, Goldman (1970, 102) construes the difference between wanting and believing as lying in whether the proposition "assented to" is itself optative or declarative. In this respect his position resembles Castañeda's (cf. Sect. 2.4.2, note 18).

[17] Kenny's analysis of what he calls "the will" builds on Geach's analysis of judgement in (1957).

are such, are components of a want's* *content*, in the case of motivational states in the form of first-person representations of their bearer. Wants* concerning other persons' actions will often be best expressed as requests or commands.[18]

It is notable that the expression of wanting*, in contrast to that of believing, requires an explicit linguistic marker. There is normally, outside children's language use, no question of us taking a linguistic representation of some content without any modal marker to be the content of a want*. Related to this is the fact that, although no reference is made to the want's* bearer in its expression, the explicitness of the modal marker makes it clear that there is some *perspectival locus* of the claim being made on the way the world is to be. The analysis of wanting*, more obviously than the analysis of belief, confronts us with the fact that there is no attitudinising without a subjective perspective from which that attitudinising can take place.[19] We can make good sense of *p being* the case without having to make reference to attitudinisers. But it is difficult to see how the idea that *p is to be* the case can be given sense if we don't presuppose the existence of a bearer of the is-to-be mode.

4.2.2 Type 2 Cases: Optative Incoherence and Irrationality

The assertoric "*p* is the case" essentially involves reference to a pre-given standard, namely that of the way things are, against which it is appropriate to evaluate the attitude's content. In contrast, the optative mode is essentially a matter of setting standards against which the way things are is appropriately evaluated. I shall return to this point in detail below (Sect. 4.5). At this stage, I want to make just one brief remark on the concept of appropriateness at work here, viz. its *attitudinally internal* character. It may for any number of reasons – prudential, moral, aesthetic – be appropriate *not* to evaluate the world with respect to the extent that it corresponds to or resists the realisation of the contents of certain of one's whims, longings or most secret desires. These are all forms of appropriateness which are brought to

[18] Requests or commands can be expressed linguistically by means of the verb "let", used as a synonym for "allow", as in "Let my people go" or "Let the prisoners go free". A purer, unaddressed expression of a want* is "Let there be light" as uttered by the god of the Old Testament. Here, "let" does not function as a content word, but as a substitute for the subjunctive mood. (Compare the French "Que la lumière soit" and the German "Es werde Licht".) Omnipotence may be precisely the lack of any causal gap between taking on a (particular sort of) optative stand and the realisation of its content.

[19] Thomas Nagel (1970, 121) has claimed that the difference between "Would that *x*" or "If only *x*" on the one hand and "I want (or wish or hope) *x*" on the other is that the former are *impersonal* expressions whereas the latter is a *personal* expression of a "desire". However, the converse would be more accurate: Nagel's characterisation of "Would that ..." as impersonal is misleading, because it ignores the perspectival locus which marks all attitudinising. As the expressive articulation of a want*, "Would that ..." is essentially tied to the perspective of its bearer, irrespective of what appears in its content. "I want ...", on the other hand, may be used descriptively as a distanced and thus "impersonal" assertion *about* one's own mental state.

bear on our attitudes from outside and which tend to involve standards that have some level of intersubjective acceptance. In contrast, the internal appropriateness of measuring the world against the contents of one's wants* is simply given with optative attitudinising. The claim that wanting* involves *subjective standard setting* itself entails nothing about the normative status of those subjective standards relative to standards that transcend the merely subjective.

The fact that wanting*, unlike believing, entails standard setting on the part of the attitude's bearer has consequences where a person has two or more wants* with contradictory contents. These consequences concern both what is psychological *usual* and what is attitudinally *rational*. Firstly, a conjunction of attitudes such as W(p)&W($\neg p$) is *far less unusual* than the corresponding B(p)&B($\neg p$). Someone who consciously both believes that she is over forty and believes that she is not over forty will only be able to maintain both states by keeping certain trains of thought separate – thoughts which justify the two beliefs and which derive consequences from them. Normally, such a conjunction of contradictory beliefs is going to be a transient affair. It is certainly highly unstable in rational persons: becoming aware of the contradictory contents normally suffices for their bearer to agglomerate them, thus taking a step into assertoric incoherence, i.e. a belief in nothing (Sect. 4.1).

In contrast, wanting* two contradictory contents is anything but unusual. Some people both want* to be over forty and not to be over forty; Des both wants* to see Amy and wants* not to see her; the dieting aristocrat both wants* to eat and not to eat. None of these people need demonstrate a particular tendency to agglomerate the wants* in question, that is, to come to want* ($p\wedge\neg\ p$). In general, they are going to have these contradictory wants* for different reasons, that is, as a consequence of other wants* whose contents are anything but contradictory: wanting* both to have certain opportunities ahead of you and wanting* the wisdom that age brings; Des both wanting* to see the woman he cares for and wanting* not to get hurt; the aristocrat both wanting* to enjoy the sensual pleasures of food and wanting* to avoid another heart attack.

In each case, both contradictory contents justified by contingent relations to these background wants* may continue to be wanted* by their bearers and felt to be reasonably wanted*. And the agents need see no reason to conjoin the contents of the two attitudes. The fact that doing so would involve them in incoherence is, rather, a reason not to do so. In contrast to beliefs, which it is rational to agglomerate wherever they are significantly relevant for each other, this is not necessarily the case with wants*.[20]

[20] According to Mele, the ability for self-control depends on the fact that even closely related "desires" don't necessarily agglomerate. He argues that an agent may act on the desire to exercise self-control in order to get back to work, although he wants to continue watching TV more strongly than he wants to get back to work. Were the various relevant desires to agglomerate, they would produce a single motivationally preponderant desire relevant to the situation which, Mele assumes, would carry the day. Without agglomeration, the desire to exercise self-control can play an independent causal role (Mele 1995a, 32ff.; 1999). If Mele is right, it would clearly not be rational for the agent to agglomerate his desires in such a case.

It can be perfectly rational to maintain two wants* with contradictory contents separate from each other. Where someone wants* p and wants* $\neg p$ for different reasons, the tension between the two optative states may be worth maintaining for instrumental reasons – things may change. Or the cost of want* shedding may be greater than the cost of continued tension. The tension may even be a deeply felt existential rift in the bearer, that is, a sense of being torn that the person experiences as an important feature of who he is. In any of these cases, the move into attitudinal incoherence which would result from agglomerating the two want* contents need not, as in the case of contradictory beliefs, provide greater clarity on the total attitudinal state in which the agent had been all along. Rather, it may involve a dissolution of tension that is either instrumentally short-sighted or existentially superficial.

In contrast, believing cannot be made sense of without the idea of coherence of what is believed. Believing is a matter of representing some proposition as being the case or true, a standard that entails coherence. Thus, anyone aware of believing contradictory propositions must be aware that there is something internally deficient about her overall mental state, if she understands what believing is. Because wanting* is a matter of subjective standard-setting, no such internal deficiency need result from a conjunction of optative attitudes with contradictory contents. Of course, problems may very well result, but these will be *effects* of the optative combination and depend on the level of motivational force and other forms of arousal that the individual wants* end up mobilising. Such typical effects are indecision, inactivity, frustration, regret, worry and perhaps psychosomatic symptoms. How one should react to such effects is itself a further question, the answer to which requires weighing up the importance of the conflicting wants* against the negative effects.

Consider a case of *OSI* wanting*. Someone might, for historical reasons, support two football teams, who end up playing against each other. We understand perfectly well what he means if he says that he both wants* team A to win and wants* them not to win. And nothing necessarily makes it advisable for him to fuse the two attitudes, thus confronting himself with the optative incoherence that rationally requires attitude change. On the contrary, as things stand, he might be able to feel some joy at the result whoever wins.

Things are no different when we turn to action wants*: a single want* of Des both to see Amy and not to see her would be incoherent, again simply because of the logical impossibility of ($p\&\neg p$). But he may well hold onto both unagglomerated wants* and rationally continue to see her, in spite of the pain it causes him. He may feel the pain is worth it – although he doesn't discard the want* not to see her either. He may, moreover, be aware that opting *not* to see her will cause the desire to see her to grow into an intense longing. This may make it rational for him to live with the tension, rather than to attempt to dissolve it.

There is no parallel to this on the doxastic side. The rationality of belief is closely bound up with the truth of beliefs' contents and therefore with their coherence. The role of coherence in the rationality of wanting* is, in contrast, mediated by its practical and emotional consequences.

4.3 Axiological Conceptions of Wanting*

Playing host to wants* with contradictory contents is neither unusual nor necessarily irrational and there is no rational pressure to agglomerate such contents as in the assertoric case. These differences ground in the differences in the way the two attitudinal modes employ standards. This constitutive difference also shows up in a further difference between the two attitudes: in their relationships to *transindividual criteria*. As beliefs "aim at truth", mirrored in the fact that the expression of the belief that *p* is simply "*p*", believing *p* essentially involves a *claim* to objectivity. But wanting* has no such component. Although "*p* should be (the case)" has a certain plausibility as a candidate for the expression of a want*, putting things this way is likely to be misleading as "should" implies that there are transindividual criteria to which an appeal is being made. But this is not necessarily the case where a person wants*, or indeed wants something.

This point is rejected by a tradition of thinking about motivation that begins with Plato's *Meno* and is encapsulated in the scholastic dictum "omne appetitum appetitur sub specie boni", a tradition that, since Anscombe's *Intention* has frequently provided a filter through which Aristotle has been read (cf. Vogt unpublished, 5). Authors influenced by this tradition have argued that there is actually a strict parallel between belief and "desire" (Anscombe 1957, 76; de Sousa 1974, 538ff.; Stampe 1987, 355; Scanlon 1998, 38f.). The idea is that "desiring" involves "aiming at the good", just as believing involves "aiming at the true", or that "the good" is the "formal object" of "desire" just as "the true" is the formal object of belief. In Anscombe's well-known formulation, wanting must involve seeing the want's object under "desirability characterisations" (1957, 70ff.).

Anscombe's – negative – argument for this conception is that failure to understand wanting* in this way results in a lack of "intelligibility". The well-known examples by means of which she illustrates the relevant lack feature the putative want* to spread all ones green books on the roof of the house (1957, 26f.) or to possess a saucer of mud (1957, 70). What Anscombe sees as decisive is the hermeneutic vertigo in the face of someone who doesn't give any further reasons for what look like an unusual whim, but simply says "I just happen to want it". Whoever talks this way in such contexts is merely producing, she says, "fair nonsense", and indeed should be seen as a "babbling loon". Now, the question for a theory of wanting* is what it is precisely that lacks intelligibility in such cases. Is it the claim that an agent can have unusual wants* not backed by further reasons or is it the person who is the bearer of such wants*? Anscombe explicitly says that the problem lies in the fact that "we could not understand such a man" (1957, 27). This is not implausible. But in so far as it is true, the problem is that we cannot understand why he might want* such things. And this problem presupposes that we can make sense of the claim that such things are wanted*. What we have difficulty making sense of is the reasons a person might have for the attitude. But that is quite simply a different question from whether we can understand that he is the bearer of certain wants*. It is, above all, one whose answer does nothing to impugn a concept of wanting* whose intelligibility it presupposes in concluding that there is something strange about the bearer of wants* of these kinds unbacked by special reasons.[21]

So much for what has been taken to be the most important argument for the axiological conception of wanting* (cf. Tenenbaum 2003, 48ff.), an argument that is supposed to work by disqualifying alternative conceptions. If we try to find evidence for the positive claim in everyday optative or desiderative phenomena, we come across two reasons why it should be rejected.

4.3.1 Want Satisfaction*

If I tell you that your taxi has arrived or that the weather forecast predicts sunny weather, you might say "Oh good!" There are other idioms in the English language such as "It's a good job ..." or "It's a good thing that ..." which are used on similar occasions. In such cases, you come to believe that something you want* has come to pass, or is perhaps likely to come about (cf. Tugendhat 1976, 521, note 18). Note that, in the taxi example, you might instead give expression to an emotion, for instance by uttering "Oh, I'm glad" or "That's a relief". These expressive uses of "good", which are responses to the belief in a want's* satisfaction, or unproblematic satisfiability, provide no support for the idea that wanting* itself requires a prior conception of the good.

Moreover, this use of "good" is weak in a clear sense. Uttering "good" in these ways has no implications that the utterer will re-apply the word in similar situations. This contrasts with a strong or genuinely evaluative use of "good", which entails precisely this. Seeing some item of type *x* as good in circumstances *c* commits the attitudiniser, at least pro tanto, to seeing other *x*s as good in relevantly similar circumstances. Put slightly differently: seeing *p* as good in the strong sense involves seeing it in the light of some *standard* that one takes to have *transcontextual applicability*. Standards for good football matches specify properties of football matches that, pro tanto, justify seeing them as good. The same is true of good works of art, good actions and good insults.[22]

[21] Anscombe conflates the question as to the formal object of wanting with the question of what makes an action intentional. This is a completely different matter, as a piece of behaviour's being wanted* by its agent is insufficient to qualify it as intentional. See below Section 5.1.3.

[22] This doesn't mean that there is some sort of deductive relationship between standards for goodness and judgements about empirical cases. Jonathan Dancy has shown the there are serious difficulties in giving a precise formulation of the way principles pick out the properties that might be thought to determine correct value judgements across different situations (Dancy 1993, 72ff.; 2004, 73ff.). But the above point requires only a fairly informal notion of transcontextual applicability or, as Bernard Williams put it, "an idea, however minimal or hazy, of a perspective in which it can be acknowledged by more than one agent as good" (Williams 1985, 58).

4.3.2 Transcontextual Criteria?

The utterance of neither “Oh good!” nor “It’s a good thing that ... ” need imply that any such standard is being applied.[23] The same is true of any utterances expressing wants*. The “standard” set in wanting* need have nothing transcontextual about it. It might be completely whimsical, perhaps the product of chance thoughts or of contingent physiological changes that are never to recur. By wanting* to eat fish one evening in a restaurant, I don’t commit myself to the same movement of mind on any other occasions in the same restaurant or elsewhere. If someone is a regular fish-eater, that will normally be so because that is what he regularly feels like eating. But where no such feeling crops up, what he wanted under earlier, similar circumstances will usually be completely irrelevant.

The additional assumptions we have to make in order to come up with exceptions make it clear that this is right. Someone like Phillip, who orders fish “on principle” whenever he eats out at Luigi’s, has installed in himself a special kind of disposition to develop a particular type of want* under specific circumstances. But it ought to be obvious that much of what we want* - certainly the objects of those wants* we label “desires”, “longings”, “yearnings”, etc – are not wanted* “on principle”. Rather, our wants* can come and go without any clear reason. They sometimes surprise us, sometimes even overwhelm us. And where they are characterised by regularity and predictability, this will often be the result of – more or less transparent – causal processes. The fact that many people desire peace and quiet after a hard day’s work is not determined by the logic of the concept of desire, but by physiological and psychological processes.

It is the dependence of the idea of goodness on transcontextual standards that enables seeing things as good to serve as a platform from which to criticise desires, including our own. This is an essential, if not *the* essential “functional role” of the idea of the good. Examples in which it fulfils this function are easily forthcoming. For one, people can experience sudden “impulses” or “urges” to do things they don’t value in the slightest, as when Ray, standing on the platform at a railway station, feels that he “must” jump down onto the track (cf. Copp 1995, 169ff.). Although some might object that the contents of such forms of motivation are not things we (really) “desire” or “want” in the everyday sense, the “must” by means of which they are typically expressed plausibly indicates that we are dealing with a species of the optative genus.

Further, wanting* – in particular, wanting or desiring – things we take to be worthless or even nefarious is by no means restricted to such cases. As Michael Stocker has persuasively argued, people sometimes develop self-destructive desires, other-directed desires out of bitterness or malice or “perverse” desire for repulsive

[23] That there is such an implication is implausibly claimed by Korsgaard (1996, 94).

goods (Stocker 1979, 744ff.).[24] Even if it were to be true that people only develop such wants* under pathological conditions, that wouldn't mean that there is anything *conceptually* wrong with them. Anyway, such wants* are surely far more common than talk of "pathology" implies. We understand perfectly well what it is to desire something with a bad conscience because we believe it isn't worth desiring.

Note here a contrast to the meta-predicate "true". Under normal circumstances, "true" and its negation are only used to criticise the beliefs of other people or our own non-contemporaneous beliefs. The assertion that one believes some proposition combined with the expression of the belief in that proposition's falsehood – "I believe that *p*, but *p* is false" – gives us something close to type 2 Moore-paradoxicality and thus denotes a highly unstable attitudinal conjunction that only occurs under extremely restricted circumstances. In contrast, value predicates and their negations – "good" and "bad" as well as their "thicker" relatives such as "(un)important", "(un)just" etc – can without paradox be employed at time *t* to criticise our own wants* at *t*. Conjunctions of dissonant evaluations and want* expressions can also express wilfulness or stubbornness on the part of their bearers. Someone in one of Stocker's examples might combine "I don't see any value in *p*" or "*p* is bad" with "but let *p* still be the case".

The gap between wanting* and value judgement speaks decisively against seeing "the good" as the "formal object" of "desire": unlike someone who "justifies" their belief that *p* by saying that *p* is true, someone who justifies or explains their wanting* some *p* by saying "because *p* is good" will (pace de Sousa 1974, 538) have said something that is at least minimally informative in so far as it excludes the relevant want* having a counterevaluative status. The gap would also be closed if Scanlon's Anscombian claim that that desiring *x* entails seeing *x* as "desirable" (1998, 38) were correct. However, the proposal is incompatible with the natural assumption that taking something to be the appropriate object of some attitude presupposes an independent understanding of what it is to take on that attitude. If "desirable" means "worthy of being desired", Scanlon's claim lands the desirer with an unfortunate semantic regress: desiring *p* involves seeing *p* as worth desiring, that is, as worth seeing as worth desiring, etc. This can hardly count as an informative explication.[25]

[24] Or in David Velleman's words, agents may be "disaffected, refractory, silly, satanic or punk" (Velleman 1992a, 99).

[25] Certain formulations chosen by Scanlon to characterise "desiring" something – as involving "a *tendency* to see something good or desirable about it" (1998, 38; my emphasis) – actually leave open the possibility of the relationship between "good" and wanting* that I am canvassing. If "desiring" only involves *being disposed* to see the object of the desire as good, then there will presumably be desire tokens for which that disposition is not triggered.

4.3.3 Temporal Specification

A final consideration relevant to the relationship between wanting* and "good" will enable a further clarification of the sense in which the optative mode constitutes a generic attitude. It has often been remarked (Aquinas, S.th. Ia2ae 30,2; Meinong 1894 §5; Anscombe 1957, 69; Kenny 1963, 115f., 119f.; Goldman 1970, 94) that we use words such as "want" or "desire" to pick out contents that are, or which we believe to be unrealised, whereas there is no problem in describing as "good" some state of affairs that we know to exist. Indeed, where some state of affairs is thought not to exist, we tend to conditionalise axiological claims, saying it "*would* be good" if it existed.

According to the optative conception of wanting*, this is a superficial grammatical difference that is irrelevant for the character of the constitutive attitude. The optative mode entails no temporal or other restrictions on its content. We can want* something to have happened, to be the case now or to take place in the future, just as we can have beliefs about propositions with any (or no) temporal indices. We use the assertoric terms "remember" and "expect" with specific temporal (and other) restrictions concerning the relationship of the time of attitudinising to the time index of the content. Similarly, we have different everyday terms to pick out different temporal relations to the contents of our wants*: "regret", "fear" and "hope" all tend to be used with a specific relative temporal indexing of their contents, in spite of having the same optative component.[26]

Note further that if our wants* were to dissolve the moment they are satisfied, it would be difficult to make sense of feelings of satisfaction (cf. Sects. 3.2.2 and 5.3.1). These generally result from the belief that the content of some want* has been realised. Importantly, such forms of affect can persist over a considerable period of time. Why should a person continue to feel "satisfied" on thinking of the proposition in question if the want* of which the proposition is the content has not survived the onset of the belief that it has been realised? This suggests that the fact that we often cease to say "I want *p*" when we believe *p* to be the case is no more than a linguistic contingency. Actually, as has been variously pointed out (Schiffer 1976, 195; Beardsley 1978, 167), there are cases in which we do use the language of "desire" to refer to states of affairs we believe to be realised. Consider "She is everything I desire" or "I am lying here in the sun and this is exactly what I want to be doing". Such linguistic facts demonstrate that there is considerable leeway in the use of the everyday terms for wanting* and that, once again, we shouldn't make anything substantial depend on how we happen to talk in particular cases.

Once we are clear, then, about the generic character of the concept of wanting*, we can see that there is no want*-belief asymmetry with respect to temporal specification. This contrasts with the highly significant asymmetry with respect to transindividual standards.

[26] Such indexings are, however, only paradigmatic, not necessary: we can fear or hope that something has already happened, just as we can regret that we will be doing something in the future.

4.3.4 Attitude-internal Standards

We *could*, somewhat misleadingly, say that to want* *p* is to take on the attitude according to which *p should* be the case, as long as "should" is understood as a "subjective claim". Talk of "claims" itself generally involves an appeal to an intersubjectively accepted, or at least acceptable *standard*. This is true whether the "should" is of the instrumental or moral variety, i.e. whether it grounds in beliefs about the suitability of means for ends or the morally right or good. Clearly, neither of these is implied by the optative "Let it be the case". One way to think of the optative posture is as a "should" minus any such elements.

However one might want to put it, there remains the difference between the assertoric and the optative attitudinal modes that the former constitutes a non-public, but literal claim, whereas the latter only constitutes a "claim" in a reduced sense. It is this reduced kind of claiming[27] that is common to both the assertoric and the optative attitudinal postures. In the assertoric case, the appeal to an objectively given standard is the core of the matter: the claim is that the content represents the way things are. Taking on the optative stand, in contrast, does not aim at satisfying an objectively or intersubjectively given standard, but is itself the *setting* of a standard to which objective givens have to correspond, should the attitude be satisfied. The symmetry between the two attitudes here thus lies not in both appealing to pre-given standards, but in their converse relationships to standards.

It is no coincidence that axiological conceptions of wanting* tend to be advanced in discussions of moral psychology. The axiological conception of "desire" can be understood as an attempt to fashion a generic notion of motivating attitudes that opens up a conceptual space for moral orientations, where these may ground in considerations – perhaps standards – external to the agent and only come to take effect through some form of understanding on his part. The bodily appetites are clearly unsuitable paradigms for a generic concept of wanting*, as Plato, Aristotle and Kant were all concerned to argue. Because "epithumiai" or Kantian "inclinations" appear too immediately dependent on sensation (GMS 413), in particular on pleasure or pain (DA 414b4-6), they are bad candidates to be the vehicles of moral motivation. However, conceiving the generic attitudinal mode in axiological terms distorts our grasp both of our basic bodily appetites, aversions and urges and of a whole set of akratic or otherwise axiologically non-aligned wants*. My claim is that the optative "Let *p* be the case" is the form of attitudinality common to both the most basic urges and the most noble moral goals of human persons.

[27] Cf. Seebass (1993) 86ff., where "claiming" ("einen Anspruch erheben") is seen as the genus of which assertive and optative attitudes are the two species.

4.4 Wants* as Mere Entailments

4.4.1 The Case for Non-existent "Desires"

A second approach[28] to wanting* developed for what appears to be its particular appropriateness to moral cases has been highly influential. According to this view, when an agent has φ-ed, it follows that she "desired" to φ in a very broad sense of "desire", although adducing the desire does no explanatory work. The basic idea is a slightly revamped version of the first premise of the Logical Connection Argument (Sect. 2.5.3). Like its Logical Behaviourist predecessor, it denies that the ascription of a "desire" adds anything to an action explanation, as to do so is merely to characterise the agent's behaviour as – depending on the precise version – motivated or intentional (Nagel 1970, 29ff.; Dancy 1993, 8ff.; 2000, 13f., 85; Schueler 1995, 34). In order to distinguish the use of this claim from that to which it was put under the immediate influence of Ryle and the later Wittgenstein, I shall label it the *Logical Connection Claim*. At stake is here no longer, or at least not primarily, the question of whether action explanation is causal, but whether it requires the adducing of a "desire".[29]

The attraction of the Logical Connection Claim in the context of moral psychology is that it allows us to say that an agent who helped someone in trouble did so because she recognised the requirement to help, not because she happened to be playing host to the desire to do so – although it remains true that she did indeed want to help, in as far as she didn't act involuntarily. According to Nagel, who first made this move, the same is true of prudential cases, where we are also faced with requirements. The recognition that I have to do my tax returns may be sufficient to move me to do so, so it seems. In neither kind of case, Nagel thinks, is a desire "a contributing influence" or "a causal condition"; rather, it is merely "a logically necessary condition" (Nagel 1970, 30; cf. McDowell 1978, 84).

In such cases, Nagel talks of "motivated desires", that is, desires explained by other attitudes, in these examples by normative beliefs. He distinguishes these from "unmotivated desires", such as the desire for food. As Dancy points out, it remains unclear whether Nagel thinks of "motivated desires" as Humean independent existences, which, although they are explained by whatever explains the action, are nevertheless necessary features of the agential machinery (Dancy 1993, 8f.). Dancy himself argues that talk of a "desire to φ" does not pick out any independent mental state, but rather summarizes the claim that the agent was motivated to

[28] In Section 7.2 of *Moral Vision* (McNaughton 1988, 110–113), McNaughton runs the two construals discussed here in Sections 4.3 and 4.4 together. This is unhelpful. If, as McNaughton claims, "desires" are actually beliefs and beliefs that the requirement is given are necessary and sufficient for the agent's corresponding motivation, then "desires" don't drop out of the picture. They are still there, but can unfortunately be described in two different ways.

[29] G.F. Schueler (2003) argues both against the necessity of adducing what he calls "desires proper" in action explanations and against the contention that action explanations are causal. However, he uses the Logical Connection Claim only to support the first point.

φ – by some belief. Following Dancy, Schueler replaces Nagel's talk of "motivated desires" with that of "intention-generated desires", which are in Dancy's words "pure ascriptions". These Schueler contrasts with genuinely independent existences, what he calls "desires proper" (Schueler 1995, 29f.; 2003, 24).[30] An example Schueler gives of a purely ascribed "desire" is the motivation to go to a really boring meeting; the motivation to stay at home and read would, in contrast, be constituted by a "desire proper".

4.4.2 Requirements and Reluctance

As far as I can see, there are three kinds of motivation for the espousal of the mere entailment view – whether in its "pure ascription" form (Dancy) or in its hybrid variant (Schueler). We can label them normative, functional and phenomenological.

The *normative* motivation, which was primary in Nagel's original argument and plays a more or less explicit role in its subsequent versions, is the concern that the obligating character of – particularly – moral requirements might appear compromised if an agent needs an additional "desire" in order to be able to satisfy them. On the face of it, it is puzzling why the normative authority of a requirement might be undermined by a rejection of the explanatory claim that its recognition can bring about actions required in the situation without optative mediation (cf. McNaughton 1988, 48). Perhaps the worry grounds in an interpretation of "ought implies can", according to which the ability in question depends on the contemporaneous presence of motivation, rather than on the capacity to generate it. If motivation were taken to require prior "desires" to perform the action required, then the purchase of normative demands would be severely restricted – in a way that, for normative reasons, we shouldn't accept. Nagel rightly rejects this last premise. His "motivated desires" (Nagel 1970, 29ff.), which require no such prior wants* from which they can be derived, are – in spite of their doubly misleading label[31] – a helpfully distinguished group. My only beef is with the mere entailment conception of such states, according to which nothing is picked out by the corresponding terms.

The reasons for this interpretation appear to be functional and phenomenological. The *functional* argument is that any such additional kind of state is simply unnecessary, as it adds nothing to an explanation in terms of the requirement's recognition. Note, however, that at least one mental move will often be required. Recognising that some action is required is only going to move me to act if I insert a

[30] This second concept would presumably not be underwritten by Dancy, whose pure theory works entirely with cognitive states (Dancy 1993, 18ff.; 2000, 85ff.).

[31] First, we wouldn't think of the relevant attitudes as everyday desires, because of their lack of hedonic qualification. Second, it seems unhelpful to label the generation of wants* out of normative judgements as a process of the former's "motivation".

directly first-person referring device into the content of the recognition (Sect. 2.3.2). So even for functional reasons, an agent needs to take a step further than merely recognising the requirement. I am claiming that the necessary mental moves are not complete until he takes on an optative stand that represents his performance of the act he takes to be required. That may look unparsimonious, but parsimony can appear appropriate at different, competing levels.[32] Dancy's pure ascription view, according to which there are simply no independent existences appropriately labelled "wants", can make the strongest claim to parsimony.[33] But once one rejects such a view, as I think we should for phenomenological reasons to which I shall come shortly, a hybrid conception has to postulate two different kinds of explanation for the cases with and without "desires proper". In such a construal, there are less want tokens, but more types of explanation. It is unclear whether theoretical elegance speaks for accepting the latter, rather than the former increase.

What I think in the end has been taken to be decisive is the *phenomenology*. In many cases in which we act because of recognising requirements, there seems to be no experiential evidence that we need to take mental steps that go beyond such recognition in order to act so as to satisfy the requirements. We can add that in such cases, typical justifications for action – "I had to help"; "I had to take part" – refer only to the requirement's content. It follows from the transparency of the assertoric mode (Sect. 4.1.1) that such justifications are the expression of beliefs. Finally, the agent's being guided by external standards would appear to preclude her setting subjective standards of her own by adopting optative stands (Sect. 4.3.4). In cases of action in the face of requirements, then, at least sometimes, the postulation of a want* appears phenomenologically "otiose" (Platts 1979, 256).

In answering this charge, I shall look at variations on a specific case that will enable me to tease out the relevant phenomenology. Such a discussion won't provide a knock-down argument against mere entailment views. Phenomenology doesn't provide knock-down arguments. However, as the strongest evidence for mere entailment comes from phenomenology, phenomenological counter-evidence has weight. At the end of the section, I will back my counter-claim with some linguistic data that I think is best explained by a realistic construal of the optative conception.

Begin with the commonplace that people often do what they "are supposed to do" in particular situations, in spite of not wanting, or "not really wanting" to do so. Take the case of a doctor who knows that a well-known chronic hypochondriac is waiting to see her. The doctor, let us imagine, feels obliged to see the pseudo-patient, but really doesn't want to. Her thus not wanting is likely to be a matter of hedonic features of her experience. Perhaps she already has a sinking feeling at the

[32] Compare the remarks at the end of Section 9.4.1, including note 12.

[33] The completed version of Dancy's "pure" theory (2000, 98ff.) is even more parsimonious. Here, Dancy radicalises his conception, claiming that the real explanatory factors behind human action are the extra-mental states of affairs about which agents form beliefs. Similar forms of explanatory anti-mentalism are advanced by Bittner (2001, 81ff.) and Stoutland (2007).

thought of seeing the person or perhaps she knows from experience that, as soon as the hypochondriac walks into the surgery, she will start to feel annoyed. Moreover, negative hedonic experience, or an expectation of some such experience is likely to go hand-in-hand with the wish that the pseudo-patient were not there. Now, as the doctor's receptionist is on holiday, she opens the door to invite the patient into her surgery herself. Let us say that the doctor's behaviour can be explained by her recognition of some feature of an ethical code of medical practice. This appears to be exactly the kind of case of which Schueler would say that, as no "desire proper" is in play, there is no need to postulate a mediating want in order to explain what the doctor does.

I have no objection to the claim that the doctor has no "desire proper" to see the patient, as having what we think of as a desire in an everyday sense plausibly involves some sort of hedonic feature – either a negative feeling accompanying the want's* non-realisation or a belief that its realisation's anticipation will be pleasant.[34] But the alternative to a hedonically qualified want* is not no want* at all. Perhaps the best way to see this is to move from introspection to an intersubjective version of the same case. In this version, the receptionist is not on holiday and the doctor has to communicate with her in order that she can perform the action the doctor performs in the story's first version. The doctor might say into her intercom, "Let him come in (even though I don't want him to come in)". In doing so, she would avoid Moore-paradoxicality by dint of the hedonic connotations of her use of "want" (cf. Sect. 4.2.1). Now, it certainly seems that she expresses an attitude with the first conjunct of her utterance. We might ascribe the relevant attitude by saying that she assents, or consents to see the pseudo-patient. She may also may also be said to be *willing* or *prepared to* see him, even *resigned to* doing so. Leaving aside the differences of detail between these terms, they all have in common that they denote an optative attitude generally taken on in the face of other countervailing wants*, wants* which are often affectively qualified. As I remarked, the doctor sees herself as acting in accordance with a requirement that she accepts as part of her professional code of conduct. Her context-specific want* is a case of her bringing her general assent to bear on a specific situation. We could characterise this as a case of subjective standard setting malgré soi. The standard of professional conduct is what imposes the requirement; her congruent optative stand, adopted in the face of countervailing wants* and unpleasant feelings, sets her personal conditions of satisfaction for what then happens in the particular situation.

Now, it follows ceteris paribus from the doctor's words that her reluctant request to the receptionist expresses an optative attitude. Do we have any reason to think that this attitude toward the pseudo-patient's entering is in play in the intersubjective version of the story, but not in the version in which the doctor acts alone? It seems to me that we don't. The relevant movement of mind may escape notice in

[34] As I point out in Sections 5.3.2 and 5.3.3, wanting* something in the absence of one of these components is sometimes thought of as a case of "not really wanting". Schueler's desires proper may then be cases of wanting* for which this characterisation would be inappropriate.

introspection. Actually, cases of reluctantly doing what you believe you should do seem to me to be excellent examples in which thoughts of the form “OK, let’s get on with it”, or similar, are introspectively accessible. The cases in which congruent optative states are least obtrusive[35] are those in which their presence seems least controversial – in virtuous action in which the agent unhesitatingly wants to do what is required.

Perhaps, however, the mere entailment theorist isn’t prepared to take at face value the apparent datum that the doctor’s request expresses a want*. Perhaps it might be claimed that the want* here has no independent existence apart from its embodiment in the request to the receptionist. It might be argued that, because expressing a “desire (want, wish)” of the speaker is the sincerity condition of any request (Searle 1979, 4), the connection between attitude and verbal action is indeed no more than a logical matter. After all, a request counts as expressing a want* even where it is insincere. However, precisely this point makes explicit that the concept of expression at work in such speech-act theoretical contexts is to be distinguished from the one we are working with here (cf. Sects. 4.1 and 4.2). The distinction between sincere and insincere requests depends on the former being given voice to in order to realise an attitude that is missing in the latter case. So the presence of an optative attitude with the same content as that of the request is not a matter of mere logical necessity. What is logically necessary is that the doctor at least purports to express a want*. What decides whether she does more than merely purporting is the presence or absence of an independently existent want* concerning the movements of the pseudo-patient. Merely purporting, on the other hand, would depend on the presence of an independently existing want* with a different content, perhaps concerning her appearing to fulfil her professional code of conduct. The doctor might say “Let him come in” in order to keep up appearances, whilst hoping that the patient slips on a banana skin whilst crossing the waiting room (thus at least ceasing to be a pseudo-patient).

Before leaving the claims of the mere entailment view, I want to say a word or two about a group of linguistic devices whose existence seems to me indicative of the optative step that mere entailment denies. There are interesting features of the English language that are used to mark transitions from the mere cognitive recognition of requirements to their underwriting by the bearer of that assertoric state. One is the move from what linguists call the semi-modal verb “to have to” (“I have to go”) to the modal “must” (“I must go”). The modal variant, significantly, has no past or future form. This is plausibly related to the fact that predicting or reporting on one’s underwriting of some requirement is not equivalent to underwriting that requirement. “Have to” and “have got to”, in contrast, leave room for the speaker to distance herself from the requirement and are sometimes clear indications of such distance (Palmer 1987, 129f.). To say “I must φ” entails an acceptance of the corresponding situation-specific requirement, whereas “I have to φ” does not. A comparable phenomenon is the lack of optative involvement on the part of the

[35] Leaving aside habitual actions for the moment. On habitual actions, see Sections 5.1.5 and 9.5.3.

speaker permitted by the use of the modal “ought to”. This is reflected in the fact that, whereas someone might unproblematically say “I ought to φ, but I’m not going to”, replacing “ought to” with “must” yields an ungrammatical sentence (Palmer 1987, 132). It is true that these distinctions are subtle and are perhaps being eroded by the development of the English language. Nevertheless, we clearly do make provision for the attitudinal step prior to any corresponding action. It is unclear what these distinctions could have been designed to indicate were the mere entailment view to be true.

My suggestion, then, is that in acting to satisfy requirements, we need to accept their applicability in some particular situation and that the step of acceptance is optative in character. In underwriting an objective or intersubjective standard with which one is confronted, one sets the standard for oneself. ”Nothing but muddle (and boredom) comes from treating desire as a mental catch-all”, it has been claimed (Platts 1979, 256). No doubt the previous discussion could have been more entertaining. My contention, however, is that “muddle” results from *not* appreciating the generic character of wanting* and its tendency to form compounds with other attitudinal and non-attitudinal factors. Doing so permits a differentiated typology of types of optative attitudes, some of which would be labelled “desires” in everyday parlance, some of which wouldn’t, all of which however are candidates for explanatory functions where agents act because they are motivated to do so.

4.5 Appendix: Direction of Fit and the Internal Normativity of Attitudinising

The expressive explication of wanting* with the help of adapted Moorean sentences brings to the fore two essential points: firstly, the compound structure that appears to characterise all optative attitudes we are familiar with, that is, their constitution as modifications of the optative core. Secondly, it helps to clarify the converse internal relationship of wanting* and believing to standards. It is because of these constitutive differences that (conscious) conjoined contradictory wants* are neither unusual nor put their bearer under rational pressure to agglomerate.

To conclude this chapter, a word or two is in order about the relationship of the expressive analysis of the attitudes to the much-discussed notion of *direction of fit*. The term, coined by Austin in 1953, was first used in its present meaning by Searle in an article from 1975[36] and given currency within moral philosophy above all by Platts (1979), Smith (1987; 1994) and Dancy (1993; 2000). The distinction designated by the term was most influentially set out by Anscombe (1957).

The boundary within the attitudes between those with one direction of fit and those with the other corresponds exactly to the division between assertoric and optative attitudes. Both terminologies are attempts to capture the essence of the

[36] Published as chapter 1 of Searle 1979.

mode in which a proposition is formed. The assertoric mode has content-to-world, the optative mode world-to-content direction of fit.[37] Two points of importance for the expressive explication of the attitudes are easily made explicit if one discusses them in terms of the two "directions".

Firstly, it is constitutive of both kinds of attitudinal mode that they establish a relation between a representation and the way things are. Both believing and wanting* are thus a matter of correspondence ("fit") between reality and a representational content. Secondly, the distinguishing feature of the two ways of corresponding, a symmetrical relationship, is rendered by the metaphor of "direction". The metaphor articulates a feature of attitudinising that is both essential and puzzling. This is a kind of normativity internal to attitudinising.

The "direction" goes, according to one formulation, from one of the two relata, the one that is "responsible" for the correspondence being established, to the other (Searle 1983, 7f.). Another way in which the same point has been put is to say that there is one relatum that is to be seen as "mistaken" or "at fault" when no correspondence is given (Anscombe 1957, 56f.; Kenny 1975, 38; Searle 1979, 3f.). A further variation on the theme is to say that the starting point of the "directed" vector is the relatum that "ought to", "should" or "is supposed to" be changed on failure of correspondence.

All these variations on normative vocabulary are ways of bringing out the fact that attitudinising, in either of these two basis modes, essentially involves standards. In taking on either of these two basic attitudes, we either *take* the way things are *as* a standard for their representation or, in representing, *set* or underwrite subjective standards *for* the world to conform to. Now, the standards set in wanting* can – and in certain cases, must – be adjusted relative to those standards accepted in believing. This is basically what happens where people come to want things in the everyday sense of the term. Without such doxastically motivated adjustments of our wants*, the chances of them ever being satisfied would presumably be significantly smaller.

From this analysis there follows a conclusion that some might find uncomfortable: taking on either of the two basic attitudes involves taking a *normative* stand. This fact is concealed within the mainstream philosophy of mind by the focus on assertoric attitudes, whose modal component is characteristically unmarked. Clarity on this relativizes – although it doesn't dissolve – the apparent oddity of the fact that optative attitudinising basically involves setting up a kind of subjective "ought". It is a natural response to this suggestion to ask where, if this is correct, the peculiar subjective normativity of wanting* comes from. The answer is that it has the same source as that of believing: from the employment of standards that is essential to both basic types of attitude.

[37] The usual terminology, following Searle, is "mind-to-world" and "world-to-mind". As the direction-setting is constitutive of the two *modes*, whilst it is the attitudes' *content* that is set in relation to "the world", I adopt a slightly modified terminology. Making "mind" one pole of the relation is a little misleading, as the mode is itself a feature of the mental state.

The normativity internal to believing can, as Davidson has pointed out (Davidson 1982a, 104f.; 1997a, 128), be read off from the inseparability of the concept of belief from that of a *mistake*, that is, from the failure to attain the standard of correctly representing what is. That wants* inherently "aim at" fulfilment of standards set in wanting*, standards that can also fail to be satisfied, is the converse form of normativity.

This means that, properly understood, wants* and beliefs are in the same normative boat. I can see no workable alternative to this claim that does not avoid it by changing the subject. To round off this chapter, I shall briefly review and reject two explications of direction of fit that attempt to avoid this conclusion, before discussing an asymmetry between the normative components of the two attitudes. An understanding of that asymmetry allows us to make sense of the claim developed in Section 2.6 that non-linguistic creatures may be bearers of motivational states in spite of not being believers.

4.5.1 Direction of Fit, Functionalist Style

Michael Smith (1987; 1994, 111ff.) has argued that the notion of direction of fit can itself be dealt with by means of a functionalist analysis. According to Smith, the difference between the two basic attitudes lies in the differing ways in which they interact with the perception of the non-realisation of their content: if a person perceives that non-*p*, a belief that *p* in the same person tends to dissolve, whereas a desire that *p* tends to endure and to dispose its bearer to bring *p* about (cf. Armstrong 1968, 155).

I have already argued against the claim that a disposition to realise *p* could be definitive of the want* that *p*. *OSI* cases, including pure spectator cases and wants* with intervention exclusion clauses (Sect. 3.1.2), along with cases of motivational inertia (Sect. 2.5.1), provide sufficient evidence against such a definition. The optative analysis and the characterisation in terms of direction of fit are, unlike the dispositional analysis, able to makes sense of all these phenomena. Smith, however, claims that differences in direction of fit are themselves analysable as differences in the proclivity of a mental state to dissolve or to endure on being confronted with a perception with a contradictory content to that of the pre-given attitude's content. Content-to-world fit is basically a *counterfactual dissolution proclivity*; world-to-content fit a *counterfactual endurance proclivity*. Smith's claim, then, is that there is a necessary identity between "a state with which the world must fit" (Smith 1994, 115) and a mental state that tends to endure in conjunction with the perception of its content's negation. There are at least three serious problems with the proposal.[38] Of these, it is the third that is decisive.

[38] Schueler quite rightly also objects to the implication that the only possible contents of the two basic attitudes are the potential contents of perceptual states (Schueler 2003, 34). Both he and Mele (2003a, 267) further cite *OSI* cases as counter-examples.

Firstly, the condition supposed to trigger the disposition is itself possessed of a direction of fit: one can be mistaken in one's perceptions – where one's perceptual representation of the world fails to match the way things are out there. Thus, according to Smith, content-to-world direction of fit characterises the kind of mental state that tends to dissolve when conjoined with another mental state with content-to-world direction of fit, but with a content that is the negation of the first. I have argued (Sect. 3.3.2) that the functionalist thesis of the necessary interdefinition of mental terms is unsatisfactory. But even if one were to accept it, surely the definition of a type of state in terms of its reaction to another token of the same generic state type involves too tight a circle.

Secondly, the "tendency" of a desire that *p* to endure on the perception that non-*p* is certainly given, but clearly it is only a tendency and one which isn't only negated in a few marginal cases dreamed up by analytic philosophers. If I want it to be a sunny day when I get up in the morning and find that it is raining cats and dogs, my desire may well dissolve, rather than endure.[39] Similarly, I may want to achieve something and fail. Whether I continue to want to achieve it and have another go or whether I resign in the face of adversity will depend on a number of further contingent factors. All in all, some people tend to persevere in their wants*, others tend to resign fairly quickly. Note further that beliefs not only in the realisation of some wanted* *p*, but also in the unlikelihood or impossibility of its realisation frequently lead to the dissolution of the relevant want*. However, this need not be the case: sometimes wants* persist in the face of irrealisability beliefs, a persistence that in certain cases can cause the want's* bearer a great deal of trouble. The point here is that whether a want* dissolves or persists when its bearer comes to believe in the non-realisation or irrealisability of its content is a contingent matter and thus has nothing to do with the attitude's direction of fit.

The third and main problem with Smith's construal of direction of fit is that, in providing a non-metaphorical and non-normative explication, it deprives us of anything that is recognisable as the concept it was supposed to be reconstructing. Why should the tendency of a representation not to disappear on being confronted with its negation constitute the *to-be-realised* character of that representation? What has an endurance disposition got to do with the *pro*-feature that an analysis should be explaining? The fact that something won't go away may lead us to get used to it, but its mere continued presence under certain conditions is quite simply a different kind of property to its optative framing. The optative mode is essentially the way in which a content is represented; what is likely to happen to a representation is something completely different. Moreover, the difference between a tendency to go out of

[39] Nick Zangwill has advanced a "normative functionalist" view that, he claims, allows Smith's proposal to be corrected by replacing talk of dispositions with that of pro tanto rationality: a desire that *p* is thus an attitude whose endurance in the face of the perceptual experience of non-*p* would be pro tanto rational (Zangwill 1998, 196). But why should, independently of reflection on the probabilities of the desire's realisation and attendant frustration or satisfaction, there necessarily be a question of the rationality of the desire's endurance at all? The most one can say here is that, in contrast to the case of belief, it would pro tanto not be irrational for the desire to endure.

existence and a tendency to endure is a gradual matter. But the difference between believing something and wanting* it is anything but gradual. Finally, there is no good reason why the fact that a certain representation is under the circumstances particularly durable should make it apt to bring about the various – perceptual, emotional, and secondarily motivational – components of the optative syndrome.

Thus, not only is Smith's reconstruction of direction of fit *circular* and not only are the conditions it specifies *unnecessary* (two points that conceptual functionalists are committed to not worrying about). Above all, the proposal provides no clarity as to the normative character of the concept. Smith's proposal does not dissolve the disconcerting, apparently irreducible normativity at the core of direction of fit. Rather, his proposal simply side-steps it.

4.5.2 Direction of Fit as a Higher-Order Attitudinal Property

In a paper that contains a penetrating discussion of the various conceptions of direction of fit, Lloyd Humberstone (1992, 71ff.) goes on to suggest that we understand the concept in terms of what he calls "controlling background intentions". These are higher-order conditional intentions, whose contents relate the instantiation of some proposition to the intention's bearer being the bearer of a first-order attitude: in the case of believing, the proposition's non-instantiation is the condition for the non-possession of the first-order attitude; in the case of wanting*, the possession of the first-order attitude is the condition for the instantiation of the proposition. In other words, a belief is an attitude whose bearer is necessarily the bearer of a higher order-intention not to have the first-order attitude toward some proposition should that proposition not be instantiated in the world. And a want* that something be the case is an attitude whose bearer is necessarily the bearer of a second-order intention that the content of the first-order attitude be instantiated whenever he is the bearer of an attitude of that type.

The motivation behind the proposal is understandable: it is an attempt to replace the normative core of the two basic attitudes with a feature that is less metaphysically problematic. The main problem with the proposal ought to be equally obvious: it simply transposes the puzzle to a higher level. The normative character of the standards at work in both attitudes is seen as deriving from the optative mode of the second-order attitude in whose content they are represented. In both cases, the higher-order state specifies that something be or not be the case under certain conditions.

Once it is clear that the suggestion makes no contribution to an analysis of the normative core of the basic attitudes, the rationale for accepting a higher-order explication of a first-order attitude vanishes. Intending has world-to-content direction of fit and so its use in the analysis, like that of perception in the functionalist suggestion, is open to the charge of circularity. Moreover, as an attitude with conditional content is more complex than a non-conditional attitude, there is something bizarre about attempting to analyse the latter by means of the former.

Finally, intentions are particularly unsuited to doing the job assigned to them here.[40] For one thing, they are themselves, as I will argue in Part II of this study, more complex attitudes than generic wants*: whether one follows the analysis I will be proposing or not, it seems intuitively clear that intentions, like everyday wants, are attitudes with the world-to-content direction of fit plus more. Part of what that more consists in is the "commitment" of the intention's bearer to realise its content. But for this reason, because wanting* entails neither the commitment component nor the reference to the content's realisation by the attitude's bearer, intention cannot help here at all. The use of an attitude with these components suggests a conflation of optative and imperative understandings of direction of fit (Sect. 4.2.1): wanting* *p* does not entail intending, or even wanting*, to bring *p* about.

4.5.3 Normativity, Subjective and Objective

The functionalist and higher-order attempts to analyse the normative component at the core of the direction of fit metaphor are two time-honoured ways of responding to philosophically puzzling phenomena: either to act as if they aren't there or to have recourse to the phenomena themselves in attempting to explicate them. Circularity may be the more obvious mistake, but failing to hit the target ensures that formal correctness is irrelevant.

In effect, however, both suggestions narrow the field of wanting* down without any particular justification. A theory of wanting* has to be able to make sense of a large range of heterogeneous cases. At the centre are paradigmatic cases, in which taking on the attitude is accompanied by positive hedonic expectations and a level of concurrent discomfort, leads to action, to the subsequent fulfilment of those expectations, to positive hedonic effects and to the dissolution of the attitude. These can be labelled *effective appetitive wants**, most clearly given where the literal appetites lead to their own satisfaction. Alongside effective appetitive wanting*, there are many cases of optative attitudinising in which one or more of the typical effects of the attitude is or are absent. Motivation derived from assent to practical requirements may be independent of congruent hedonic symptoms or even be given in spite of affective opposition. *OSI* cases, on the other hand, usually entertain strong causal connections to hedonic experience without producing primary motivational effects, i.e. actions directed at the realisation of the want* contents. Other components of the optative syndrome are most noticeable in *OSI* cases where no action occurs. However, particularly where action is delayed or prevented, secondary epistemic and expressive want* formation, fantasy, perceptual salience and emotional reactions are all characteristic symptoms. All of these phenomena can be made sense of in the light of the expressive explication of

[40] Humberstone (1992, 75) more or less admits this.

wanting* as a generic optative attitude, that is, as essentially a matter of the setting of one's own subjective standards, against which reality is measured.

In Section 2.6, I rejected Davidson's claim that the normative dimension of belief entails the impossibility of attitudinising for non-linguistic creatures. I also rejected his claim that the two complementary basic attitudinal modes necessarily emerge together. Instead, I argued that, as the one-way triangulation case of Brian Hare's subordinate chimpanzee shows, creatures can play host to motivational states even if they are not bearers of full-blown beliefs. The one-way triangulating chimpanzee doesn't require the distinction between appearance and reality, but simply the distinction between representations from different perspectives. We should understand the subordinate as behaving as a result of his motivational states because his behaviour is flexibly modifiable depending on what he sees the dominant as seeing or having seen. Nevertheless, what he sees the dominant as seeing doesn't qualify as the content of a belief if the subordinate has no conception of objective reality, against which the various representations can be measured. And the behaviour of the chimpanzee can be explained as motivated without having recourse to such a conception.

We are now in a position to explain that conclusion on the basis of the analyses of believing and wanting*. The two basic attitudes involve their bearers taking on complementary relations to standards. However, the standards at work in the two kinds of attitude differ in how demanding they are. Whereas the standard applied in wanting* is subjective in the strict sense that it is set by its bearer, the standard aimed at in believing can be made no sense of without some procedure by which the bearer's perspective is transcended. Moreover, the standard cannot be simply some other perspective that is equally relative to the position of its bearer. What is so demanding about the standard of the way things are is its independence from any perspective or, put another way, it's being the common object of varying perspectives. Hanjo Glock has suggested that the conception of a perspective-independent reality may be attained independently by particular individuals as a result of their moving round objects and perceiving the different appearances of what they come to take as the same thing (Glock 2000, 58). If this were correct, no reciprocal processes of attention direction and mutual adjustment would be necessary. This strikes me as implausible – for empirical, not for priori reasons. I suspect that Davidson is in fact right that a conception of objectivity can only come into being for creatures that have ways of stabilising a common view of the way things are. Only if such a view is sturdy enough will it obtrude into an animal's thoughts even as things appear in a different light. And that sturdiness, Davidson assumes, only comes about through interactive processes which themselves build up self-evident fixed points of orientation. Finally, Davidson assumes that the only sort of interactive practice that enables the sedimentation of such self-evident forms of common orientation is language.

There are three points I want to hold onto here. First, Davidson's story identifies correctly the more demanding feature of belief's normativity relative to that of wanting*, but illegitimately assumes that the holism of the mental forces us to extend claims about the conditions of believing to the conditions of wanting*. Second, triangulation is a plausible candidate for an empirical – phylo- and ontogenetic – structure necessary for that feature, viz. the distinction between appearance and reality, to come into being. As I argued earlier, however, triangulation itself may come in less than fully fledged forms and may, in a one-way variant, justify the attribution of wants*, but not of beliefs in the full sense. Third, it is an open question whether, if triangulation should indeed be the only way to achieve the normative distinction essential to belief, language is the only way that triangulation structures can be sufficiently stabilised for the idea of a perspective-independent reality to take root.[41]

Evidence that a creature is processing information in the light of the normativity essential to belief is, as Davidson has argued, provided by its working with the category of a mistake. Language easily allows us to check for this. There is anecdotal evidence from the ape language projects that encultured chimpanzees can correct their use of symbols after observing a conspecific react to that symbol use in a way they had apparently not wanted him to (Savage-Rumbaugh and Lewin 1994, 82). A methodologically key question here is whether we could have sufficient non-linguistic behavioural evidence for a creature taking itself to have made an error. Colin Allen reports experiments with pigs trained to respond differentially to pairs of objects depending on whether they differed in shape, size or colour. The pigs, whose choices had a near-90 % correctness level, sometimes "backed out" of a choice they had made before receiving feedback, such backout behaviour generally only occurring where the choice was incorrect. Allen claims the experiments demonstrate at least that there can be non-linguistic evidence of "endogenous error detection capacity" (Allen 1999, 38).[42]

A further kind of evidence for creatures possessing the concept of error is plausibly provided by cases of deception, in particular cases of counter-deception. In a case of the latter kind reported by Whiten and Byrne, a chimpanzee with access to food showed no interest claiming it in the presence of a dominant, who himself left the scene only to hide behind a tree, observe the subordinate claiming the food and return to relieve the him of his prize (Whiten and Byrne 1988, 220). If the subordinate's behaviour could be explained by simple operant conditioning, having learned that it's rewarding to sit still when near food if a dominant conspecific is present, this is not true of the behaviour of the dominant. Peeking out from behind

[41] Conceptions of the attitudes as defined by their linguistic expression may also appear particularly suggestive of lingualism. For the analogical conception's internalisation variant (Aune 1967, 218ff.), this is an obvious consequence of the logical priority of language over mentality.

[42] He doesn't commit himself in the question of whether the pigs in the experiment genuinely show the capacity to be guided by understanding mistakes. Disappointingly, it seems the empirical work he draws on here was never published. Allen and Beckoff have also suggested it may be possible to gain behavioural evidence that an organism "is subject to an illusion yet can make choices that depend on rejecting the illusory properties" (Allen and Beckoff 1997, 152). If a non-linguistic animal could do this, that would show it to have a capacity not only to perceive, but also to have beliefs about the same thing. I am not aware of any experiments that have demonstrated these abilities.

a tree is, as Byrne remarks, not something a chimpanzee would coincidently do and then find rewarding (Byrne 1995, 134). The only plausible explanation seems to be that he suspects that he is being led into error and takes it that leading the deceiver into error is the best way to counter any such attempt. If this is correct, it means that the dominant here must have some idea that there are ways things are that both his own and his competitor's beliefs might fail to represent correctly.

The actions of the counter-deceptive dominant are, so it seems, to be explained by wants* that are flexibly modified by fully fledged beliefs. I have argued that want* modification is also possible in the light of subdoxastic representations, as might be the case in one-way triangulation. If there are indeed agents whose behaviour is guided by such attitudinal constellations of wants* and sub-doxastic representations, these would be agents who couldn't *check* to see whether their wants* have in fact been fulfilled. Their behavioural dispositions to change response once the organism registers the satisfaction of subjective standards set in wanting* would ensure that motivational states fulfil their functional role: if motivating standards were set in the absence of any mechanisms to orientate behaviour in line with those standards and prevent continued striving to attain what has already been attained, the bearer of those standards would be massively dysfunctional. But discrimination between *a* and *b* is not sufficient for the belief that *a* is not *b*, because the capacity to categorise stimuli into perceptual equivalence classes is a characteristic of most organisms (Allen 1999, 36). Pace Searle, an animal's being able to "tell whether its desire is satisfied or frustrated" (Searle 1994b, 212) need not be a matter of belief. The creature in question doesn't have to have any thoughts concerning want* satisfaction, it just needs to get on with behaviour appropriate to the relevant information registered, where subjective standards determine relevance.

Although, then, wanting's* entailing the possibility of non-satisfaction parallels belief's entailing the possibility of mistake and although this means that both attitudes are normative at core,[43] the less demanding character of the standards set in wanting* suggests that wanting* may be present in species incapable of belief and developmentally prior to believing in humans. Studies by the developmental psychologists Bartsch and Wellman also provide evidence that this could correspond to the normal ontogenesis of children's mental capacities. Bartsch and Wellman report that talk of "desires" precedes talk of beliefs on average by 7 months (Bartsch and Wellman 1995, 96). Using the term "desire" to summarise uses of the words "wish", "hope", "care (about)", "afraid (that)" and above all "want" (but not "desire"), they found that children begin to use these words by 1½ – 2 years (1995, 93), whereas reference to "thoughts" and "beliefs" begins infrequently at the beginning of their third year, only becoming established just before their third

[43] Zangwill (cf. note 38) also claims that the attitudes "have a normative essence" (Zangwill 1998, 190). However, according to Zangwill's normative functionalist conception, the standards at work in wanting and believing are equally objective. He argues that the normativity relevant to desire is that of standards that constitutively relate tokens of the attitude to action and to other attitudes. On normative functionalism, see below, Section 7.3.

birthday (1995, 63f.). The ability manifested in passing the false belief task – the ability to ascribe someone else a belief about some object with a content incompatible with that of one's own belief on the matter – is frequently taken to be necessary and sufficient for the possession of the full-blown concept of belief (Wellman et al. 2001). There is also widespread agreement that this capacity isn't acquired before the middle of children's third year. However, there is evidence that children as young as 18 months are able to ascribe others "desires" that differ in content from their own (Repacholi and Gopnik 1997). Indeed, Henry Wellman has proposed that we see children between the ages of two and three as thinking in terms of "a simple desire psychology", a precursor of "belief-desire psychology", which only comes into being around about the age of three (Wellman 1992, 209ff.).

Now, the date at which children begin using certain mental terms is not necessarily the date at which they acquire the capacities to which those terms refer. So these data from theory of mind experiments would allow for different explanations. The data nevertheless fit nicely with the conception I have been elaborating. If I am right, it may well be that want* talk is more readily available to small children simply because wanting* is less demanding than believing.[44]

[44] Hannes Rakoczy has offered a different explanation. According to Rakoczy's proposal, beliefs that don't match what (one takes to be) the way the world is are more difficult to ascribe because doing so requires inhibiting the default presupposition that beliefs one is ascribing are true. Such inhibition in turn requires the development of executive functions, capacities for which come into being around the time of success at the false belief task (Rakoczy 2010). Rakoczy's explanation is compatible with the construal of wanting* developed here. I have suggested that a time lag between the capacities for ascription of wants* and beliefs may be down to the difficulties involved in acquiring the concept of truth, whereas Rakoczy proposes that the problem lies in the inhibition of its application.

Chapter 5
Wanting*, Consciousness and Affect

Galen Strawson (1994, 283) has argued that "desiring" must be understood as entertaining an "internal" or "constitutive" relation to one of three other concepts. The first is that of behavioural dispositions. He rejects this possibility, as I have done (Sect. 3.3). The remaining alternatives he sees as the connection either to hedonic dispositions or to the capacity for what he calls "conscious episodes of desiring or wanting". I agree that a theory of wanting* needs to tie the concept closely to one of these other concepts. However, unlike Strawson, I believe that it is his third alternative, and not his second, that is correct. In this final chapter on the generic concept of wanting*, I shall attempt to vindicate this claim.

The first two sections discuss the intimate relationship between optative attitudinising and consciousness. Wanting's* linguistic expression is the articulation of a structure that is paradigmatically conscious in character: we could make scant sense of the idea of expressive articulation if we had no conscious experience of taking the stance conveyed by the relevant utterance. However, as I pointed out in Section 3.4, it is a central requirement on a theory of wanting* that it be able to make sense of the fact that wants* are characteristically, but not necessarily directly accessible to their bearers in consciousness. Thus, to take up Strawson's terms, the relationship between wanting* and consciousness may, in some sense, be accurately described as "internal"; however, it cannot be "constitutive" in the sense of necessary and sufficient in individual cases.

In Section 5.1, I discuss one way in which the relationship between the two concepts might be made particularly tight, whilst still allowing for wants* that are not conscious. The conception, which requires for every want* the conscious tokening of an optative thought at, or prior to the time of want* possession, is rejected as unable to adequately explain two kinds of phenomena. In simple cases of what I call "subintentional action" and in cases in which considerations of self-esteem seem to prevent an agent's access to his genuine motivation, we need,

N. Roughley, *Wanting and Intending*, Philosophical Studies Series 123,
DOI 10.1007/978-94-017-7387-4_5

I argue, to postulate non-conscious variants of those optative postures we are familiar with in our conscious mental life. I support this conclusion with a look at some empirical work on goal priming.

Although there is no necessity that all individual wants* have involved some conscious occurrence, conscious optative thoughts are, I argue, sufficient for wanting*. In Section 5.2, I attempt to show how this claim can be reconciled with the possibility of meaningfully denying that someone "really wants" something they sincerely profess to want. A number of different phenomena can be denoted by this expression. Moreover, their investigation puts us on the trail of various forms of want* strength that should be distinguished. Among these are forms of hedonic strength.

In the final section of the chapter (Sect. 5.3), I discuss the different relationships that hedonic experience can entertain to wanting*. That these are two distinct states has been presupposed in previous sections of this study, particularly in Sections 3.2.2, 4.2.1 and 4.4.2, where I made use of the idea that wanting* and affect tend to co-occur in particularly salient constellations picked out by the terms of natural languages. Obviously, for wanting* and affect to be able to fuse into such compounds, there must be a clear sense in which they are different kinds of states. The characterisation of wanting* as an attitudinal structure articulated by expressions such as "Let it be the case that *p*" hardly provides a plausible candidate for an explication of pleasure.[1] In Section 5.3, I argue, conversely, that no use of the notion of affect can account for the phenomena that a theory of wanting needs to cover.

5.1 Conscious Occurrentism

It is by and large easy to know what you want*. This knowledge is characteristically non-deductive: most of the time, we don't have to observe our behaviour, attend to features of our experience or be informed by external observers in order to be able to infer what it is we are after. The typically unmediated accessibility of the contents of our own wants* consists in our ability to give them expressive articulation, where that articulation involves an – audible or inaudible – linguistic performance. This capacity, in turn, is explained by the fact that wanting* is a matter of taking on an attitudinal stand structurally analogous to that of an optative utterance.

Now, the idea of expressive explication makes no explicit reference to consciousness. There does, however, seem to be a very close connection. It is

[1] Not surprisingly, in view of the intimate relations between wanting* and hedonics, there have not only been attempts to define the former in terms of the latter, but also to define the latter in terms of the former. The fact that *both* attempts have a certain initial plausibility is in itself a prima facie indication that both phenomena are fairly robust and are thus unlikely to fall victim to reduction at the hands of the other. On unsuccessful attempts to reduce pleasure to "desire", see my articles referred to in Chapter 3, note 17.

plausibly a necessary feature of conscious attitudes that they available for expressive articulation.[2] Conversely, if in a psychoanalytic scenario someone is driven by an unconscious want* to perform certain actions in a way that is intransparent to them, they may be said to be expressing the want* in their action. Moreover, it may be possible for them to self-ascribe the want* if they believe what their psychoanalyst tells them. However, neither the want's* behavioural expression nor its linguistic formulation count as its expressive articulation. This relation, constitutive of unmediated access, can only be instantiated where the relevant attitude is conscious.

It follows that expressing an attitude by articulating it necessarily involves consciousness – either because the linguistic articulation involves reference to an antecedently existing conscious state or because the conscious attitude is generated in the course of the utterance. I argued in Section 4.2.1 that there is no optative attitudinising without some perspective from which the relevant standard is set. This perspective is aired publicly in linguistic utterances that articulate the attitude. And the perspective to which that attitudinal stand contributes can only become clear to its bearer if he rehearses it consciously. For this reason, conscious wanting* is wanting's* paradigmatic form: without it we could make little sense of the relation by means of which we get a conceptual grip on what it is to have such attitudes.

Nevertheless, the occurrence of a conscious optative thought at any time *t* clearly cannot be a necessary condition of wanting* at *t*: we don't want* only those things we are consciously thinking about. Indeed, most of the wants* of a person are not consciously occurrent at any particular time. A both empirically and experientially plausible conception of wanting* should therefore explain how this can be both true and compatible with the fact that our understanding of what it is to want* something grounds in the first-person perspective, from which we have expressive access to our attitudes.

Are there any connections between individual episodes of wanting* and consciousness? Searle has claimed that, necessarily, an attitude is at least potentially conscious (the "connection principle"), a characterisation whose modal component has understandably appeared to require substantial spelling out.[3] In spite of the uproar that Searle's claim caused, particularly among the cognitive scientists who it was primarily aimed at, the principle is compatible with *none* of an agent's attitudes actually being conscious. I therefore want to discuss a principle that is considerably stronger, a principle that, if it were true, would establish a substantial relationship between conscious and non-conscious wanting*. Although I shall be arguing that

[2] Pretty much this claim is argued for convincingly by Finkelstein (Finkelstein 1999, 91ff.; 2003, 119ff.). The difference between his claim and mine is that Finkelstein isn't working with a conception of expressive articulation, but with a notion of expression by self-ascription.

[3] Searle argues for the principle in a number of texts (Searle 1990; 1992, 155ff.; 1994a). Worries about the sense of "potentiality" loom large in the comments Searle received when he first aired it in the 1990 article in *Behavioral and Brain Sciences*.

the principle is false, I think that the reasons why this is so shed some light on the status of conscious relative to non-conscious wanting*, a question generally ignored in standard discussions.

5.1.1 CO

Anthony Kenny suggests a position on the relationship between wanting* and consciousness that I shall label "conscious occurrentism" (*CO*). Wanting* involves, according to Kenny, either "saying in one's heart" (1963, 218ff.) or "having said in one's heart" (1974, 32) "Would that *p* were the case".[4] If we were to take "saying in one's heart" to entail having a conscious thought,[5] we could fashion the following disjunctive position out of his claim:

(CO)
X wants* *p* at *t* only if either:

1. a thought in the optative mode with content *p* is consciously occurring to *X* at *t*, or
2. a thought in the optative mode with content *p* has consciously occurred to *X* at some time t_{-n}.

CO fulfils the requirement of giving conscious optative thoughts a conceptual role to play whilst allowing for their absence at a particular time. If the relevant thought is not occurring now, it must, according to *CO*, have occurred at some earlier point in time.

The requirement of characteristically non-deductive, yet possibly deductive access is also satisfied. Where someone has forgotten or "repressed" having had the relevant thought – perhaps because of some process adequately described in Freudian terms as "repression" – they may have become incapable of giving sincere expression to a want* to which they are nevertheless still playing host. In such cases, wants* could acquire an external guise, only being accessible via their effects on our perceptions, thoughts, feelings or actions. In such cases, manifestations of the

[4] For Kenny, this is only true of those "wants" he labels "volitions", which, in contrast to "desires", require language (Kenny 1975, 49ff.; 1989, 35ff.). This distinction, which is modelled on Aristotle's distinction between "boulesis" and "epithumiai" (cf. Sect. 1.3), is supposed to solve the problem of how attitudes identified by their linguistic expression could be ascribed to animals. The answer is that they cannot, so that "desires", which are ascribable to both humans and non-linguistic animals, turn out to be quite different kinds of states to other wants. Kenny doesn't reveal what it is in virtue of which volitions and desires nevertheless both count as wants.

[5] Kenny doesn't. He talks in various places of intention formation as consisting in a "mental utterance" (Kenny 1963, 216; 1975, 31), which, even when distinguished, as it must be, from its – possibly subvocal – linguistic expression, sounds very much like a mental event. However, he explicitly denies that such an "utterance" might be necessary for a "volition": "to say 'He has said in his heart ›Let me do *A*‹' is simply to say 'He is in a mental state expressible by ›Let me do *A*‹'(Kenny 1975, 35). Talk of "saying in one's heart" would then do no substantial philosophical work. Perhaps it is for this reason that it is missing in Kenny's later *Metaphysics of Mind* (1989).

optative syndrome take on a role that, although not criterial, provides the person with good reasons to suspect that she is the bearer of some (at least momentarily) otherwise inaccessible want*.

CO would also provide a way of establishing an important distinction in theories of attitudinising: that between being disposed to become the bearer of an attitude and being its dispositional bearer. For instance, a person who has grown up in Scotland and suffers from rheumatism may be disposed to develop the desire to live in southern Italy, because the climatic conditions would be conducive to his overall well-being. As he has never thought of this solution and as no representation of his life in Italy determines any forms of his – mental or physical – behaviour, it would be wrong to see him as wanting* to live in Italy. However, were someone to point out the benefits to him, he might well develop the desire to do so. The criterion for the difference proposed by *CO* is quite simply that the relevant optative thought have been consciously tokened. Where this has happened, we can see why he might suddenly begin to manifest the optative syndrome, i.e. agential and non-agential symptoms of having had the thought.

Finally, *CO* could also be seen as helping to make sense of agents optative relations to the components of complex sequential actions such as making a cup of tea or driving to work. All of these include component actions: taking an individual step, turning the tap on and putting your foot on the clutch. It would obviously be phenomenologically inaccurate to claim that each of these component actions requires a separate conscious thought for its instigation. The conscious occurrentist could claim that the motivated character of these behavioural segments is secured genetically by the fact that, when the agent learned to do these things, the relevant optative thoughts did occur consciously. The corresponding optative structures can then be seen as having been "installed" in the agent in such a way that they only need triggering by the want* to perform the complex action. These would be genuine *dispositional wants**, as opposed to the mere *disposition to want** that characterises the Scot in the previous paragraph.[6]

5.1.2 Insufficiency

CO only claims to provide necessary conditions for someone wanting* something at a particular time. The fact that it doesn't give us sufficient conditions is thus obviously in itself no objection. Perhaps, one might think, we simply need to supplement *CO* with some further requirements in order to get a complete description of the relationship between consciousness and wanting*. These would have to deal with the question of want* persistence. Quite simply, many of our

[6] Robert Audi argues convincingly for the importance of this distinction in relation to belief in his (1994). In his (2003) 31ff. (but contrast Mele 1995b, 396f.), Mele makes a comparable distinction between "standing desires" and "dispositions to desire concurrently".

wants* – whatever role consciousness may play in our acquiring them – are only short-term features of our psychology. Wanting* something at t_1 is obviously, and fortunately, no guarantee of still wanting* it at t_2. Otherwise, wanting* would be cumulative affair which left us permanently saddled with all the wants* we have ever had.

Non-persistence of wants* can result from various different types of process, for instance: (a) from internal physiological or external changes in the world leading to their dissolution, (b) from their becoming irrelevant, (c) from their being satisfied, (d) from their being permanently forgotten or (e) from a change of mind.

A word on each of these: (a) someone who at t_1 is desperate for an ice cream may, in spite of not getting what he longs for, find that he has no such desire at t_2, either because the weather has gone cold or because he has developed some sort of nauseous illness. (b) A person who would like to go to a particular museum whilst he is holidaying in some city may simply drop the attitude if he leaves without having got round to realising it. (c) Although, as I argued in Sections 4.3.3 and 4.5.1, the causal connection between want* satisfaction and the (at least temporary) disappearance of the relevant want* is not a matter of conceptual necessity, the realisation of a desire's content is obviously a typical cause of its dissolution. (d) Although forgetting a want* need by no means entail its dissolution, longings or intentions, for instance, do indeed sometimes just dissolve as a result of not being consciously tokened over a longer period. Clearly, cases of type (d) may also be cases of type (a) or (b); and (a) and (b) may not always be easily distinguished.

(e) Finally, an optative change of mind is the replacement of some want* by another determinate optative attitude concerning the proposition in question. We usually only talk of "change of mind" where the relevant replacement has come about as a result of conscious thought, cases frequently characterised by an active phenomenology. The result of this process can be plausibly interpreted as the explicit renouncement of "saying in one's heart": "Let it be the case that *p*", replacing *p* with *q*. Note though, that as (a) to (d) show, even if there is a certain plausibility to the claim that wanting* is necessarily inaugurated consciously, there is none at all to the claim that it must be consciously brought to an end.[7]

Someone might be tempted to try and expand *CO* so to also cover sufficient conditions, listing the exclusion of (a) to (e) – and possibly other defeating conditions. This would not only be messy. In order to look like an analysis, it would need to specify a criterion whose fulfilment means that the want* has, in any of the types of cases listed, dissolved. Success in providing such a criterion is, however, likely to make *CO* redundant. Moreover, the more interesting problem with *CO*, is not its insufficiency, but the evidence for its non-necessity.

[7] Kenny defends the sufficiency of the (true) utterance "I have said in my heart 'Let me do *A*'" by appealing to the fact that British English speakers only employ the perfect aspect where the content of the verb remains relevant at the time of utterance (Kenny 1975, 32). Linguistically, this ignores the fact that the relevance up to the time of utterance may simply be a matter of perspective on the period of time during which the event is located (Palmer 1987, 47ff.). More importantly, this explanation assumes that we have an independent criterion that allows us to assess whether the saying-in-one's-heart idiom can still be relevantly used.

5.1.3 Subintentional Action

The first kind of evidence involves extremely banal cases of behaviour which we have good grounds to see as caused by wants* in spite of the absence of conscious thoughts concerning their contents. These are cases of what I shall call *subintentional action*. In these cases people do things because they want* to do them, but without doing them – or anything else in the course of which they end up doing them – "on purpose" or intentionally.[8]

A person sitting at their desk working might, as we would say, "do" any number of things without being aware of doing them: stretch their legs out, scratch their head, fold and unfold their arms, brush a piece of fluff to the floor, etc.[9] These are forms of behaviour in which we are more or less continually involved. The decisive point is that at least some of these are best understood as being motivated by wants* of the person. Where someone, after sitting in one position for some time, stretches out their legs under their desk, the explanation for this movement is that the first position was starting to get uncomfortable, which led them to want* to change position, as a result of which they did. And all this seems perfectly correct, independently of whether the subject had a conscious thought along the lines of "Would that this state of discomfort were to end".

In response to such cases, there are two moves which would be open to a conscious occurrentist. The first would be simply to deny that we are dealing with wants* in such examples. The basis for want* ascription seems relatively thin here; it is, so it may seem, no more than a certain bodily movement.[10] Should we then perhaps see cases such as these as nothing more than the mere triggering of bodily dispositions? Their phenomenology certainly does not suggest their assimilation to simple reflexes. They don't appear to be direct responses to specific stimuli triggered independently of whether the agent wants those responses to take place or not – as when something causes a person's knee to jerk or their eyelid to blink. Granted, reflexes can become subject to control by wants* where people, for instance, wearers of contact lenses, are intent on suppressing them. But if someone becomes conscious of her leg stretching and decides to stop it, she has no sense of suppressing the functioning of an automatic physiological mechanism.[11]

[8] If such cases exist, then Davidson's suggestion that all action is necessarily intentional under some description (1971, 46ff.) is false.

[9] Kent Bach (1978, 363ff.) calls such pieces of behaviour "minimal actions". However, he postulates for their explanation sub-attitudinal representational states that are different in nature to, not just non-conscious variants of, the conscious attitudes that generally cause intentional action. Robert Audi (1986, 28f.) also suggests tentatively that we might see such movements as non-intentional actions.

[10] Ironically, these borderline examples of action are the cases in which the major premise of the Logical Connection Argument (Sect. 2.5.3) – "our *entire* criterion for saying what he wanted ... to do, is what he in fact did" (Taylor 1966, 52; my emphasis) – is most plausibly fulfilled.

[11] Gottfried Seebass has suggested in conversation that such movements are analogous to our tossing and turning in our sleep, movements made without our wanting* to make them. But is

The reasons for explaining many such examples of leg-stretching by the causal efficacy of instrumental desires to stretch one's legs lie in the *analogies* and the *continuities* of these cases with cases where conscious wanting* is at work. Both points become manifest when we consider what happens when the movement in question encounters resistance. There are, so it seems, basically three things that can happen: (i) the movement can simply come to a halt; (ii) the behavioural process can circumnavigate the obstacle through a change of direction or overcome it by an increase in pressure; or (iii), particularly if neither of these – non-conscious – strategies is successful, conscious attention might click in, in the form of thoughts like "What on earth is that under my desk?" In the second case, the claim that the movement was goal-directed is given support by modifications which themselves appear teleological. In the third case, where consciousness clicks in, it will tend to immediately set in with means-ends reflection: what's in the way? Can I push it to one side? etc. The person will not under normal circumstances begin by asking herself why her legs were moving in the direction they were at the moment of contact. The analogy with conscious wanting* consists in what looks very much like the non-conscious search for alternative means to attain a desired end. The continuity consists in the fact that conscious attention may simply carry on where non-conscious striving appears to leave off.

Note further that, in contrast to movements caused by reflexes or by the triggering of other bodily dispositions, the bodily movements in question here are not movements that we would ascribe to a *body part*, as opposed to the *person* herself. Whereas we say that a person's eye blinks, her knee jerks or her face twitches, we say that *she* scratches her back or adjusts her sitting posture.

Perhaps the most plausible strategy open to a conscious occurrentist for these kind of cases would not be to deny that we are dealing with wants*, but instead to insist that there *must have* been some corresponding conscious occurrence. Such an optative occurrence may have been barely registered because of being non-focally conscious. Certainly, not all consciousness of something involves focussing on that particular. The purview of consciousness can take in and individuate propositionally a considerable amount of data at any one moment. Certain things are attended to explicitly, others are registered without focal attention. This strategy gains some support from the fact that what plausibly triggers leg-stretching, back-scratching or eye-rubbing is a *sensation* of discomfort that must itself be consciously felt, however non-focally. As William James insisted (1890, 174f.), the only mode of existence of a feeling is that of being felt.

If the discomfort is registered consciously, must the want* that then leads to the relevant movement itself not have been consciously formed? I don't think so.

it really so clear that none of the movements we make in sleep might be caused by wants* to remove discomfort? Compare the movements you might make when lying wide awake, and then those made whilst slipping into half-sleep. There seem to be continuities here which suggest that, at least sometimes, the same kinds of representational processes may be at work in controlling the movement of our bodies during sleep as do so when we are conscious of what we are doing.

Take a person who finds himself scratching his back or is surprised to be told that he was rubbing his eyes a lot this morning. We cannot make sense of the relevant actions if we don't assume the agent felt some unpleasant sensation, if only vaguely and in a way that need leave no trace in conscious memory. If it is felt, then it must be at least minimally conscious. But it seems phenomenologically implausible to postulate a further conscious thought of the form "Let it be the case that I rub my eyes/scratch my back, in order to reduce this unpleasant sensation". You may actually find yourself rubbing your eyes and only be able to infer what must have led to your doing so because of your knowledge of what generally causes people, you in particular, to engage in eye-rubbing.

If this is correct, then there are actions that we perform without intentionally doing so, because the wants* that cause them are, at the moment of their efficacy, outside the purview of conscious access.[12] Introspective phenomenology is of course bound to be controversial. However, the unlikelihood of providing clear proofs in some area doesn't mean that there are no facts of the matter, even if the evidence that one can gather is bound to remain inconclusive. There is an obvious barrier to becoming aware of the fact that one's actions are being caused by wants* of which one is at that moment not conscious. Nevertheless, the picture that one acquires in retrospective reflection seems to me to offer considerable support for the claim that there are subintentional actions.

5.1.4 Motivated Want Inaccessibility*

The first group of examples that tell against *CO* ground in evidence that we sometimes develop and realise wants* for minor changes in our hedonic situation although our consciousness is completely preoccupied with other matters. The second kind of case has acquired a certain familiarity since Freud. Here, agents develop wants* without being aware of them because – so it seems – such awareness would endanger their self-esteem, that is, it would prevent them believing they are the sort of person they want* to be. In the first kind of case, it appears that consciousness is not engaged because we don't need it in order to get what we want*; in the second kind of case, consciousness seems to be avoided because we are motivated not to be aware of what it is that we want*. It is cases of this latter kind to which I now turn.

Two examples are furnished by Jeff and Harold. Jeff is a jealous husband who repeatedly rings home for a whole gamut of ostensible reasons, whilst his real motivation is the fear that his wife is unfaithful. Harold has suffered a career setback at the hands of a colleague, Betty, but tells his friends sincerely that he accepts it as part of a competitive situation and bears Betty no hard feelings. Nevertheless, when he hears that things are going wrong for her, he notices feelings of pleasure creeping

[12] In this point I am in agreement with Brandt and Kim (1963, 430f.).

over him. This leads him to conclude that he must have developed a desire that she suffer harm without having been aware of it. Jeff performs intentional actions, such as ringing up to discuss the shopping. However, so the example goes, this intentional action is also correctly describable as an attempt to glean information that might provide clues as to his wife's extra-marital activities – although Jeff is unaware that it falls under this description. Harold is of interest because his desire is causally manifest not in action, but in affect.

As before, a conscious occurrentist would have to either deny that the relevant causes are indeed wants* or insist that there must be a conscious optative thought with the relevant content somewhere along the line. The first line of defence is particularly unconvincing where various features of the optative syndrome occur together in such a way that they seem best explained by the standard common cause. When Harold acquires the belief that he wants Betty to suffer harm, he might remember such things as having noticed with unusual frequency the behaviour of others construable as threatening for Betty and having behaved in an agitated manner at a party where she was the centre of attention. Certainly, we are disposed to perceive, feel and behave in certain ways independently of what we want*. Where, however, such features cluster, their best explanation is likely to be in terms of an attitude we see as explaining them when it is consciously available for expressive articulation. This would appear to be confirmed in some cases of renewed self-examination: on the basis of the hedonic, perceptual and behavioural evidence, Harold may ask himself again what his optative attitude towards his colleague really is and find that he is in fact the bearer of a desire that she suffer harm after all.

Once again, the more promising line of defence of *CO* would be to insist that there *must have* been some conscious optative thought. According to Sartre (1943, 52f.), Jeff and Harold should be described as consciously desiring the relevant information or the colleague's harm, but as refusing to acknowledge to themselves that this is the case. For Sartre, so-called unconscious desiring is inauthenticity or "bad faith" in the face of desires one would rather not have. And there is a certain plausibility to his main argument for this conclusion: if the person is not conscious of the want*, how is he to recognise the danger it spells for his self-esteem? If the threat to the agent's desired self-image is to explain the want's* non-conscious character, then this seems to presuppose that the relevant optative thought have occurred consciously.

C.D. Broad has proposed a fairly simple two-step mechanism for the functioning of "bad faith" in such cases. The first step consists in our consciously and deliberately averting our attention from certain conscious desires, a step taken in order to preserve our self-esteem. At the same time, he suggests, we refuse to acknowledge that we are doing so, in order not to undermine the effect of averting our attention (1962, 366f.). The offending desires are thus "quite literally conscious". In a second step, such aversion of attention can become a habitual response to certain recurrent desires, a habit that may become so strong that it resists deliberate influence.

Broad's proposal raises several questions: are the agent's desires not to undermine the effect of averting her attention themselves conscious? If so, can the conclusion be avoided that the agent requires an unending set of desires not to attend to desires motivating attention aversion? Concerning Broad's second stage: how is the mechanism of habitual response supposed to function? Do we need to have attended to the offending desires before we recognize them as the ones from which we divert our attention habitually? Compare the fact that you seem only able to ignore someone, even habitually, once you have recognized him as the person you are supposed to be ignoring. The Sartre-Broad strategy for dealing with motivated want* inaccessibility looks fairly unpromising as an attempt to vindicate *CO* theory.

5.1.5 The Auto-Motive Model and Goal Priming

Perhaps, however, conscious occurrentism can look to support from an unexpected source. In the psychological literature on routine actions such as getting dressed and cleaning ones teeth, it is sometimes claimed that such actions are brought about as a result of non-conscious equivalents of conscious motivational attitudes (Bargh 1990). According to such conceptions, the presence of the non-conscious motivational states leads to the action being triggered "automatically" – i.e. without any relevant conscious thought – by features of the environment. The basic idea is that component actions in complex sequences are originally learned and carried out with conscious intent before habituation allows those attitudes to retreat from the reach of consciousness. The capacity for action mediated by non-conscious motivational mechanisms ("auto-motives") is claimed to be a central feature of an economically functioning psychological system, which thus frees the "resources" of consciousness to deal with other tasks (Bargh 1989, 24ff.; Bargh and Barndollar 1996, 462f.). The claim that such a structure might have been selected for in evolution fits well with the phenomenology of habitual action. One might wonder whether such an evolutionary explanation is equally plausible where an agent's wants* appear inaccessible for motivational reasons: it is certainly not obvious what the evolutionary advantage is of freeing consciousness from the burden of registering its bearer's own motivation where that motivation poses a threat to self-esteem. If genes are the units of selection, then the "repression" of sexual jealousy is, on the contrary, likely to be counterproductive. Nevertheless, whatever the precise aetiology of its object, if the auto-motive theory's explanation of habitual action is correct, then explanations of the same structure could be applicable where the inaccessibility of motivation is itself to be motivationally explained.[13]

There are, however, decisive differences between the two kinds of case. Whereas, under normal circumstances, we see the component actions of routines as under the agent's control, actions with Freudian-type unconscious motivation reveal

[13] Cf. Chartrand and Bargh (2002, 15f.), where the authors illustrate how they believe their findings on automaticity are applicable to Freudian-type cases.

something about the agent's motivation in spite of himself.[14] This difference derives from the way in which the action sequence is typically initiated. Routine action sequences are generally set in motion by conscious optative thoughts concerning the goal of the complex action of which they are a part.[15] In motivated inaccessibility cases, in contrast, it is the subordinate actions that are consciously initiated, whereas the want* whose means they represent is not, or is no longer accessible to consciousness.

The important point for our purposes is that, in spite of the auto-motive theorists' anti-consciousness rhetoric, a model of this kind looks like grist on the mill of the *CO* theorist. It assumes that the relevant unconscious wants* have had a previous life as conscious optative stands before retreating into dispositionality: we are the bearers of dispositional wants* to clean our teeth under certain circumstances because we originally acquired such wants* consciously. The automaticity of a going-to-bed routine grounds in the installation of want* dispositions that can be activated sequentially in a state of complete absent-mindedness.[16]

If this were to be an acceptable model for the explanation of action by wants* that are inaccessible for motivational reasons, it would support *CO*: certain wants*, once they have crossed an agent's mind, can retreat from consciousness – for one reason or another. However, before accepting such a parallel between habitual and "repressed" motivation, we need to consider more precisely the genesis of the attitudes responsible for the behaviour of agents such as Jeff and Harold. To this end, I want to discuss briefly some findings from the impressive body of work in empirical psychology that goes under the title of "goal priming". It has been shown that presenting subjects with words belonging to certain semantic fields in scrambled sentences tests has systematic consequences for their behaviour in ostensibly unrelated follow-up tasks. In particular, the presentation of terms belonging to the classic psychological "motives" (cf. Sect. 2.4.2, note 20) – "strive",

[14] The latter actions are, under the description as thus motivated, unintentional. What we should precisely say about the "intentionality" of the components of routine actions is less obvious than is often assumed. For an analysis of routine action and the relevance of Bargh's "auto-motive model", see below Section 9.5.3 and Roughley 2007a.

[15] Someone who goes running regularly will generally have relevant optative thoughts before getting changed and setting off. What will require no conscious thought are the routinized components of the composite action of running. In non-standard cases, habitual action sequences can also be initiated non-consciously. So-called action slips are unintended action sequences apparently set in motion by the non-conscious triggering of dispositional wants*, for instance, a bus driver "absent-mindedly" pulling into a bus stop whilst taking his family out in the car. Cf. Heckhausen and Beckmann (1990).

[16] Searle has argued that in routine actions the non-consciously produced component actions are no longer the effects of attitudes, but of mere dispositions. In his view, the body takes over in these cases (1983, 150ff.). In the article mentioned in note 14 above, I offer arguments why this cannot be right. Central among these is the continuity between non-consciously and consciously controlled routine actions, such as tying one's shoelaces. Where some such "automatic" action breaks down, we are generally able to carry on where we had left off, without having to ask ourselves what it was that we wanted to do. Compare the remarks on subintentional action (above Sect. 5.1.3).

"success" (achievement) on the one hand, or "friend", "reliable" (affiliation) on the other – has remarkable effects on whether the subject thus primed cooperates with, or competes against other subjects in joint tasks (Bargh and Gollwitzer 1994, 85ff.; Bargh and Barndollar 1996, 469; Bargh et al. 2001, 1016ff.). Moreover, the effects of such "goal priming" turn out to be precisely those that occur where the subjects are explicitly instructed to form an achievement or affiliation goal. Finally, "debriefing" after the experiments demonstrated that the participants had had no conscious goal that matched the marked increase in competitive or cooperative behaviour.

Now, John Bargh and his colleagues are clear that they see goal priming not as instilling completely new aims in the relevant agents, but as "activating" a goal they already possess. So these results might also appear to be consistent with *CO*. However, it is obvious that the "goals" thus activated have a content of such a high level of abstraction that their representation doesn't pick out any particular action. The actions performed in the service of such abstract "motives" in the joint tasks require the formation of far more specific subordinate wants* that specify ways of cooperating or competing. Now, these newly generated wants* to make the individual moves in the games are themselves likely to be conscious. Nevertheless, it is here that the decisive question for the conscious occurrentist looms: how precisely are we to conceive the *generation* of the situation-specific wants* from the general "motives" activated through priming?

The obvious answer appears to be that the cooperators and competitors play host to some form of instrumental reasoning. But if they do this on the basis of premises of which they are at the time unaware, then such reasoning processes must themselves take place below the level of conscious awareness. Indeed, the strong plausibility of explaining certain actions by means of unconscious mental states derives precisely from the appearance that the familiar structures of practical inference are at work in cases in which the agents have no expressive access to them.[17] The primed subjects appear at the very least to be bearers of the beliefs that making certain moves would be ways of realising their overarching competitive or cooperative goal. But once this is accepted, it is arbitrary to stipulate that there *cannot* be further wants* mediating between the basic motives and the specific action wants*.

5.1.6 Extending the Optative Conception

Return to Jeff and Harold. Their desires – for information as to possible activities of one particular person and for harm to another – are plausibly also secondary relative more basic wants* (cf. Sect. 3.2.1): perhaps to the fear of being humiliated and to

[17] This is again the Humean point that is crucial for justifying ascriptions of motivating states to non-linguistic animals (cf. Sect. 2.6, note 34).

the desire not to lose out to a member of “the opposite sex”. Their motivational condition relative to the objects of their concern thus seems to result from their coming to believe that there is some feature of their present situation in virtue of which it falls under the description of something they want* not to be the case. Again, if the indexical belief that mediates between two such wants* need not be conscious, it is unclear why the same should not be true of the wants* themselves. In the case of Harold: perhaps he has never had the conscious optative thought with the content that he not lose out to a woman. Maybe this want* is itself derived from the fear of being humiliated and the belief that being beaten by a woman instantiates the feared content. It doesn’t seem inconceivable that Harold may move mentally from the once conscious fear of humiliation to the desire that Betty suffer some harm without ever having played host to a conscious optative thought that he not lose out to a woman.

The view of Sartre and Broad that the causally efficacious wants* of Jeff and Harold must be conscious depends on the claim that only their conscious occurrence can explain why they are seen as undesirable and are therefore “repressed”. The goal priming experiments, on the other hand, provide evidence both for the efficacy of nonconscious wanting* and for the existence of nonconscious inferential mechanisms. If the attitudes realised in our actions can be generated in this way, there can be no a priori reason why nonconscious mechanisms might not withhold from consciousness access to types of motivation the agent would rather not have. However, even if we accept the claim that, in such psychodynamic cases, the mechanism responsible for the diversion of apparently threatening contents from conscious awareness has to be triggered by conscious awareness of the apparent threat, the Sartrian conclusion that any motivationally relevant wants* must be conscious still does not follow.

An alternative explanation might draw on the fact that certain *types* of thought contents trigger negative hedonic reactions before the details of those contents become clear to their bearers. It seems experientially plausible that we sometimes feel afraid, worried, apprehensive or in some less clear manner uncomfortable about something without being clear what it is. Might not such unpleasant qualia in turn trigger mechanisms that withdraw consciousness from the relevant thought? Earlier experiences of situations of a certain type, conscious optative reactions to them and their hedonic consequences might be necessary aetiological features. If something like this is plausible, then that would open up the conceptual possibility of *token* wants* – fears, desires for socially undesirable contents – occurring under the cover of the retreat of consciousness from the hedonic accompaniments of their *type* of content. Such mechanisms would permit “censorship” of the contents of optative thought contents without requiring that the specific thought have occurred consciously.[18]

[18] There is some evidence for mechanisms of this kind in the study of “motivated reasoning”. Cf. Baumeister and Newman (1994) 15f.

This suggestion is derived from speculation on the basis of some phenomenological and experiential evidence. That puts it pretty much on a par with Sartre's sweeping claim and Broad's model. There can, I think, be little doubt that Sartre and Broad are right about significant portions of our optative lives. The decisive question is whether they are right about the entire optative sphere and, above all, whether they are *necessarily* so. That would be the case if no sense could be given to the idea that a state of a person *would be* accurately expressed by an optative utterance where that state is not, and never has been conscious.[19] Certainly, an advocate of the optative analysis cannot simply switch to purely functional criteria here. Nevertheless, functional properties do take on an important role as symptoms: where a person finds herself acting, feeling, thinking and consciously wanting* in ways she doesn't understand, the hypothesis seems justified that she may be playing host to additional attitudinal occurrences of the kind she would express in optative language.[20] This is, as it should be, an *extension* of the concept of wanting* beyond its paradigmatic instances (cf. Goldman 1970, 122ff.): we can only make sense of the idea of unconscious wanting* because we possess the primary concept that is tied to conscious occurrences.

The truth of ascriptions of nonconscious wants*, I suggest, depends on the truth of counterfactuals concerning optative utterances an agent would produce if he were to gain an understanding of what is causing initially perplexing behaviour and were to wish to be sincere about the matter. Certain forms of overt behaviour, of perceptual salience and hedonic reactions, along with the generation of conscious wants* and beliefs that we would otherwise be at a loss to explain can *warrant* the ascription of unconscious optative states. However, it is of the nature of unconscious wants* that there can be *no strict criteria* for their correct ascription. If at t_2 an agent thinks the evidence through for his having wanted* some p at t_1 and, as a result, ends up expressing the relevant optative stand, that increases the plausibility of the claim that he had at t_1 wanted p. However, because of the time lag involved, we cannot exclude the possibility of the agent having acquired the want* retrospectively in the process of his reflection.[21]

[19] The space of possibility that has to be straddled by such a counterfactual seems fairly narrow for linguistic creatures that habitually take on such expressive relations to features of their orientation in the world. In contrast, there is an enormous gap to be bridged if we are to make sense of Robert C. Roberts' claim that construals of situations by non-linguistic animals such as squirrels "have a structure that *would* be expressed in the corresponding sentences or their rough equivalents" (Roberts 2009, 223). This counterfactual covers the addition of whole layers of competence that give sense to talk of expression in the first place.

[20] Contrast William James (1890, 174ff.), whose conception of mental states as qualia leads him to explicitly reject the ascription of attitudes on the basis of what he also calls "functional" criteria. According such a view, which models desires on feelings, want* tokens are necessarily tied to conscious occurrences. In contrast, the criterion of optative expressibility leaves undecided whether the paradigmatically conscious character of the states thus expressible can allow for non-conscious variants.

[21] This is a methodological, in principle unsolvable problem for most kinds of psychotherapy. It is an advantage of the optative analysis that it can explain it.

It should be noted that the wants* in subintentional, goal-priming or "repression" cases cannot be unconscious in the same sense as mere dispositional wants*. The latter are presumably brain structures, typically bearing the trace of conscious optative thoughts, the triggering of which tends to cause both further optative thoughts and other features of the optative syndrome. In contrast, the wants* we have been discussing are concurrently nonconscious and causally efficacious with respect to actions, perceptions, hedonic experience or further attitudes.[22] Unconscious wants*, thus understood, are located on the *occurrent* side of the divide between dispositional and occurrent attitudes*.[23]

The first part of this chapter began from the assumption that the expressive explication of optative attitudinising ties wanting* closely to consciousness and went on to ask how closely. The discussion of cases that reveal the non-necessity of relevant conscious occurrences will obviously appear superfluous to the functionalist, as to the physicalist or language-of-thought theorist. However, it shows that an analysis which emphasizes wanting's* connection with consciousness need not deny the significance of the non-conscious occurrences such other theorists wrongly take to be paradigmatic. Sometimes we pursue goals without being aware of doing so and sometimes we generate further motivation through non-conscious forms of inference. This is compatible with a conception that takes the capacity for expressive articulation to be what gives us a grip on the idea of wanting* in the first place, where that capacity is inconceivable independently of concurrent conscious mental events.

5.2 Not Really Wanting

Jaegwon Kim has distinguished two dimensions of the doctrine of the mind's transparency (Kim 1996, 17ff.). Both of these involve epistemic relations between attitudinising and believing or knowing that one is thus attitudinising. According to the first, mental states are *self-intimating* to their bearer. Applied to wants*, this would be the case if, necessarily, if *X* wants* *p*, *X* believes she wants* *p*. *CO* differs from the claim that wants* are self-intimating in two respects. Firstly, it doesn't concern higher-order assertoric attitudes, i.e. beliefs about wants*, but conscious optative occurrences. Secondly, it requires that such an occurrence be either contemporaneous or prior to the moment of correct ascription. *CO*, as we

[22] For this reason, it is at least misleading to characterise unconscious wants* as "dispositions" to have conscious thoughts (Searle 1992, 161f.; 1994a, 850, 854).

[23] Goldman introduces the distinction between "standing" and occurrent wants in terms of whether the relevant want content "occurs to" its bearer, i.e. whether it is a conscious thought (1970, 86ff.). If however, as is generally the case, we take "occurrent" wants* to be thus named because they are occurrences, i.e. events (or perhaps processes), rather than states, then Goldman's equation of "occurrent" with "conscious" is fallacious (cf. Mele 1995a, 316; 2003a, 30). Goldman's own thoughts on unconscious wanting, which are largely consonant with my own, make this perfectly clear.

have seen, is false. This is presumably at least part of the explanation of why wants* are not self-intimating. An agent who adopts a conscious optative stand towards some object will be disposed to believe that he wants* that object, should he have reason to develop a belief about the matter. The lack of any such conscious thought on the matter removes the only direct form of access of an agent to her wants*.

Now, I have argued that there are conceptual ties both between wanting* and optative utterances and between the capacity for expressive articulation and consciousness. This commits me to seeing wants* as characterised by a relation that is at least comparable to what Kim calls *infallibility*. Infallibility is for Kim the converse relation between belief and attitudinising to that of self-intimacy. Thus understood, an agent's knowledge of her optative attitudes is infallible if, necessarily, if *X* believes she wants* *p*, she wants* *p*. In what follows, I shall be arguing that, although we are not infallible relative to our wants*, a conscious thought of the form "Let *e* happen" is sufficient for its bearer to want* *e* to occur. We are fallible relative to our wants* because of the conceptual and empirical gap between conscious wanting* and beliefs about wants*. The more interesting claim concerns cases in which a person has a conscious optative thought concerning her φ-ing, but of whom we nevertheless seem able correctly to say that she "doesn't really" want to φ. The explanation for this, I claim, is that everyday talk of wanting picks out features of the agent that go beyond wanting's optative core.

5.2.1 "I want *p*" as a Fallible Thought

Because of the characteristic transparency of the assertoric mode, the belief that *p* is typically expressed by the utterance "*p*", whereas a thought or utterance of the form "I believe that *p*" tends to indicate some sort of distance to the attitude's content on the part of its bearer (Sect. 4.1.1). Such an explicit self-ascription of a belief could also be its expression under restricted circumstances, say if a group of people agree explicitly to declare their beliefs one after the other. Under more usual circumstances, the higher-order structure involved in such self-ascription makes itself felt. It is, for instance, an appropriate way of ascribing oneself a belief merely on the basis of behavioural evidence that could equally be garnered by an observer. It might also be used appropriately by someone ascribing herself a dispositional belief on the basis of inference by exclusion. The forgetful shopper, who assumes that one of the washing powders on the shelf before her must be the one she believes is able to remove all those resistant stains, may conclude she believes that Kleen is the powder with that property (Sect. 4.1.1, note 2). In either of these cases, the self-ascriber may be wrong about what she believes.

Wanting* looks to be different from believing in these respects because there seems to be no general option of expressing an optative attitude without explicitly articulating its mode. The thought or utterance "I want *p*" will therefore frequently be more than a mere self-ascription, equally expressing – without articulating – the relevant optative attitude.[24] It would be phenomenologically absurd to claim that a small child spontaneously crying out "Oooh, I want an ice cream too!" must be the bearer of a higher-order belief concerning her want*. A little attention to the phenomena actually makes it clear that wants* can actually be expressed without even mentioning the optative mode. The child might express the same thought by simply exclaming "An ice cream!" The lack of modal, or indeed propositional articulation of the utterance doesn't entail that the thought thus expressed have no such structure.[25] Rather, should the relevant mental state be such as to be fulfilled if the child comes to have an ice cream, then she has expressed an attitude that would be most explicitly articulated by a sentence in the optative mode.

Sometimes, however, the utterance "I want *p*" may, like "I believe that *p*", be a mere report of a person's mental state. Something analogous can also be true of the corresponding thought, which may simply be matter of ascertaining that one is playing host to the attitude. Returning to the forgetful shopper: after concluding that Kleen is the powder with the impressive powers she was after, she will naturally develop the mistaken belief that it is Kleen she wants to buy. The mistake is not cancelled by her then going on to form the subordinate want to buy Kleen. That is a further movement of the mind, not entailed by the belief that gives rise to it. Similarly, Harold (Sect. 5.1.4) may, as a result of observing his actions and emotional reactions, come to believe that he wishes Betty harm. The fact that his access to his want* is mediated by self-observation, and perhaps by the testimony of others, prevents his thoughts on the matter being expressions of the want* and makes them categorically susceptible to being mistaken.

"I want *p*" can thus be used fallibly where it does not function to express a person's optative attitudinising. Whether the verb is being used expressively or descriptively may not at first be apparent. Unlike what one might suspect,[26] this is true of both first- and third-person uses. A sentence such as "Colin wants a coffee" is likely to be as descriptive as "Colin has a cold". There are, however, also circumstances in which the former sentence is to be understood as expressing a request or a demand on behalf of Colin (cf. Hare 1968, 47f.), so that the structure of the mental state conveyed by the utterance would be accurately articulated as: "Let it be the case that Colin's want* (that it be the case that he have a coffee) be satisfied". For another clear example of a first-person descriptive use of the verb, take the case of the person responsible for gathering and passing on the computer wishes of the members of a department. He might run down the list saying to himself "Jones wants *x*, Smith wants *y*, I want *z*". He could turn out to be just as mistaken about himself as about Smith or Jones. The descriptive utterance again expresses a belief and the belief could be false. In contrast, sincere utterances or thoughts of the

[24] Like Alston and Finkelstein, I see the claim that expression and self-ascription are mutually exclusive as mistaken. Cf. Alston (1967, 16) and Finkelstein (2003, 93ff.).

[25] On ways of expressing the modal component of an attitude without thereby articulating the full attitudinal structure, cf. Vendler (1972, 6f.) and Rosenthal (1989, 315f.).

[26] Wittgenstein suspected as such. Compare *Zettel* §472.

form “Let it be the case that *p*” – as opposed to hierarchical thoughts of the form “It is the case that I am playing host to the attitudinal stand (Let it be the case that *p*)” – are, I am claiming, sufficient for the attitudiniser’s wanting* that *p*.

5.2.2 Four Ways to Not Really Want

If this sufficiency thesis is to be phenomenologically plausible, it will have to be compatible with the fact that agents can find themselves legitimately confronted with assertions to the effect that they “don’t really want” something, although they have expressed a positive optative attitude toward it. There is a short explanation for this: wanting is not just wanting*. The longer explanation involves showing how the features that supplement wanting’s opative core are picked out by talk of “really wanting”. There are, I shall claim, three such kinds of feature. These bequeath us three meanings of the “really want” idiom. Before coming to these, there is a first kind of case we should note that has seemed particularly important in the discussion of altruistic motivation.

We have touched on a specific variant of this kind of case in our discussion of unconscious wanting*. In cases of the relevant kind, the agent suffers from what might appear to be a form of optative self-deception. In such cases, she believes, or is disposed to believe – incorrectly – that a certain conscious optative attitude motivates some action of hers. Helen, for instance, may think of herself as acting in order to help someone when she is really acting in order to appear admirable. And it may appear that this mistake can be manifested in the optative attitude itself, not merely in higher-order beliefs about it. Instead of a representation of Helen’s helping someone, what actually belongs in the attitude’s content, so someone might think, is a representation of her being admired.

Here, we simply need to be clear on the distinction between having an attitude and justifying or explaining it.[27] Taking on an attitude doesn’t guarantee that you know why you have adopted it, for instance, whether your want* is subordinate to some further optative attitude and if so, to which. This point, which obviously holds if there are unconscious wants* from which some conscious wants* might be derived (Sect. 5.1.5), holds equally where we are dealing with conscious wants*. You may well want* to help someone and believe that want* to be *intrinsic*, that is, believe you want* its object for its own sake, whereas you in fact only, or primarily want* it as a means to satisfy some other want*. Where there is something dubious about the want* in the background, we tend to talk of an “ulterior motive”. However, the latter wish being the reason for, or cause of the former does *not* mean that you don’t want* to help (cf. Goldman 1970, 122). What it does mean is that the desire

[27] The neglect of this distinction is at the root of the guise of the good conception of wanting (Sect. 4.3).

to help is instrumental and is likely to dissolve as soon as giving assistance appears no longer able to contribute to your more deeply held aim.

Helen might find herself confronted with the accusation, "You don't really want to help, you just want to be admired". We should interpret this as meaning: you don't want* to help intrinsically; what you intrinsically want* is to be admired. And it is perfectly possible that an observer might, in saying this, tell Helen something of which she in some sense wasn't aware. There are three different ways in which this might be the case. The observer might be correct in ascribing a desire to which the bearer momentarily has no expressive access, that is, a desire that is *genuinely unconscious*. Alternatively, the want* ascribed might be conscious, but have been *unattended* to, perhaps for the sort of reasons suggested by Broad. Finally, the attitudiniser might be aware of both her desire to be admired and her desire to help, *without recognising the causal connection* between the two. For each of these three reasons we can certainly be mistaken both as to whether some want* of ours is intrinsic or instrumental or as to which further want* has given birth to some instrumental want*. None of this tells against the sufficiency of conscious optative thoughts for wanting*.

Of the three other meanings of "really wanting", I briefly mentioned two in Section 3.3.4. According to the first of these, the motivational use of the expression, denying that someone "really wants" some proposition is more or less equivalent to asserting that they are insufficiently motivated to adopt the means or other course of action they deem necessary to bring it about. There can indeed be cases in which such a denial pinpoints a mistake on the part of an agent.

Little Pia's parents, for instance, have told her that they will pay for her piano lessons if she "really wants" to learn to play, that is, if she is prepared to practise regularly. She declares that, yes, she "really, really wants" to learn to play. A troubled year later, the piano lessons are cancelled and Pia's parents are convinced, in spite of their daughter's protests to the contrary, that she "didn't really want" to learn to play after all. Here, the original question as to whether she "really wanted" to learn was indeed a question about a matter of fact that involves more than simply whether Pia was the bearer of an attitude expressible by "Let it be the case that I learn to play the piano". Her parents were asking her for an estimation as to whether that want* was backed by sufficient motivational force. And in that point, so it seems, she was mistaken. Clearly, there is no sense in which the mistake could itself be optative. It is, as it has to be, a simple mistaken belief about a matter of fact. People can lead others to believe, or can even believe themselves, that some want* of theirs is backed by greater motivational force than is the case, just as they can believe or lead others to believe that a want* is intrinsic, when it in fact owes its existence to the belief in its instrumental value for the realisation of some other want*.

A *third* feature that an agent may be missing if he fails to "really want" something is a certain level of *stability* of the attitudinal posture he is tokening in thinking the corresponding optative thought. Just as we need not be aware how far we would be prepared to go in order to realise a want's* content, there is no guarantee that we know how long a certain want* is going to persist. In an example of Goldman's, Mary, who is wondering whether she "really wants" to marry a particular man, is concerned as to whether her present desire will last (Goldman 1970, 97). This is a further feature of wanting* to which the bearer of the attitude has no privileged access. She will often be best placed to make a judgement on the matter, but this need not be, and sometimes clearly isn"t the case.

Reflection on the diverse senses of "really want" allows a revealing pattern to emerge. Both the latter two meanings discussed correspond to ways in which optative attitudes can be furnished with "strength". The willingness to adopt a course of action subordinate to the want* in question marks a particular level of *motivational strength*. The form of "really wanting" about which Mary is unsure concerns her having a want* with a certain level of *persistence*. These two types of strength no doubt often co-occur. But there are such things as desires that are brief and overwhelming as well as wishes that accompany us over long periods of time without having a particularly strong influence on our action. These distinctions will be of particular importance in the discussion of the "intentional syndrome" in Section 7.2. What is important at this stage is that these two types of "strength" are, like the details of a want's* genesis, features of a wanter's* condition about which she can be mistaken and which, for this reason, can be predicated of her by means of the expression "really want".

A *fourth* and final feature whose absence can justify seeing an optative attitudiniser as not "really" wanting the attitude's object was also mentioned in Section 3.3.4. This feature is *hedonic*. We can be mistaken in our belief that we will feel better if we "get what we want". Where this turns out to be the case, we may, after the event, also say we "didn't really want" whatever it was. However, things are somewhat more complicated here, as there are several distinct, and in part incompatible uses of the expression that relate wanting* to hedonic experience. This diversity of usage derives from the manifold relationships entertained by wanting* to affect, the variety of which I only hinted at in outlining the non-agential features of the optative syndrome (Sect. 3.2.2). In the history of philosophy, one or other of these relations has been repeatedly seen as *constitutive* of desire or wanting. Before bringing the first part of this study to a close, I offer a brief overview over the various relationships between wanting* and affect. In doing so, I will sketch arguments as to why these relationships cannot count as essential to wanting*, although they can support talk of "really wanting".

5.3 Wanting* and Affect

5.3.1 Affect Dispositions

A first characteristic connection between wanting* and affect is that we generally tend to experience some level of hedonic gain if we acquire the belief that the content of a want* of ours has been realised. This has led certain philosophers to equate wanting or desiring with being hedonically disposed. According to Galen Strawson, for instance (1994, 283), "a reliable tendency to react with pleasure or displeasure to various occurrences may surely be held to be sufficient for the attribution of genuine likings, preferences or pro-attitudes".[28] Such a construal would support a literal reading of the expression "to really want" in the sentence "I didn't really want to see her after all", uttered by Alf, who feels awful after going out for a meal with his ex-wife. What we should be saying instead is, I think, that we simply have a want* minus a characteristic feature of want* satisfaction, but whose absence doesn't impugn the presence of the want*.

There are two distinctions to be noted here. It seems likely that the failure to recognise either of them will contribute to the plausibility of this version of conceptual hedonism. A first distinction is between two ways in which we are generally disposed to experience a hedonic improvement when we get what we want*. One way derives from the fact that a great deal of the propositions that are made the contents of our wants* become optatively framed *precisely because* we believe that their instantiation will bring us pleasure. Because such beliefs about our own hedonic proclivities tend to be correct, there will generally be a correlation between these proclivities and what we want*.

This point should be kept separate from the fact that people tend to experience pleasure on the realisation of their wants'* contents and displeasure on their non-realisation. This tendency is independent of whether the person had been disposed to react with positive affect to the realisation of the relevant proposition *prior to* forming the want*. Whereas the first type of disposition can provide a reason for want* formation, the second presupposes the want* to which it is a reaction.

We should, then, distinguish *want*-independent* and *meta-optative* hedonic dispositions.[29] Dispositions of the former sort, such as the tendency to experience distress on being burned or stabbed, lead naturally and automatically to the formation of corresponding wants*. Because in such cases the hedonic and the optative dispositions tend to be activated under the same conditions, they may be difficult to separate phenomenologically. Nevertheless, there is a clear conceptual difference manifested in the fact that the ascription of mere hedonic dispositions doesn't create intensional contexts. However the insertion of a short sharp metal instrument into the flesh of some human or animal is described, the event will activate a disposition to negative hedonic experience. In contrast, meta-optative affective dispositions are conceptually structured: satisfaction, delight, disappointment, regret, pride are all phenomena with content the correctness of whose description is sensitive to the descriptions under which the relevant content is wanted*. For related reasons, meta-

[28] To be fair to Strawson, he only talks of a "primary linkage" between affect dispositions and desire, and does not claim explicitly that in individual cases a person desiring something requires that that they be the bearer of a corresponding disposition.

[29] Tim Schroeder's conception of pleasure as the representation of a net increase of desire satisfaction relative to expectation relies on dismissing this distinction as insignificant (Schroeder 2001, 523ff.; 2004a, 97ff.). For a critique of Schroeder's theory, see Roughley unpublished a.

optative affective dispositions tend to be significantly modified by accompanying beliefs about the probability of the realisation of what is wanted*. Delight is more likely to occur if the realisation of what is wanted* was unexpected, just as serious disappointment tends to be caused by the non-realisation of contents whose realisation was strongly expected.

The second distinction that should be noted here is necessary because of the double duty done by the term generally used to designate the typical positive affect caused by attaining what one wants*. That term, "satisfaction", is also frequently used simply to refer to the realisation of a want's* content, an event that may or may not in turn be responsible for some such affect. Both the event of realisation and its hedonic consequences can thus go under the same label. Failure to keep *semantic* and *hedonic satisfaction* separate may be a second source of conceptual hedonism. However, this distinction ought to be clear merely from the fact that the former can occur without the latter.

One obvious price that a hedonic-dispositional conception of wanting has to pay is that it sacrifices the possibility of explaining action by means of an agent's wants. Were Alf's feeling awful after seeing his ex-wife to entail that he had not wanted to see her in the first place, we would need to explain his going out with her by different means. In fact, such a conception suggests a special variant of motivational cognitivism. On such a view, actions would be generally caused by *beliefs* about the agent's "desires", that is, about her affective constitution. This would be action theoretic cognitivism without its main attraction: its perceived adequacy to external reasons such as moral or prudential requirements (Sect. 4.3).

Moreover, the hedonic-dispositional conception shares the serious problems of other dispositional theories. First, it is unable to explain our characteristic non-inferential access to our wants*. Like the behavioural-dispositional theory (Sect. 3.3.2), it conceives our relationship to our wants as mediated by self-observation. Second, it leaves us wanting* things of which we have no concepts or indeed any kinds of representations at all (cf. Sect. 3.3.4). I may be disposed to experience particularly pleasant sensations if teletransported to Alpha Centauri, but it would be absurd to claim that I want to experience the things there that would trigger such sensations as long as I have no idea what they are.

5.3.2 Expected Pleasure

A second version of conceptual hedonism does not require people to want things they have never heard of. Moreover, it allows us to say that someone may still have wanted something to be the case even if she is disappointed on its realisation. According to this construal, wants are constituted by "the prospect of pain or pleasure".[30] In conceptions of this kind, for a person to want that *p* is for her to believe that, should *p* occur, it would bring her some form of hedonic gain.

The traditional response to this claim is that we want – or want* – things we won't be around to experience, such as our own fame or the well being of our children after our deaths (cf. Gosling 1969, 9ff.). This response still seems to me to be perfectly appropriate.

A second response is, I think, of primary importance for an understanding of why no hedonic conception of wanting can be right. This consists in pointing out the obvious fact that *instrumental* wants* are also wants*. We often want* things for the sake of other things. In fact, *most* of the things we want* are wanted* for the sake of other things.[31] Even where pleasure is the ultimate goal, the motivational transfer that takes place from end to means is by no means necessarily accompanied by the same transfer of hedonic expectations: wanting* *q* in order to bring about *p* simply does not entail expecting pleasure from *q*, just because *p* is expected to be pleasurable.

Jock, for instance, may well expect more pleasure, or less discomfort, from a life which is physically healthy and believe that unpleasant things like jogging before breakfast and cold showers are means to achieve this aim. If he therefore indulges in these practices regularly, he may even come to find them pleasant and thus to expect to find them pleasant. That transformation is, however, obviously no conceptual consequence of his instrumental wanting, but is rather a possible contingent result of his regularly performing the actions that are the means to his end. And, of course, that hedonic change may not set in at all, in spite of the fact that Jock manages repeatedly to muster the motivation to do what has to be done.

Positive hedonic expectations, then, are not constitutive of wanting*, although they may both explain want* formation and be closely bound up with want* states thus formed. Indeed, where someone wants* something and expects that its realisation will bring her a significant amount of pleasure, she may, again, refer to the compound optative-assertoric state as one of "really wanting". This use of the phrase generally requires a contrastive context, i.e. a situation in which wanting* qualified by positive hedonic expectations is set against a form of wanting* that is either hedonically unqualified or qualified by negative hedonic expectations, i.e. what I called "assent" or "willingness" (Sect. 4.4.2). A paradigmatic expression of

[30] The quotation is from Hume (T II, iii, 3). As I argued in Section 1.5, the main thrust of Hume's doctrine does not specify the *doxastic* relation to expected pleasure as constitutive of the passions. Rather, what Hume sees as generally essential is that such a belief elicits in its bearer a *present affective colouring* of the state of affairs thus represented. The fudging of the modal issue is not unusual among conceptual hedonists. Another prime example is Mill's claim that "to desire anything, except in proportion as the *idea* of it is pleasant, is a physical and metaphysical impossibility" (U 173; my emphasis). Compare the modal unspecificity of "idea" with that of James' use of "thought" (Sect. 2.4.3).

[31] Quite often, of course, what is ultimately being aimed at is pleasure. But even if pleasure were to be the ultimate *object* of all our wants (explanatory or psychological hedonism), that wouldn't make a relation to pleasure *constitutive* of what it is to want (conceptual hedonism).

this contrast is “It’s not (just) that I *have* to go. I really want to!” Note that “really wanting” in this expected pleasure sense is perfectly compatible with “not really wanting” in the affect disposition sense.

5.3.3 Present Discomfort

There are two further ways in which the optative and the hedonic can be bound together that have been elevated by philosophers to a constitutive status. The first corresponds to the conception advanced by Locke in the *Essay*, Book II, chapter xxi (§§29ff.). According to what seems to be the main strand of Locke’s thought on these matters, “desiring” that p is equivalent to experiencing discomfort at $\neg p$.[32] Thus understood, desiring p is a matter of feeling discomfort or “unease” at p-lessness. One can capture the idea by saying that, where beliefs “aim at truth”, desires “aim at relief”.

If this has a certain plausibility for what Locke (E II, xxi, §34) calls “the uneasiness of hunger and thirst”, it is hardly generalisable. It makes acting on our wants always a matter of conforming to what Nowell-Smith dubbed the “itch-scratch pattern” (Sect. 1.6). But if Holly gets up on the first morning of her holiday and ponders what to do on that day, she doesn’t do so by checking herself for signs of discomfort. The mental process gone through is, on the contrary, likely to involve the positive imagining of possible goals for the day. Wanting* can be a matter of pure forward-looking “pro-jection”. There is no plausibility to the idea that what persons are motivated to do is necessarily to change things they are uncomfortable with. A great deal of our social practices would become completely incomprehensible if we were to try to understand them in these terms.

Note that there is a second kind of actual discomfort associated with wanting or desiring, which is clearly no candidate for a constitutive condition, but whose widespread co-occurrence with wanting* might be in part responsible for the Lockean theory. This is a form of *second-order discomfort* at the non-realisation of some wanted* p, what one might call the “pain of exclusion” (cf. Duncker 1940/41, 417; Plato Phil 36a).[33] The significance of such optative-hedonic conjunctions is shown by the set of terms by means of which we refer to them, terms such as “to crave”, “to hunger” or “thirst for” and “to be dying for”.

This does also seem to be another context in which the really wanting idiom might be used. If so, it may pick out cases of either causal constellation: wanting*

[32] There are passages in Locke, for instance *Essay* II, xxi, §37, which may look more like evidence for explanatory, rather than conceptual hedonism.

[33] This presumably results from the triggering of meta-optative affective dispositions of the kind discussed above in Section 5.3.1. Disappointment looks like a specific version of the pain of exclusion, triggered by the belief that the relevant want* has not been realised within some context or time frame within which its realisation had been expected or longed for.

p because of the discomfort at *p*-lessness (really wanting a drink) or discomfort because of the conjunction of non-*p* and wanting* *p* (really wanting to see someone). Note that, as present discomfort – in contrast to one's motivational or hedonic dispositions – isn't something about which an agent can be mistaken, the feature of reality ("really") picked out by the idiom must be the causal connection of either kind.

At times, it may be difficult to ascertain whether the discomfort is of the want*-independent, proto-optative variety or whether it is a case of second-order "pain of exclusion". The test is whether the discomfort could be eradicated by eradicating the want* itself. If relief can be gained either by destroying the want* or by (semantically) satisfying it, then the discomfort in question is a form of the pain of exclusion.

5.3.4 Imaginative Pleasure

A final attempt to define wanting in terms of affect was proposed by Moritz Schlick (1930, II.4; VIII.4; cf. Fehige 2001). According to Schlick, a person wants *p* if they gain pleasure from imagining *p*. Schlick argued that we can be motivated to bring about an event it is pleasant to imagine in spite of believing that the event itself will be unpleasant – for instance, dying a heroic death. For this reason, he claims, the *imaginative*-hedonic conception of wanting[34] is more plausible than any position that ties wanting to expected hedonic gain.

Actually, the theory is not even particularly plausible on its own terms. As it is the positive feature of pleasure that is supposed to explain motivation, the theory needs to explain why imaginative pleasure doesn't lead to behavioural inertia: if I can get pleasure from thinking about *p*, why should I take the trouble of trying to bring *p* about?

Moreover, in spite of the fact that enjoying imagining the realisation of some state of affairs will often lead someone to want* to really experience it, there is no necessity that wanting* to do or experience something involves imagining enjoying that thing. Even when the reason we want* some *p* is that we expect pleasure from it, we can want* that *p* without imagining the experience of *p*. Concetta might want to go to a concert because it has been recommended to her by a friend, who has told her he is sure she will enjoy it. But, as she knows nothing more than this about what she can expect from the concert, Concetta is hardly likely to even attempt to imagine the experience.

Conversely, deriving pleasure from imagining a proposition is by no means sufficient for wanting it to really happen. It is surely a mark of the peculiar human life form that we can gain pleasure from the imaginative experience of things whose

[34] Actually, Schlick saw this as an explanation of motivation, rather than of wanting.

real occurrence would horrify us. Fans of horror films would presumably try to avoid any experience of (non-philosophical) zombies, were they to crop up in their real life.

It is instructive here to compare imaginative and perceptual pleasure. There are things we enjoy perceiving without them necessarily stirring any kind of desire in us. For instance, we are apparently biologically wired so that the sight of the bodily proportions of young children or animals causes a surge of warm feeling in us. But surely that doesn't entail us wanting* anything in particular. If you find the sight of some baby cute, there is no implication that you either want such a baby or even want to continue looking at it. In this respect, perceptual and imaginative pleasure are, as far as I can see, of a piece.

5.4 End of Part I

In his *Pleasure and Desire* (1969, 122ff.), J.C.B. Gosling argues that we would have difficulty seeing ourselves sharing a life form with creatures whose "wanting" is completely cut off from hedonic relations.[35] This seems to me to be perfectly correct, not however for conceptual reasons, but because of the mental constellations of which our wants* are characteristically key components. The hedonic elements of these constellations vary considerably along the dimensions I have discussed in this last section, often qualifying wanting's* optative core in several such ways simultaneously. Creatures that wanted*, but never experienced related pleasures or displeasures would be as alien to us as creatures that wanted*, but were never motivated to act accordingly. Nevertheless, creatures of both kinds are conceivable,[36] just as there are empirical examples of wanting* that lack either motivational or hedonic dimensions.

The optative analysis of wanting in terms of a linguistic structure analogous to that of the attitude entails that to want is to be the bearer of a mental state capable of conscious articulation. It leaves open the extent to which there may be wants* that are never manifest in consciousness. The empirical evidence, I have argued, constrains us to extend the optative conception to make conceptual room for such nonconscious wants*. A second implication of the optative analysis is that a thought with optative structure is sufficient for the thought's bearer to want* the content in question, although it will generally be insufficient for their "really wanting" it.

[35] "[P]eople who to any considerable extent fail to want$_{\text{P}}$ [i.e. want "with pleasure connotations"] are peculiar at least in the sense of being unlike most human beings" (1969, 123).

[36] A description of creatures of the latter kind is provided by Strawson in his (1994, 251ff.).

Part II
Intending

Chapter 6
Intention, Belief and the Irreducibility Thesis

6.1 Introduction: The Irreducibility Thesis and the Role of Belief

6.1.1 Eminently Practical Attitudes

Since Aristotle, the philosophy of human action has distinguished between some kind of basic, proto-practical attitude and a kind of attitude that is practical in a more emphatic sense, having a more intimate relation to the action of its bearer. Aristotle distinguished "prohairesis" from the three non-deliberative forms of "orexis" (Sect.1.3); for Hobbes, only an "appetite" or "desire" that issues from deliberation can count as a person's "will" (Sect. 1.4); Locke (E II, xxi, §§5, 28–30) saw voluntary actions as caused by "volitions", a kind of self-command given after weighing against one another the forms of uneasiness that in his view constitute "desires"; and Mill (U 173) declared the "will" "a different thing from desire", although "originally an offshoot from it".

Certainly, some distinction of this kind has to be established, if a theory is to be able to lay claim to phenomenal and explanatory adequacy. Hume's conception of "the will" as a mere epiphenomenal feeling, experienced when certain kinds of actions are "knowingly" carried out (T II, iii, 1), is a blatant dismissal of everyday self-understanding in this point. Whilst Hume's dismissal was perhaps in part motivated by a rejection of Scholastic faculty psychology, there is overwhelming everyday evidence for the fact that a distinction of this kind is not simply a remnant of misguided theory.

On the contrary, although "willing" and "the will", not to mention "volitions", are terms with a fairly restricted use in everyday language, "intention" is one which plays a highly significant role in our self-understanding, as well as in our legal practices. And that role is clearly bound up with our conviction that the state

N. Roughley, *Wanting and Intending*, Philosophical Studies Series 123,
DOI 10.1007/978-94-017-7387-4_6

designated by the term entertains a particularly intimate relationship to our actions. A key question posed by an investigation of the referents of any of the above terms is whether the *specificity* of the phenomena thus identified grounds in a psychological *irreducibility* of some special attitudinal feature.

Hobbes' notion of the will as simply whatever desire happens to survive deliberation and then lead to action[1] is a minimal recognition that everyday understanding distinguishes certain attitudes as eminently practical, combined with the conviction that the distinction thus established is of no substantial interest. In contrast, Locke's claim "that desiring and willing are two distinct Acts of the mind" (E II, xxi, §30) is the contention that the two mental properties thus instantiated are of completely different kinds.

6.1.2 The Irreducibility Thesis and the Structure of Part II

It seems obvious that intending to do something involves something other than desiring to do it, in the everyday sense of "desire", and something more than wanting* to do it, in the technical sense I have been employing. The central conceptual question raised by the relevant range of phenomena for a philosophy of practical mind is what that "other" or that "more" consists in. According to the *irreducibility thesis*, the relevant supplementary component resists analysis in terms of any other psychological concepts. Instead, intending is at core a unique kind of *executive* attitude, whose specific effects we can catalogue and at which we can gesture by means of metaphors that evoke the phenomena in agents familiar with them. What is impossible, however, according to the irreducibility thesis, is a componential analysis that can show intention to be constituted out of other psychological building blocks.[2]

Characteristic effects of intending are the initiation, sustaining, guidance and termination of more or less complex actions (Brand 1984, 173ff.; Adams and Mele 1989, 513ff.; Mele 1992a, 130ff.; McCann 1991, 197); the longer-term constraining of practical deliberation and further intention formation (Bratman 1985, 222ff.;

[1] Hobbes doesn't distinguish between the two criteria.

[2] The irreducibility thesis is advanced in Hampshire (1970, 131), Harman (1975/76, 432; 1986a, 78ff.; 1986b, 367), O'Shaughnessy (1980II, 310), Brand (1984, 126f.; 1986, 221), Bratman (1987, 10, 20, 110, 121; 1999a, 110), Heckhausen and Gollwitzer (1987, 103), Bishop (1989, 113ff.), Mele (1992a, 154ff.; 1995a; 71ff.), McCann (1986b, 128ff.), Tuomela (1995, 54) and Holton (2009, 17ff.). The introduction of the irreducibility thesis into contemporary philosophy of action seems in part to have resulted from the reception within the mainstream causal action theory inaugurated by Brandt and Kim (1963), Davidson (1963) and Goldman (1970) of the orthogonal conceptions of Sellars (1966, 17) and Castañeda (1975, 274, 290). According to Sellars and Castañeda, desires or wants are definable as dispositions to intend, intending being the primitive, genuinely practical attitude that can be most closely seen to parallel belief. Both Brand and Bratman contributed early essays on intention to a volume presented to Castañeda (Tomberlin 1983). Cf. also Brand's defence of Sellars against Davidson in (Brand 1989, 424ff.).

1987, 16f.; 28ff.; Mele 1992a, 138ff.); the formation of specific kinds of beliefs; and a general tendency to uphold and give priority to the realisation of the intention's content over and above those of other optative attitudes.

There are three principal metaphors used in the literature to draw out our everyday understanding of what it is to be in the relevant kind of executive state: forming an intention, or deciding, it is often said, is essentially *settling on*, *setting oneself to* or *committing oneself to* a course of action. Correspondingly, the resultant state of intending is characterised as a matter of *settledness* or *being set on* realising, or a *commitment to* realise the intention's content.[3] These are all terms that are admirably suited to pick out the fact that a special kind of step has been taken in coming to intend. And it seems clear that the step is a step beyond the mere "Let it be the case that ..." constitutive of wanting*.

Part II of this investigation asks whether the irreducibility thesis is true. The answer I shall give will be negative. Justifying this rejection requires an analysis that can account for a considerable amount of phenomenological, causal and normative evidence, evidence that has been primarily brought forward by the proponents of the irreducibility thesis. This is evidence that standard "belief-desire" analyses are unable to account for adequately. Perhaps on the reductionist side, the conviction that intention is, at core, "nothing special" has led to an insufficient sensitivity to the breadth of the relevant phenomena, indeed to their highly specific nature.[4] But a detailed look at the causal, experiential and deontic features associated with intending shows that the discussion cannot concern the question of whether or not intending is a *specific* attitude or not. There can be no doubt that this is the case. The question at issue can only be *what explains this specificity*. And in order to answer this question, one first has to take seriously the job of detailing what that specificity consists in.

That is the aim of the next two chapters. This chapter deals with intending's special relationship to believing; Chapter 7 with further characteristically related phenomena, some of which are causal and some of which are normative. The conjunction of these phenomena constitutes what I shall call the *intentional syndrome*, which in certain respects differs markedly from the optative syndrome detailed in Section 3.2. The full extent of the challenge for the reductionist is only clear once this syndrome has been described in detail. It turns out that it is the normative phenomena associated with intending, what I label the "intention-consequential requirements", that raise the most serious difficulties for a reductive approach.

Chapters 8 and 9 then offer a constructive conception of intention that claims to account for the intentional syndrome. A key move on the way is the rejection of

[3] Von Wright (1971, 96ff.), Harman (1975/76, 438; 1986a, 94; 1986b, 368ff.), Kim (1976, 255f.), Bratman (1985, 223; 1987, 4, 10, 15ff., 27, 61f., 108ff.; 1999a, 2), Heckhausen and Kuhl (1985, 150ff.), McCann (1986b, 131; 1991, 197f.), Velleman (1989, 111f.; 2000b, 25), Gollwitzer (1990, 57; 1996, 292; 1996, 493), Mele (1992a, 158ff.; 1995a, 71f.; 2000 100) and Wallace (2001, 107).

[4] Exceptions are Audi (1986; 1991) and Ridge (1998).

the unity of intention thesis: the claim that all intentions are attitudes picked out in terms of their realisation of the same conditions. Intention, I will be arguing, is a *disjunctive* concept: although all intentions are optative attitudes on which a unique practical status has been conferred, that unique status can be conferred in one of two different ways. The paradigmatic way of its conferral is by decision; secondarily, it can be conferred by a conjunction of a conscious thought and motivational properties. Either way, it is a concept that only makes sense against the background assumption of the possibility of practical deliberation. In the last chapter, I shall go on to argue that the subjection of agents to the intention-consequential requirements is a consequence of the role of intention formation as the *anchor of responsibility*. The requirements specify the conditions whose non-satisfaction would undermine the intelligibility of that step.

6.1.3 Intention and Belief: This Chapter

A good starting point for the investigation of intending's specificity is its relation to believing. It is a striking fact that the typical linguistic forms with which we give voice to intentions have, unlike most want* expressions, the surface grammar of purely assertoric, predictive usage (Anscombe 1957, 4; Broome 2009, 78). This seems bound up with there being a particularly close relationship between intending to perform some action and believing that one will perform it, a relationship that might appear to provide the key to intention's unique practical status. Intending to φ, believing one will φ and φ-ing form a typical syndrome. The question is why.

Two kinds of answers naturally suggest themselves, answers that fit nicely with either rejecting or advancing the irreducibility thesis. According to the first, intending is a matter of taking on a conjunction of optative and assertoric attitudes, where both concern one's performance of the action wanted*. On such a construal, intention's peculiarly intimate relation to action is taken to result from the pairing of an optative attitude with a belief that one will perform the wanted* action: perhaps because the expectations we have about our own actions are by and large fairly accurate or perhaps because such expectations somehow bring about a special fixation of our motives.[5] On an alternative construal, according to which intending involves taking on an irreducible executive attitude, the relation between intention and self-prediction is naturally seen as less tight: we tend to believe we

[5] The most straightforward theory of this kind was advanced by Monroe Beardsley (1978, 176ff.), for whom an intention is "a co-referring want-belief pair". A more complex proposal is that of Robert Audi (1973b, 64f.; 1986, 18; 1991, 362), who sees the basic intention-constitutive attitudes as the agent's belief that she will perform the relevant action and a want to do so that is motivationally unrivalled, except where the status of some motivationally equal or stronger want is epistemically obscured. A third "belief-desire" reduction, proposed by Wayne Davis (1984, 147), sees intention not as a conjunction of the two attitude types, but as an expectation concerning one's future action caused by a desire to perform that action.

will perform those actions towards whose performance we have this specifically executive attitude. We might do so either because we are aware that being the bearer of the executive attitude generally leads to us performing the relevant action or else because there might seem to be something irrational about taking on the attitude in the absence of such a belief.[6]

The discussion of the relationship between intention and belief thus contains indirect pointers to the irreducibility question. I will argue that attempts to distil "commitment" out of an assertoric component are unsuccessful. The linguistic and phenomenological facts about the relation between intention and belief are on the antireductionist side of the alternative just set out. However, the alternative is clearly not exhaustive. A belief component could be part-constitutive of intending, whilst the decisive supplementary feature could still be irreducible. A number of authors have also claimed that intention is a purely doxastic or epistemic state.[7] Finally, it is conceivable that belief and intention are more or less contingently related, whilst intention is subject to a form of reduction that doesn't depend primarily on doxastic components.

It is for a position of this last type that I shall be arguing. Not only, I shall claim, is the irreducibility thesis false; so is any straightforward belief-"desire" conception of intending. There is, as irreducibility theorists have argued (Adams 1986, 288; McCann 1986a, 205ff.), no positive doxastic condition that has to be fulfilled for someone to intend to do something,[8] although there will turn out to be a somewhat recondite conceptual role for belief to play. The arguments of this chapter should thus, at least up to a certain point, be congenial to proponents of the irreducibility thesis.

I begin Section 6.2 with a review of the various ways in which an intention's bearer can refer to his mental state, focusing in particular on the surface-grammatical assertoric form characteristic of the linguistic means of intention expression. This is plausibly interpreted as the articulation of intention's eminently practical character. I then discuss and reject (Sect. 6.3) attempts to understand this "supra-optative" component in doxastic terms, which take the assertoric form of typical intention

[6] Irreducibility theorists with a position along these lines with regards to belief are Bratman (1987, 37ff.) and Mele (1992a, 140, 146ff.), both of whom claim that there is at most a weak doxastic requirement on intending. Holton (2009, 47ff.) suggests that intention irreducibly conceived may require no belief condition at all.

[7] The former position is illustrated by Harman (1975/76, 432ff.; 1986a, 90ff.). Fishbein and Ajzen (1975, 12f.), Davis (1984), Velleman (1989, 109ff.) and Setiya (2007, 664) all suggest that intentions can be identified with beliefs with a specific function, content or genesis. Anscombe (1957, 87), Hampshire and Hart (1958, 11ff.) and Hampshire (1970, 134) see intention as intimately bound up with a form of "knowledge". Hampshire sees the relevant knowledge as possessed "in virtue of having formed firm intentions" (Hampshire 1975, 53) and thus as presupposing intention. He takes the concept of intention as "fundamental and unanalysed" (Hampshire 1970, 131). Anscombe, in contrast, may be identifying practical knowledge and intention (cf. Anscombe 1957, 87), although she thinks of neither as mental states.

[8] The most radical rejections of any doxastic condition are Thalberg (1962) and Ludwig (1992).

expressions literally. Other ways in which belief and intention could be related are then reviewed. I conclude that we should distinguish three doxastic features at work in intention and intention's environment. First, there is a doxastic *conceptual* condition. However, the condition is both weak and restricted in its application. Second, there is a closely related *rational* requirement on paradigmatic forms of intending. In spite of the close relationship between the two types of condition, it is important that they are distinguished. Third, there are also positive beliefs concerning an intention's contents which are typically caused by, and are thus mere *characteristic symptoms* of intending. These are not even plausible candidates for the status of conceptual constituents (Sect. 6.4). These doxastic effects, if we are to believe certain social psychologists – and I think we should – are in part produced subpersonally by our decisions or by our planning to implement decisions. A significant increase in the subjective probability an agent would assign to his doing something he intends is no doubt also a result of induction from experience. I conclude (Sect. 6.5) that the discussion of belief, rather than uncovering the completing conditions that transform an optative into an executive attitude, on the contrary, indicates that the step into "commitment" will have to be explained by different means.

6.2 Intention Expression

The discussion of Moorean optative sentences in Section 4.2 showed that there is an important distinction to be drawn between the articulation of wanting's* mode from the internal perspective of its bearer and the ascription to the bearer of the relevant attitude from an external perspective, even where the wanter* and the ascriber are the same person. When we turn to intending, we can see both that the same kind of distinction is to be made and that the expression of intention has what may seem to be a puzzling proximity to the expression of belief.

There are *three* linguistic forms that can plausibly be seen as *expressions* of an intention. The first two are first-person forms of the semi-modal "going to" (cf. Anscombe 1957, 1ff.) and of the present progressive aspect, usually with the mention of some future time. "I'm going to meet Antonia" and "I'm meeting Antonia tonight" both express Tony's practical perspective on a future action of his: where Tony utters either of these sentences, it seems he expresses a positive optative stance on his meeting Antonia tonight, in conjunction with the further state of being "settled on" meeting her.[9]

The use of either linguistic construction expresses the attitudinal framing of a proposition indexed in the future, a framing that confers on the event thus represented what one might call *practical imminence*. That such "practical imminence"

[9] This decisive additional component is neglected by both Kenny (1963, 218) and Hare (1968, 55), who equate intending to φ with "saying in one's heart 'Let it be the case that I φ'".

goes beyond the perspective established by a proposition's optative framing is strongly suggested by the fact that in other contexts these linguistic constructions are used to express assertoric attitudes[10]: to be precise, expectations of some event for whose future occurrence there is evidence at the time of speech. For instance, "It's going to rain" predicts rain on the basis of some present evidence such as the black clouds in the sky. Things are complicated by the fact that, in the case of predictions asserted by means of the present progressive, the evidence drawn on is necessarily a present arrangement, plan or intention (Quirk et al. 1985, 215). "They are leaving tonight" describes a future action of others made likely by arrangements they are known to have made.

What ought to be clear is that the practical imminence that can be expressed where these constructions are used in the first person is distinguishable from the theoretical certainty of predictions that may use the same words. "Tony is meeting Antonia tonight" or "Tony is going to meet Antonia tonight" could be assertions. If so, they could turn out to be false. However, if Tony tokens his intention using either of these idioms, but ends up not meeting Antonia, then his thought is not falsified. If, for instance, he simply changes his mind, he would not see himself as having made a mistake in tokening his earlier, now revised intention (cf. Hampshire and Hart 1958, 11).

Of course, if Tony *says* to you "I'm meeting Antonia tonight" and you then organise your evening in some way that depends on Tony's meeting Antonia, for instance, burgling her flat while she's out, you may feel you were misled if Tony changes his mind and cancels the meeting. But what you precisely should be inferring from such a remark on his part will depend on further facts you know about Tony, and perhaps about his relationship to Antonia. Note that, if he uses a linguistic form apt to be understood as a mere self-ascription, rather than an expression of the intention, by saying "I *intend* to meet Antonia tonight", that may weaken your tendency to rely on him so acting.[11]

The lexical items by means of which an agent generally *asserts that* someone, perhaps she herself, intends to do something, are the verbs "to intend", "to mean", "to aim" and "to plan" to do something.[12] "I am planning to φ" may be used in a fairly extended sense to indicate that the bearer is additionally in the process of

[10] John Broome claims that the possibility of using the same indicative sentence to express a belief or an intention is sufficient evidence that expressing an intention is also expressing a belief (Broome 2009, 80). This is much too quick.

[11] David Velleman has claimed that a cognitive conception of intention is supported by the fact that hearing an agent express an intention licenses the hearer to expect a corresponding action (Velleman 2007, 207). Such license need, however, by no means depend on a *conceptual* connection. On Velleman's particular brand of cognitivism, see Section 6.3.1 below and Roughley (2007b).

[12] As with other optative attitudes (cf. Sect. 5.2.1), we cannot read off the expressive or ascriptive character of an utterance referring to intentions from mere linguistic form. The distinction between the linguistic means of intention expression and self-ascription only concerns characteristic cases.

forming intentions subordinate to the intention to φ.[13] It can, however, also be used in a minimal sense in which it is equivalent to "I intend to φ".[14]

We can combine one of the two grammatical devices we use to express intentions with a negated assertion using one of the terms just mentioned in order to produce Moorean sentences for intention. "I'm meeting Antonia tonight, but I don't intend to meet Antonia tonight" involves a kind of type 1 Moore-incompatibility. However, "I'm going to meet Antonia tonight, but I don't intend to meet Antonia tonight" permits of a non-paradoxical reading, if the first conjunct is read as the expression of a belief: I might just believe I will inevitably run into her.

There is a third important linguistic construction that plays a central role in intention's expression. This is typically the abbreviated modal construction "I'll φ", sometimes employed in its full form "will". Infrequently these days, "shall" also plays this role. Again, both of these are constructions otherwise used predictively. Comparing these modal articulations of intention to the progressive and going-to constructions reveals an interesting fact, namely that English provides separate lexico-grammatical devices for expressing the formation of an intention and for expressing an intention possessed antecedently. If Phyllis rings up Philip and asks him whether he'd like to come to the cinema this evening, he might express his spontaneous intention even as he forms it by saying "Yes, I'll come". A thought of the same form might also pass through Philip's mind when he is browsing through the paper looking for a film he'd like to see and chances on one that catches his fancy. Notice that the linguistic devices used to express or attribute antecedently possessed intentions are completely inappropriate here.[15]

[13] In Gollwitzer's use (1996, 287ff.; 1999, 493ff.), "planning" seems to be more or less equivalent to the formation of subordinate intentions of a highly specific type: what he calls "implementation intentions". These are intentions whose content is indexed with fairly concrete specifications of the conditions – in particular the time and place – under which the superordinate intention is to be realised.

[14] Michael Bratman's analysis of the concept of intention runs under the heading "the planning theory of intention", according to which to intend is necessarily to plan. Bratman (1987, 29) defines plans as "mental states involving an appropriate sort of commitment to action". At the same time, he sees plans as more complex than "relatively simple intentions". These two claims are compatible from a functionalist perspective, for which the essence of intending consists precisely in its characteristic involvement with farther-reaching dispositional – and, in Bratman's conception, normative – structures seen as constitutive of the commitment at the core of plans.

A second sense of "plan", in which plans are sometimes claimed to be constitutive of intentions is that of a representational structure involving both an *aim* and a *way* of realising that aim. According to this suggestion, one of the peculiarities of intentions is precisely that their *contents* are necessarily plans. In other words, the representation of the way or means in which the aim is to be achieved is not the content of a further subordinate intention. For this use, see Brand (1984) 153f., (1986) 213ff. Mele (1992a, 144) follows Brand. However, he also claims (1992a, 136) that the "plan" component that constitutes an intention's content can be the mere representation of a basic action. This would seem to make "plan" simply the term for intention's content, whatever that might be.

[15] W. Sellars (1966, 105ff.) proposed the technical use of "shall" as an expression of what he took to be the irreducible attitude of intention, which he contrasted with "will" as the expression of a

To conclude this brief sketch of our ways of conveying intentions, it is worth holding on to two linguistic points that easily remain unnoticed alongside the striking parallels with the expression of belief. The first is the peculiarity just mentioned: that language furnishes different means of expressing the *formation* and the *antecedent possession* of an attitude. This could be seen as an indication that the aetiology of the attitude is of particular importance for understanding of intentions. I shall argue that this is indeed the case.

Secondly, alongside the predictive uses of "will", "'ll" and "shall", linguistics text books (Quirk et al. 1985, 229; Palmer 1987, 138ff.) generally distinguish a usage they call *willingness*, where "I'll φ" expresses an optative attitude without the supplement of "practical imminence". "I'll do it if you like" is not a conditional intention, but expresses preparedness or willingness (cf. Sect. 4.4.2) that could be transformed into an intention if the antecedent is fulfilled. This might also make us suspect that intending is not that far removed from other forms of optative attitudinising. Again I shall argue that this suspicion is not only justified, but also correct.

6.3 The Conceptual Marginality of Belief

6.3.1 Commitment as Expectation

One straightforward explanation for the use of assertoric linguistic forms to express intentions would be that the supplementary component constitutive of supra-optative commitment is a belief. "I'll φ tonight" could thus be thought to express the conjunction of a want* to φ and some belief concerning one's φ-ing. According to an analysis of this kind, a person moves an attitudinal step nearer to φ-ing by coming to believe that she will φ. And obviously, people do often believe they will do what they intend to do.

However, in spite of an initial plausibility to such a position, an agent's belief that she will φ is a poor candidate to complete the conditions of her intending to φ. Such a belief seems neither to be necessary nor to complete sufficient conditions for intending.

Certainly, an agent's belief that she will φ appears to be too strong to be required in order for her to be accurately describable as intending to φ. Tim might intend to

predictive belief (cf. Grice 1971, 271). In both points – that is, in the combination of the technical device for intention expression and the irreducibility thesis – Sellars has been followed by McCann (1986b, 134) and Brandom (1994, 245ff.). Using "shall" as the general expression for the state of intending is, however, somewhat misleading because of the fact that it is characteristically used to express intention *formation*, whereas we use other devices to express or refer to antecedently formed intentions.

get to church on time, whilst believing, on the basis of his punctuality record, that he *might* not make it. Were the belief that one will φ to be constitutive of the intention to φ, such a scenario would be incoherent.

Conversely, a corresponding belief-want* pair can be given without the attitudes' bearer intending to perform the action that is represented in the contents of the two attitudes. Del, for instance, knows he ought to stay at home this evening and work, but at the same time strongly desires to go to the pub. On the basis of his normative judgement, he forms the intention to, indeed he is "determined" to, stay at home. However, aware of his frequent failures to resist burning desires, he believes he will eventually crack and go for a drink. But his belief together with the desire it is based on doesn't imply a corresponding intention. Indeed, the belief-desire pair appears compatible – if uncomfortably so – with the intention not to realize their contents.

One problem that this counterexample exploits is that we don't need to infer anything about how likely we are to act on the basis of previous experience in order to form intentions. A moment's conscious thought about whether to φ often suffices without us needing to reflect on our behavioural dispositions. Moreover, the phenomenology of intention formation would be equally misrepresented by a position according to which intention requires reflection on present motivation. Again we don't need to introspect until we realize what it is we want* most, before committing ourselves to perform the action that would bring that about.

There is a response to objections of this kind that goes back to Anscombe and Hampshire (Anscombe 1957; Hampshire and Hart 1958, 1ff; Hampshire 1970, 131, 1975, 53ff.). It consists in the claim that the relevant belief – or knowledge[16] – that one will φ has to be non-inferentially acquired. But this is very difficult to make sense of. Certainly, there is a lot to be said for the traditional views according to which we have specific kinds of non-inferential beliefs, perhaps knowledge, concerning our present state – that we exist or that we are in pain. But how doxastic or epistemic states whose contents concern our future actions could have this character appears quite mysterious.

There is, nevertheless, a strand in Hampshire's presentation of his case for this particular brand of non-inferential knowledge that is worth dwelling on for a moment. He presents such practical "knowledge" or "certainty", as he and Hart call it, as the result of deciding after a process of practical deliberation (Hampshire and Hart 1958, 4ff.; Hampshire 1970, 108; 1975, 76). It ought to be phenomenologically uncontroversial both that an agent has unmediated access to the content of her decisions to act and that the kind of mental state in which that access consists is to be distinguished from that acquired by discovering some feature of herself. But it seems equally phenomenologically clear that deciding to act is not a matter of developing a belief, whether inferentially or non-inferentially. Certainly, once an agent is acting intentionally, there is an everyday sense in which she generally "knows what she is doing". Even if the agent is actually mistaken, she no doubt has beliefs concerning her action, beliefs that might qualify, at least to some extent, as non-inferential. But

[16] Cf. note 7 above.

the contents of decisionally formed distal intentions are badly qualified to be the object of such non-inferential beliefs.[17] As mentioned at the beginning of Section 5.2, being the bearer of an optative attitude brings with it the disposition to believe that one wants* whatever is that want's* content. If one has played host to the event of active attitude formation that is a decision, it is presumably more likely that related beliefs will follow. But such beliefs are no more unmediated than any other higher-order beliefs. Moreover, it is a contingent matter whether the agent does end up developing them. Where the intention in question is not the product of a decision, the likelihood of an accompanying belief may well be significantly lower.

The claim that we arrive at a certain kind of – practical – "certainty" through decision is, I think, not only correct, but important for an understanding of intention. A theory of intention should tell us both what such certainty consists in and how it relates to cases of intending that do not result from deciding. A theorist who claims that that certainty is epistemic is faced with a dilemma: if the relevant belief is taken to be supplementary to an optative core of the attitude, it seems that it must be in some way be dependent on the agent's doxastic relation to his own motivation and thus inferential. The construal then faces objections of the kind illustrated by Del. If, on the other hand, the belief is taken to be non-inferential, it seems to be independent of the agent's optative attitudes. It is thus natural that construals of intention as involving non-inferential beliefs take intention to be exclusively doxastic or epistemic. But such construals are faced with the question of *why* agents should acquire such beliefs. Deliberation may lead you to conclude that you have the conclusive reasons to φ. But how do you get from there to the belief or knowledge that you will φ? It seems that we need to add a premise concerning your reliability in doing what you believe you have conclusive reasons to do. But such a premise can only be acquired as a result of registering your own track record. Moreover, adopting such a premise doesn't only involve making an inference. It involves an inference that not everyone will be able to make. Such a position therefore makes the ability to intend entirely contingent and presumably restricts its applicability to certain sorts of people.

David Velleman has advanced an ingenious theory built around an answer to the question of why an agent should acquire an epistemically understood intention, which he equates with "the state of being decided upon one's next action" (Velleman 1989, 112). According to Velleman, we are motivated to form expectations of our own actions because, where they are true, such expectations contribute significantly to our self-understanding and because the desire for self-understanding is central to being a human agent (1989, 35f., 49f.; 61–4; 1992b, 141f.). Intentions are thus beliefs of a special kind, formed not on the basis of sufficient evidence for their truth, but on the basis of the belief that their formation will lead to their truth.

[17] Both Anscombe (1957, 1) and Hampshire (1970, 102) assume that the same criteria must be at work in the concepts of intentionally acting and prior intention. This assumption is anything but self-evident.

Because their believed self-fulfilling character gives them a special status among assertoric attitudes, Velleman has in more recent writings abandoned talk of beliefs and come instead to label them “directive cognitions” (1996, 195) or “cognitive commitments” (2007, 209). Put simply: we form intentions because we believe that we would otherwise be “baffled” about what we end up doing, and that is something to which we are strongly averse. According to Velleman, being possessed of and motivated by this aversion is constitutive of agency.

Velleman’s position appears appropriate to the phenomenology of intention *formation* in several respects. Firstly, although he accepts that there may be an inferential component to the formation of our intentions, he denies that this is necessary and, above all, that it is sufficient. Secondly, his conception provides an explanation of why, although the content of our intentions is frequently what we are, prior to deliberation, most motivated to do, that need not be the case (cf. Sect. 8.2.3). The want* to do what will enable us to make most sense of ourselves brings in a supplementary source of motivation that can sway us to do something we were less strongly motivated to do before deliberating.

The latter point is a particular strength of the position because it offers an explanation for our sense that our taking a stand in deliberation on the options at issue can make a difference to what we end up doing. However, this strength with respect to deliberation’s upshot is achieved at the price of skewing the relationship between the attitude whose tokening terminates deliberation and the reason we enter into deliberation in the first place. We deliberate, as Hampshire and Hart claim, because we are uncertain about what to do and see deliberating as the way of resolving that uncertainty. But uncertainty about *what to do* is not uncertainty about *what one is going to do*.

Moreover, if intending to φ really were a cognitive commitment to the proposition that one will φ, it would be unclear why an intender should concern herself with making sure that she realises her intentions. Sometimes realising an intention’s content can turn out to be a strenuous affair. Where this is the case, an agent has to work at making that content true. But if she believes all along that the proposition that she will φ is true anyway, why should she bother to make the effort? Velleman’s answer is that such effort is itself the *effect* of the belief in conjunction with the agent’s desire to know what she is doing (1989, 44–51). This fundamental desire will be satisfied if the belief that she will φ turns out to be true. So we work at realising our intentions in order not to have been mistaken in our estimate of what we are going to do. But that makes sticking to your intentions the manifestation of a peculiarly self-opinionated stance. The tendency to stick to an opinion merely because it is one’s own is, one would think, not a particularly desirable character trait. Surely, expending energy in order to make sure you stick to an opinion for that reason is neither rational nor reasonable. However, sticking to our intentions, except where we take ourselves to have a good reason to re-enter deliberation, is characteristic of rational agents (cf. Sect. 7.2.3).

Finally, the phenomenological advantages of Velleman’s conception of intention formation pale in the face of the counterintuitive nature of the construction with which he supports it. Everyday agents neither believe that their intentions are self-fulfilling beliefs nor that they realise them in order to secure self-knowledge. On the first point: sports psychologists and self-help coaches claim that autosuggestion can lead to success in the relevant areas of life and there is presumably something to their claim. However, where someone enrols in some such training programme and benefits from it, that will be because the special techniques involved – amongst other things attention focussing and regular repetition of relevant phrases (cf. Orlick 1990, 113) – effect a change in the person's abilities. But if intentions were self-fulfilling beliefs, we would all have such abilities already. Everyday experience tells us, however, that this is not the case. Secondly, none of us believe that all our everyday actions are significantly motivated by a want* for self-knowledge. Indeed, not only do we not have the belief; we are also, as Rae Langton has pointed out (2004, 258), disposed to believe that no such want* is playing any such role in the genesis of our behaviour. This doesn't disprove Velleman's theory, but the phenomenology – where its strengths at first appear to lie – turns out to be stacked against a position according to which we are all systematically confused about what we are doing whenever we perform an action.

6.3.2 Weaker Doxastic Conditions

In spite of the fact that people often expect to do what they intend to do, such an expectation is no essential component of intending. This negative result raises the question of whether intending might involve weaker doxastic conditions. There has been a considerable amount of ink shed on how strong any such doxastic condition would have to be. I remarked in 6.2 that what an addressee should conclude from Tony's expression of the intention to meet Antonia tonight will depend on various contingent facts about Tony and his relationship to Antonia (as well as on his relationship to the addressee). This is, I think, because only highly attenuated doxastic conditions attach to intending. The only condition that has a conceptual role to play, I will be arguing, is *negative*, namely the lack of a belief that one won't perform the action intended. Before giving an account of the status of the condition and suggesting how we should explain it, let me first briefly give reasons for rejecting other candidates.

It is helpful to visualise the scale on which the doxastic candidates are localised. Positively, someone's intending to φ might seem to involve:

(B1) the belief that she[18] will φ;
(B2) the belief that there's a significant probability (perhaps over 0.5) of her φ -ing;
(B3) the belief that it is possible for her to φ.

[18] The personal pronoun in the content of the agent's belief refers directly, unmediated by the representation of properties, to the attitude's bearer. This is the same representational format at work in being motivated (cf. Sect. 2.3.2).

For each of these, there are corresponding exclusions of defeating doxastic conditions:

(NB1) the lack of the belief that she won't φ;
(NB2) the lack of the belief that there is a significant probability of her not φ -ing;
(NB3) the lack of the belief that it is impossible for her to φ.

To these we can, for the sake of completeness, add beliefs with two further related contents sometimes suggested. The first concerns the agent's ability to perform the action, the second her knowledge of a way or a means of doing so:

(BA) the belief that he is able to φ,
(BM) the belief that he knows how to φ.[19]

BM can be easily dismissed. A belief that one has a way or means of doing something one intends is clearly an attitude that might be only acquired as a result of considering how to realise an intention one already has. Excluding the negation of such a belief is also unnecessary. Someone might acquire the intention to bring about *p* whilst being fully aware that she as yet has no idea of what means are suited to bringing *p* about. All that seems required is that she doesn't believe herself incapable of acquiring the relevant knowledge. A bearer of the latter belief must, if she is thinking at all straight, conclude that it is impossible for her to do the deed, thus contravening the weakest doxastic condition, *NB3*. Such straight thinking would also lead her to conclude that she won't φ, thus infringing *NB1*.

An analogous point is relevant for the rejection of *BA*, in as far as it is distinguished from *B3*. A claim that someone is able to do something at *t* is fairly naturally understood as a claim about the actions open to them at *t*. But an agent may at t_1 be unable to φ at t_1, whilst being capable of acquiring the ability to φ at t_2. And in such cases, where the agent possesses the mere second-order capacity to develop the ability to φ, he may still intend to φ. A person at present unable to speak some foreign language can intend to have an elementary conversation in that language at the end of the summer once she has completed a language course. Thus, the belief in a relevant present first-order ability is no condition of intention. This argument is equally applicable to the exclusion of a corresponding belief with negative content.

[19] Most of the positions that are logically possible have been advanced. For *B*1: see Sellars (1966, 126), Kaufman (1966, 39), Chisholm (1970, 644ff.), Grice (1971, 278), Fishbein and Ajzen (1975, 12), Harman (1975/76, 432ff.; 1986a, 90ff.), Kim (1976, 259f.), O'Shaughnesy (1980, II, 305), Pears (1985), Velleman (1989, 109; 1996, 195; 2007, 204ff.), Setiya (2007, 663f.), Broome (2009, 79ff.); for *B*2: Audi (1973b, 65; 1986, 25ff.; 1991, 362), Davis (1984, 134ff.); for *B*3: O'Shaughnessy (1980II, 305), Adams (1986, 288); for a weakened variant of *B3*, "the belief that it may be possible to" φ: Peacocke (1985, 71); For *BA*: Armstrong (1968, 149f.), Hampshire (1970, 134) and Wallace (2001, 105f.); for *BM*: von Wright (1971, 101ff.) (who doesn't differentiate between *BA* and *BM*). For negative conditions, see Mele (1992a, 146ff.), Bratman (1987, 17f.; 38ff.; 1999a, 241) and Brand (1986, 214ff.).

Turning to the various estimates of probability codified in *B1* to *B3*: the divergences within both philosophical and everyday opinion should make us wary of seeing intuitions here as the key to any substantial philosophical thesis. Intuitions as to what can coherently be said apparently diverge radically (cf. Harman 1986b, 366f.). Certainly, people *tend to feel some sort of resistance* to intention ascription where success seems a very thin possibility, a matter of luck or largely dependent on others (Mele 1989, 102; 1992a, 148; 2003b, 131). People don't normally say they intend to win a lottery. But why not? Is such an intention ascription *incoherent* or simply *misleading* as a result of conversational implicature? The problem seems to me to be of the latter kind.[20] If someone were to express an intention to win a lottery, we surely wouldn't be at a loss to understand what she *meant*, but would tend to wonder why she opted to express it in this particular way.

Standard cases of intention expression are presumably located somewhere within the purview of *B2*. But it is hardly plausible that we can identify an even vaguely determinate point on the subjective probability scale at which statements of intent are suddenly deprived of meaning. "I know I've only got a one in ten chance of sinking the putt, but I still intend to" seems comprehensible enough. I suspect that different individuals allow their intention expressions to coexist with different subjective probabilities, perhaps even varying according to their moods. If this were correct, it would obviously indicate that we are not dealing here with genuinely conceptual constraints.

Where we want to indicate a lack of confidence that we will succeed, we often tend to say we merely *hope* to φ; where our confidence only extends to our ability to perform some action subordinate to φ-ing, we often talk of *trying* to φ (cf. Sect. 3.2.1). Harman suggests, plausibly enough, that hope "apparently excludes belief" (1975/76, 437), by which he presumably means that someone who hopes that *p* cannot believe that *p* (cf. Searle 1983, 32). However, this doxastic condition, which Harman thinks is one big difference between hope and intention, is nothing of the sort. Rather, it is a basic conceptual condition common to hoping, intending and trying. You can not only not hope that *p*, you can neither intend nor try to bring it about that *p* if you already believe that *p*. However, once we turn to levels of subjective probability below certainty, what we find are, on the one hand, characteristic usages, on the other hand, usages that seem odd, inappropriate or misleading without being false.[21] Saying you "hope" to achieve something may in one context give rise to the belief in your interlocutors that you believe you have little chance of achieving it. Nevertheless, a person may, so it seems, hope to achieve something they think they will almost certainly achieve – for instance, where achieving it would be particularly important to them. Moreover, as Grice pointed out explicitly, we often ascribe "tryings" – to others and ourselves – after unsuccessful

[20] Here I agree with Adams (1986, 287f.) and Holton (2009, 23, 37).

[21] These are the kinds of case that led Grice to introduce the pragmatic concept of implicature (Grice 1967/87, 4, 9). I stated, very briefly, in Section 3.2.1 why I don't think that the perspective-relative doubt brought into play by talk of trying is merely implicated, rather than conceptually entailed. This will be important in Section 7.2.1.

attempts, irrespective of how likely we had thought success beforehand. Something similar seems to be true of retrospectively self-ascribing “hopes”. There can thus be no question of separating neatly intendings, tryings and hopings on the basis of different subjective probabilities. On the contrary, there are many action-guiding states of mind that can be equally appropriately described or expressed *by any or all* of these terms.

The only clear constraint seems to be that revealed by utterances of the form: “I’m φ-ing this evening, although I won’t be able to φ this evening”, “I intend to φ, although it is impossible that I will φ” and “I’m going to φ, although I won’t φ”. All of these are genuinely incoherent (cf. Hampshire 1970, 134), as we can simply make no sense of what kind of mental state the utterer thinks he is expressing or ascribing to himself. It seems, then, that it at least has to be excluded that someone intending to φ might believe the negation of *BA* (that he has the ability to φ), of *B3* (that it is possible for him to φ) or of *B1* (that he will φ). An agent’s lacking the ability (of the relevant order) to φ entails the impossibility of their φ-ing, which in turn entails that they won’t φ. Conversely, one can believe that one won’t perform an action in spite of believing that one is able to perform it and that no other factors, such as lack of opportunity, make one’s performing it impossible. Therefore, negations of the modal beliefs are covered by *NB1*. Thus the lack of the belief that one won’t φ – for whatever reason – seems to suffice as the doxastic condition we are after.[22] When turn to an explanation of the condition in the next section, however, we will see that even this is too strong as a general conceptual condition on intention.

In the literature, two kinds of example have been employed to justify a negative, rather than a positive requirement. The first kind of case can be termed *postdeliberative agnosticism*. It seems possible that an agent might, after reflection, decide to perform an action, whilst having no definite opinion as to whether she will achieve, or even try to achieve what she intends (Bratman 1987, 37f.). Someone who knows herself to be forgetful, particularly when she is under stress, might intend to φ whilst being aware that her forgetfulness could lead to her not taking the necessary steps at the appropriate time. This person might simply not weigh up the factors that would support the belief that she will φ against those that speak against it. If so, then it is possible to intend to φ whilst withholding any assertoric stand as to the probability of one’s φ-ing.

Secondly, there are surely intentions formed by agents – perhaps spontaneously or habitually – *without their having even considered the question* of whether φ-ing is possible for them. Here, the absence of belief is not a matter of a considered withholding of an assertoric stand, but the absence of any thought that constitutes a relevant assertoric attitude. There may be no question that the agent would develop the relevant feasibility belief were she to reflect on the matter. But if no such deliberation has taken place, that step may simply not have been carried out. For instance,

[22] The content of the excluded belief can, alternatively, be thought of as specifying a form of agent-relative modality: the impossibility-for-me of my φ-ing against the background of certain assumptions I have no disposition to question. We will return to this point in the discussion of the intention-consequential requirements in 7.2.

it is surely not unlikely that someone unexpectedly seeing an old friend approaching on the street will acquire the intention to say "hello" without considering what the chances are of him failing to get the word out (cf. Mele 1992a, 147ff.).

The plausibility of such cases supports – although it obviously doesn't prove– the claim that any doxastic condition attaching to intending will be merely negative. Richard Holton has suggested that not having considered the question whether one will φ is incompatible with intending to φ (Holton 2009, 28). This may appear to be the case if the only intentions one considers are formed as a result of practical deliberation. Certainly, intentions with a deliberative aetiology are central, indeed paradigmatic cases. However, they are not the only intentions we have. And where not even minimal deliberation precedes intention formation, it is unclear why considering whether one will perform the action in question should be required. It seems then that, of the two arguments that tell against a positive doxastic condition, one applies to deliberative cases, the other to nondeliberative cases. This is, I think, telling. A theory of intention should come to grips both with the importance of the deliberative genesis of many intentions and with the fact that not all intentions have such an aetiology. This will be one of the main aims of the constructive theory to be developed in chapters to come.

Before I turn to the important role of practical deliberation for an understanding of intention's relation to belief, it is worth recalling that in Chapter 3 we already encountered negative doxastic conditions attaching to certain optative attitudes. There (Sect. 3.1.1) I argued that everyday talk both of "being motivated to φ" and of "wanting to φ" implies, or at least implicates the absence of pretty much the same defeating doxastic conditions. Note, moreover, that this is even true of merely "hoping to φ". Hoping that *p* generally suggests that the possibility one assigns to *p* coming about is significantly low – at least relative to the level of importance *p* has for us. Nevertheless, we don't hope for something of which we are certain that there is *no* chance of it coming about.[23] That is where the semantic switch to (at least one use of) "wishing" takes place. If the merely proto-practical attitudes of hoping and (everyday) wanting already have some such minimal confidence condition, that again supports our conclusion that the eminently practical character of intending is unlikely to depend substantially on its relation to belief.

6.3.3 Intention, Belief and Practical Deliberation

The move to such attenuated doxastic conditions shifts the import of the question of belief's role relative to intention. Whereas the main attraction of the conception

[23] The same is surely true of actions appropriately seen by their agents as tryings. This is argued by O'Shaughnessy (1980II, 40), Hornsby (1980, 40f.) and Adams (1986, 288). Mele apparently believes this is not so: he thinks that an agent can (intend to) try to hit each of two targets, where he knows it is technically impossible to hit both, as in Bratman's video arcade case (Mele 2003b, 130).

of intention as expectation lies in its apparent potential to explain the idea of commitment, there is little plausibility to the claim that this function might be fulfilled by weaker beliefs. Why, then, might there be any doxastic conditions associated with intending at all?

I think we get a little help answering this question if we turn to a brief argument from the third book of the *Nicomachean Ethics*. According to Aristotle, “sane” deliberation (“bouleusis”) can only be about “things that are in our power” or “doable through the agency” of the deliberator (NE 1112b 31ff.). As the objects of our deliberation and of our decision (“prohairesis”) are the same (NE 1113a3; 1113a10), “there is no decision for impossible things” (NE 1111b20).

Three brief remarks before we proceed: firstly, although Aristotle’s formulations are ontological, concerning the possibility of performing the relevant action, I think we can still get Aristotle to help us with our epistemic question. “Sane” deliberation will not concern questions one believes to be outside the purview of one’s agency. Should it turn out that a state of affairs had, unbeknown to me, been immune to my influence, that would invalidate a presupposition of the deliberation, but not compromise its sanity or rationality. Second, Aristotle’s formulations are primarily positive, that is, they correspond broadly to *BA*. Nevertheless, I am going to simply apply his basic idea to the negative condition I have argued is relevant. An epistemic version of the last quotation in the preceding paragraph is what I shall be working with. Third, as remarked at the end of Section 1.3, Aristotle’s terms “bouleusis” and “prohairesis” both pick out specifically ethical sub-sets of what we understand by the terms “deliberation” and “decision”, i.e. deliberation carried out, and decisions taken with the aim of living an “eudaimon” life. I will simply ignore what appears to be the implication of this view, viz that non-ethical variants are somehow structured differently.

Now, if the relevant doxastic condition holds for practical deliberation, if, as Aristotle claims, the objects of practical deliberation and decision are identical and if deciding is a paradigmatic way of generating intentions, it follows that there are doxastic restrictions on what we can come to intend in those paradigmatic cases in which intentions are generated by decisions. Such restrictions derive from restrictions already in place with respect to those want* contents that are *mere candidates* for intention contents. In as far as there are doxastic conditions on paradigmatic cases of intending, then, they exist because they carry over from conditions on practical deliberation.

This insight, I want to suggest, should be combined with close attention to the difference between conceptual and rational conditions on the generation of decisional intentions. Close attention reveals that this difference is a further point at which the importance of conscious attitudinising should be recognised. My claim is that *NB1* can function either as a conceptual or a rational constraint on decisional intention formation and decisional intending. The difference between the two kinds of constraint hangs on the question of whether the attitudes involved are conscious or not. This difference also carries over from constraints on practical deliberation.

If someone consciously believes that she has a zero probability of φ-ing during the period of time under consideration, she simply cannot deliberate practically about whether to φ. She might wonder whether she would have good reasons to φ or whether it would be best if she were to φ, but these reflections will, strictly speaking, be species of theoretical inquiry.[24] The essence of practical deliberation is the enlistment of conscious thought in the service of our action.[25] So we cannot take ourselves to be practically reflecting on whether to do something we consciously believe lies outside the purview of our influence.

It is, however, not incoherent to deliberate on whether to φ if you believe that you cannot φ, as long as that belief is unconscious, for instance, if the belief has temporarily slipped your mind. Deliberating on whether to φ whilst being the bearer of a temporarily inaccessible belief that one won't φ or that one's φ-ing is impossible is, however, a rationally unfortunate state to be in, indeed a state that we are rationally required to avoid. If this is correct, then practical deliberation on whether to φ both entails and rationally requires the lack of belief that one won't φ, the difference being that the conceptual requirement concerns conscious belief, whereas the rational requirement involves no such restriction.

On the basis of Aristotle's point, both the conceptual and the rational conditions carry over to the genesis of those paradigmatic intentions formed as a result of deliberation. It is impossible to come decisionally to intend to perform some action you consciously believe you won't perform and it is rationally required of you that you don't decide to do what you think you won't. This has consequences for decisionally formed intentions themselves.

However, the consequences are not completely straightforward. This is because, although decisional intentions are necessarily formed consciously, the attitudes thus formed will frequently not themselves remain conscious. If an agent decides in the morning to go shopping after work, but in the meantime acquires the belief that the meeting he has to attend in the afternoon will last until the shops are shut, he might retain his intention for a time, indeed he might retain it for too long, causing him and others various problems. Such cases show that the intention isn't necessarily revised as a result of acquiring a belief in its unfulfillability, although it rationally should be. However, it would be impossible for an agent to be the bearer of both mental states, if they were both conscious. Such an agent would, like the putative deliberator with a conscious belief that the object of her deliberation is unattainable, lapse into incoherence.

NB1, then, is a rational constraint on decisionally formed intentions, but a conceptual constraint only on conscious intentions thus formed where the relevant belief is itself conscious. The incoherence of sentences such as "I'm going to φ, although I won't φ", in which the first conjunct expresses an intention and the second a belief, is explained by this construal. It grounds in the fact that the belief

[24] For an explanation of the difference that I am appealing to intuitively at this stage, see Section 8.5.3.

[25] The "unconscious inferential mechanisms" discussed in Section 5.1.6 should thus not be seen as driving forms of practical deliberation. There is inference without deliberation.

whose content is asserted in the second conjunct cannot be a non-conscious belief of the speaker at the moment at which he is sincerely asserting its content (cf. Sect. 4.1.1). These kinds of sentences therefore give us a certain insight into relations between intentions and beliefs. The relevant relations are, however, only necessarily relations between the attitudes under certain, restricted conditions.

Aristotle's construal of the relationship between the doxastic conditions on deliberation and intentions acquired through decision tells us something significant about what it is to intend. Aristotle was right that the incoherence of "setting yourself to" do something you – consciously – believe impossible is of the same nature as the incoherence of deliberating on whether to do something you – consciously – believe to be impossible. This gives the negative belief condition considerable more weight than in the case of the distinction between "wanting" and "wishing", where it appears to be a mere linguistic contingency. If someone says they "want to have gone to the concert yesterday", we would understand what they are saying, but mentally note that they have failed to master a detail of the way the English language works. If, however, they were to say that they are deliberating about whether to φ, whilst they believe they have literally no chance of φ-ing, we would be forced to conclude that they simply don't know what they are talking about. This would be equally true for agents who claimed to intend to φ whilst believing that they won't φ.

We are making progress with decisional intentions. However, this raises problems for an understanding of non-decisional cases. If decisionally and non-decisionally formed intentions are attitudes of a type defined by its tokens meeting the same conditions, that is, if the unity of intention thesis is true, then the conditions on the deliberative cases also apply to nondeliberative cases. If, however, the explanation of the specific rational and conceptual doxastic conditions on (some) decisional intentions is their deliberative aetiology, then it would be a big coincidence if the same conditions were to apply to intentions acquired in other ways. Should it turn out that the unity of intention thesis is false, there would be little reason why the same doxastic conditions should apply in both cases. I will be suggesting that it is and that they don't.

6.4 Doxastic Symptoms of Decisional Intending

I have rejected the claim that intending to φ entails believing one will φ and argued both that forming the decisional intention to φ excludes consciously believing that one won't φ and that playing host to the decisional intention thus formed rationally requires the lack of any such belief, conscious or not. If that is correct, then a theory of intention should be able to offer some explanation as to why the positive conditions *B1* and *B2* have appeared so plausible to so many philosophers, whose command of the English language is surely no less well developed than that of their unbelieving colleagues. I have already mentioned two possible sources of this error. The first is the failure to distinguish between expression and assertion, as a result of which it can appear that "I'll φ" is an assertion about the speaker's future action. The second source, at least for *B1*, might lie in the quest for some further psychological component that would make sense of the "commitment" component that transcends mere optativity.

There are, however, two further relations between intention and belief which may have added to the plausibility of a stronger conceptual connection than can be vindicated. Both have the status of symptoms and may be counted as part of "the intentional syndrome". Both are, moreover, most clearly given in paradigmatic cases in which intentions are acquired decisionally. The first appears to belong to a *subrationally produced* constellation of effects of deciding. The second is the result of rational *inference from experience*.

6.4.1 Postdecisional Doxastic Bias

According to a considerable body of evidence that has been amassed by motivational psychologists testing the so-called "Rubicon model" of action phases, beliefs on the lines of *B1* and *B2* fit into an overall pattern of change in the doxastic profile of agents who have decided to perform some action. According to this model, deciding is a "point of no return", which marks the transition from one "mind set", "cognitive style" or "type of psychological functioning" to another (Heckhausen and Gollwitzer 1987, 103; Heckhausen 1991, 175ff.; Gollwitzer 1991, 38ff.; Achtziger and Gollwitzer 2008, 273ff.). Deliberation prior to decision is said to be governed by the standard decision-theoretic principles of establishing the desirability and probability values of options to be weighed up against each other, a process during which the person is receptive to any information that could turn out to be relevant for the decision. In contrast, bearers of a postdecisional "implemental mind set" take on a significant doxastic bias that marks both the *selection* and the *processing* of information.

As regards *selection*, it seems that deciding, particularly where the decision is supported by the planning of implemental steps,[26] installs in the intender a "doxastic filter". This filters out a great deal of information that could plausibly be detrimental to the realisation of the decision's content: in particular information that could conceivably justify the reconsideration of the decision, either on the basis of desirability or feasibility. As regards the *processing* of the information the agent has to work with, this leads to the formation of beliefs that favour the implementation of the decision. This latter feature has been labelled "illusionary optimism" (Gollwitzer and Kinney 1989, 540; Gollwitzer 1990, 77; Taylor and Gollwitzer 1995, 225;

[26] As the bulk of recent research here has focussed on the effects of detailed planning, there is some unclarity as to how many of these effects result from the mere decision that sets the end, the means to which are subsequently planned. The "implemental mind set" is "induced" through mere decisions in the early studies Heckhausen and Gollwitzer (1986, 1073), Heckhausen and Gollwitzer (1987, 104), Beckmann and Gollwitzer (1987, 265ff.) and Gollwitzer and Kinney (1989, 532ff.).

Gollwitzer 2003) and has been shown, at least under certain laboratory conditions, to include the agent's beliefs about her own more general abilities and about her control over events around her. For instance, it was demonstrated that subjects who had just committed themselves to some goal significantly over-estimated the extent to which the switching on of a light under experimental conditions was caused by their own button pushing (Gollwitzer and Kinney 1989, 536ff.; Gollwitzer 1990, 67ff.; 1996, 300ff.; Gollwitzer and Bayer 1999, 405ff.; Gollwitzer 2003, 262f.).

Even assuming that these empirical results are borne out in further studies, they obviously don't answer conceptual questions. In our context, nevertheless, they are important for two reasons. Firstly, they show that there is coherence between a whole set of postdecisional doxastic phenomena and the kind of beliefs that have often been thought to be in some sense definitive of intention. Secondly, they provide further symptomatic evidence for the special nature of the step involved in committing oneself *by deciding*.[27] Heinz Heckhausen, who first developed the "Rubicon model", indicated this by means of a terminological distinction that for a motivational psychologist appears somewhat paradoxical: he terms the postdecisional attitudinal profile "volitional" in order to distinguish it from the agent's predecisional condition, which he labels "motivational". (Heckhausen and Gollwitzer 1986, 1072; 1987, 103; Heckhausen 1987, 123; Achtziger and Gollwitzer 2008, 276). Were the motivational-volitional distinction to be an exclusive disjunction in nature, this would entail that intending is not a motivational state.[28] Whether or not Heckhausen intended such a reading, it would be congenial to irreducibility theorists.

6.4.2 Inductive Beliefs

If certain doxastic features of the "implemental mind set" are indeed "illusionary", then this may only be the case up to a certain point. We humans may tend to overestimate our abilities and to underestimate obstacles after deciding, tendencies that are by and large useful features of our psychological economy. But we also have *grounds* to believe that there is a higher probability of us doing what we have decided to do than there is of us doing something we only desire to do.

One way to see this is to compare the beliefs we bring about in others if we express the hope, the desire or the intention to do something. We would generally be likely to bet more on Tony meeting Antonia tonight if he has said he is "going

[27] This point is neglected in the otherwise perspicuous use Richard Holton makes of the data on "illusory optimism" (Holton 2009, 6f.).

[28] It should be noted that the same body of literature contains formulations that suggest the opposite view. Take the claim of Taylor and Gollwitzer (1995, 213): "The implemental mindset is assumed to induce participants to muster motivation, resources and cognitions in service of goal-directed actions". If, as I have claimed, the mustering of motivational force is the function of wants*, such a claim can be seen as supporting the thesis that intending is indeed a special kind of wanting*.

to" do so than if he says that he "hopes" or "wants" to do so. This is because we generally take the first phrase to indicate that he has taken an attitudinal step beyond everyday wanting or hoping that makes its bearer more likely to bring about the attitude's content.

Notice, though, that the tendency to such belief formation is defeasible, depending, amongst other things, on who we are dealing with. In the case of certain individuals und particular circumstances, we may not lay very much store at all on them coming to do what they intend to do. There is thus a certain level of induction involved in our tendency to form beliefs about the future behaviour of people to whom we ascribe intentions. It seems plausible that the expectations we form about our own future behaviour on the basis of what we intend are also up to a point subject to influence by induction on the basis of past correlations.[29]

A general, and generally rational effect of deciding to φ is the acquisition of a belief that one will φ with some level of probability, where the probability in question is higher than the probability the person either assigned to her φ-ing or would have assigned to her φ-ing before taking the decision. The belief in question may be explicitly comparative, but it certainly need not be.

This means, by the way, that we cannot expect any precise information as to what level of probability an intender will assign to her realising the content of her attitude. The probability level will depend on a whole set of contingent factors, not only on the level she would have assigned before her decision, but also on how difficult the intended action is thought to be and how confident she happens to be feeling at the time. This may be one of the reasons why there is such variation in the estimations of the strength of the doxastic requirement.

A final reason why this contingent connection to intending may be felt to be necessary is that there is perhaps a certain empirical necessity at work here. To see what this might consist in, imagine a person whose intention formation generally makes no difference to what she believes she is going to do. It seems highly plausible, from what we know of human nature, that such a person would with time lose a hold on what intending is. Perhaps someone suffering from depression might have lost all confidence in the power of her decisions to move her to act. Then her episodes of postdeliberative intention formation might leave her doxastically unmoved. Should this causal connection remain interrupted over a longer period, then perhaps she may find that she no longer understands what she is doing when she decides and thus lose the capacity for decision. If this speculation were correct, then there would be a certain kind of necessary connection between intention

[29] It is primarily these postdecisional effects, rather than the predecisional doxastic conditions, of intending that play a significant role in the inter- and intrasubjective coordination of our actions that Bratman (1987, 31f.; 38f.) sees as essential to intending. Grice (1971, 277ff.) claims that an inductively formed belief is conceptually required for intending. The belief in question is the belief that one will φ, formed on (i) the evidence of one's having an "unrivalled" want* ("willing") to φ, (ii) previous experience that φ-ing results from one's willing to φ in the absence of any interference and (iii) the belief that nothing will interfere with one's φ-ing. The conjunction of the unrivalled want* and the belief it causes he terms "intention".

and beliefs concerning one's future actions. It would be determined by empirical conditions of the possibility of upholding the capacity for intention formation over time.

6.5 Summary: Postdecisional Commitment and Belief

Examination of the relationship between intending and believing has yielded the following results: first, intention formation through decision, like deliberation, excludes the agent's conscious belief that she won't realise the intention's content. Second, decisional intentions are subject to the rational requirement to avoid conjunction with any such beliefs that their contents won't be realised. As deliberation generally includes reflection on questions of feasibility, it is not unlikely that the agent will be also the bearer of an explicit belief in her capacity to realise that content. Third, intention formation, as a matter of empirical fact, tends to increase the probability with which the agent believes she will do the intended deed. This is presumably at least in part a result of induction from past experience. Finally, if the empirical hypotheses associated with the Rubicon model of intention formation are correct, the formation of beliefs of this kind is also strongly favoured by a general postdecisional increase in optimism as far as the intender's general agential capacities are concerned. These latter two points plausibly go some way to explaining why intention is often accompanied by the willingness to predict that one will accomplish the relevant deed.

It is clear from this discussion that the gap between the generic concept of wanting* and that of intending cannot be explained by doxastic conditions. For postdecisional belief, explanation runs in precisely the opposite direction: the step to "commitment" or "settledness" taken in deciding at least partly explains the postdecisional shift in subjective probabilities. That the prior absence of defeating doxastic conditions is only a precondition of intention formation can be seen from the fact that the same conditions have to be absent for the person to enter into meaningful deliberation, irrespective of whether she comes to a decision on the matter. The positive assertoric attitudes that are characteristically connected with having taken a decision are *symptoms* of the presence of commitment; they are no *part* of the commitment itself. The reductionist, then, has to look elsewhere.

Chapter 7
The Intentional Syndrome: Characteristic Causal Features and Rational Requirements

The discussion of intention and belief has led to the conclusion that, in spite of intending's characteristically strong effects on its bearer's beliefs, intention has no doxastic requirement with conceptual status, although conscious, deliberatively formed intentions do need to satisfy a weak belief condition. Such a minimal condition is not going to help explain what it is about intentions that makes them appear so different to everyday desires. Clearly, if this condition were all that were to be added to wants* in order to generate intentions, then intending would not even *appear* to be a candidate for the status of an irreducible attitude.

But intending is such a candidate. The doxastic requirement on rational intending together with intention's characteristic doxastic effects are first indications that this is so. Although, unlike mere "wishing", "wanting" also involves the absence of defeating doxastic conditions, it doesn't seem that it typically generates positive beliefs concerning the probability of its content coming about. These doxastic features are a first group of components in a set of phenomena that can be dubbed the *intentional syndrome*. The purpose of this chapter is to catalogue the symptoms of intending, to describe them with some precision and to mark the ways in which they stand out against the syndrome of phenomena associated with the generic attitude of wanting*. This will allow us to distinguish those features that pose a genuine challenge to the project of reduction from those easily accommodated within the broader optative syndrome.

Phenomena of two basic kinds are candidates for the role of symptoms that distinguish intention from other kinds of wants*. The first concern intention's causal environment, the second its normative consequences. In the first part of the chapter, I shall discuss features of the former kind under the heading of *intention strength*. Approaching these phenomena by way of the distinctions between kinds of want* strength discussed in Section 5.2 enables important differences from the optative syndrome to emerge: on the one hand, there is one kind of attitudinal strength characteristic of wanting* that the English language doesn't pick out by talk of

N. Roughley, *Wanting and Intending*, Philosophical Studies Series 123,
DOI 10.1007/978-94-017-7387-4_7

"intention strength"; on the other hand, three further forms of attitudinal strength attributable to wants* are generally instantiated fairly high up relevant scales in the case of intentions.

Taking these features together suggests that intention's characteristic causal environment has a certain level of specificity. However, this doesn't seem sufficient to warrant the postulation of an entirely distinct attitude. Stronger support for the irreducibility thesis is provided by the *normative constraints* to which an intender is necessarily subject. In the second part of the chapter, I discuss the strictures that apply relative to intention's tendency to *lead to action*, its tendency to *pervade* its bearer's mental life and its tendency to *persist*. All these dimensions are covered by requirements of rationality, deontic structures that have been intensely discussed in the recent literature. I will argue for specific formulations of the requirements, along the way rejecting the claim that they necessarily supervene on the mind, but supporting the claim that they take "wide scope". The normative embedding of intentions in such specific structures proves to be the primary datum that the reductionist has to explain.

7.1 Characteristic Causal Features: Dimensions of Intention Strength

In Section 5.2.2, I distinguished three broad sorts of want* strength. Two of these, the tendency of a want* to lead to action that realises its content and the tendency of a want* to persist, can be thought of as distinguishable, although frequently complementary notions of motivational force. The third, hedonic strength, is a matter of connections between wanting* and either affective experience or tendencies to affective experience. Causal or dispositional connections of these kinds also characterise intentions. However, the language by means of which we express and ascribe intentions picks out relatively specific constellations of these factors. Up to a point, this may be because, when we talk of intentions, certain features of wanting's* causal environment are of no interest to us. However, it also seems clear that intentions are optative attitudes that are typically characterised by particular sorts of strength.

7.1.1 Hedonic Intensity

It has been claimed, in particular by defenders of the irreducibility thesis (Meiland 1970, 76f.; Kim 1976, 254; Brand 1984, 125), that intention, unlike "desire", does not permit of gradation at all. This claim has in turn been seen as identifying a symptom of intention's special status. Intending, it has seemed to some, is an all-or-nothing affair: one enters into the state of intending by taking a step, paradigmatically by deciding. Correlatively, so it has seemed, we don't get there by simply increasing want* strength.

We need to see what is correct about this claim, but also why it doesn't take us where the irreducibility theorist wants to go. Certainly, a number of adverbial qualifiers we apply to "want" are not applicable to "intend": we cannot intend "desperately" or "badly" to do something. However, the absence of certain adverbial qualifiers by no means shows that the state denoted by the verb "to intend" is not gradable. Indeed, one does not need to look very far to realise that there are other adverbial intensifiers that do locate intending on a scale of more or less. Most obviously, people sometimes say they "strongly intend" to do something. Moreover, we can also "fully" or "firmly intend" to perform some action.

The explanation is that there is one particular kind of strength characteristic of optative attitudes that is not ascribed or expressed by the language of intention, namely hedonic strength. To be precise, it is intensity of the higher-order forms of hedonic discomfort Duncker termed the "pain of exclusion" (Sect. 5.3.3). Someone who occurrently "desperately" or "badly" wants to do, experience or have something, or, as we can also say, "craves", "longs for" or "is dying for" whatever it is, is experiencing a significant intensity of negative hedonic affect at[1] the non-realisation of his want's* contents. Talk of "strongly desiring" something often picks out wanting* some *p* accompanied by negative affect at what one takes to be *p*'s non-realisation. But we see this experiential dimension as irrelevant where we are ascribing intentions and their characteristic forms of strength. In other words, "strongly" at least sometimes picks out different kinds of properties when used in the phrases "strongly intend" and "strongly desire".

This point is closely related to another fact that has sometimes over-hastily been taken show intention's irreducibility. This is that intending to φ does not entail being the bearer of an everyday desire or want to φ (Brand 1984, 122f.; cf. Locke E II, xxi, §36). The reason there is no such entailment is the same reason why neither "being willing to" nor "consenting to" φ in instrumental, moral or institutional cases implies everyday desiring to φ (Sect. 4.4.2). Whereas the everyday term "desire" generally refers to a compound of wanting* and hedonic features consonant with the want*, this is the case neither with "willingness" nor with "intention". We can consent to a course of action without desiring to adopt it just as we can intend to do something without desiring to do it. And we can come to intend to do something we don't desire to do, because we have consented to do it. Note though that, whereas "willingness" tends to be an optative stand in the face of countervailing wants*, usually wants, talk of "intention" is completely neutral in this respect.

Does, then, the inapplicability of the adverbs "desperately" and "badly" to the verb "to intend" tell us something about what intentions are? I think caution is in order here. Certainly, talk of "intention' does not allow us in the same verbal phrase to pick out the hedonic property of experiencing second-order discomfort at the current non-satisfaction of an optative attitude. But that doesn't mean that intentions

[1] I leave open the interesting question of whether this "at" is genuinely intentional or merely causal.

cannot stand in the same causal relation to “pains of exclusion” as many other optative attitudes, for instance everyday desires. If, for example, it is true of Des both that he intends to see Amy and that he desperately longs to see her, then Des is implausibly the bearer of two distinct states, one of which is hedonically gradable, whereas the other is not. On the contrary, the optative conception of wanting* suggests that we understand him as the bearer of a want* that is compounded both with discomfort at its present non-satisfaction and with whatever it takes to give us the commitment component of intention. What I am proposing, then, is that, when we talk of Des’s “intention”, we are simply picking out a constellation of these factors that is different from, yet overlaps with those picked out by talk of his “longing”. In fact, it is perfectly conceivable that Des “setting himself” to see Amy might itself be the source of a *specific, particularly cutting feeling of discomfort* during the time in which he can’t see her. The fact that we don’t have a specific term for cases of intending accompanied by a particular pain of exclusion from what the intender is committed to bring about seems philosophically unimportant. The point is that we *could* have one, as the compound state it would designate is not only conceivable, but surely occurs empirically.

Although the English language has no one term to refer to compounds of intention and affect, we are fully aware that *intentional action* tends to lead to particular forms of pleasure and displeasure. Two such hedonic phenomena were termed by Sidgwick the “pleasures of attainment” and the “pleasures of pursuit” (ME 46ff.). Pleasures of *attainment* are the particular feelings of satisfaction (Sect. 5.3.1) that may result from attaining an end one has set oneself. Perhaps these do have a specific “feel” to them as compared to the “satisfaction” felt when something one wants* is brought about by other means. Certainly, this is plausible for the pleasures of *pursuit*, forms of positive affect that tend to accompany one’s active striving to reach some goal. Much of the pleasure that can be gained from playing sports is of this kind. The characteristic increase in motivation in the course of a sporting activity often seems to result from the particular pleasure gained in striving to achieve some aim such as running a certain time, reaching a certain peak or beating one’s opponent. It is surely at least in part due to the specificity of the pleasures thus gained that sportsmen and -women will not generally be prepared to exchange them for pleasures gained in other ways.

It ought to be obvious, however, that there is no general connection between such pleasures and the conscious pursuit of intended goals. Indeed, the phenomenon of “flow”, the experience of complete immersion in some activity, often cited as the consummate experience of goal pursuit, is plausibly understood as resulting from the activity’s detailed control being handed over to non-conscious directive processes.[2] In other words, what the psychologist Csikszentmihaly has characterised

[2] David Velleman has emphasised this point, which he sees as supporting his cognitive conception of intention (Velleman 2007, 213f.). As I have argued in reply (Roughley 2007b, 227f.), I believe that the data and theoretical gloss on flow provided by the psychologist Csikszentmihaly are far less hospitable to Velleman’s theory than he thinks.

as the optimal experience of sporting performance (Csikszentmihalyi 1990) seems to be achieved by minimising the contribution of conscious intention at the time of performance, instead delegating much of what one does to automatic processes.[3]

There would certainly be no plausibility to a claim that there is some generally determinate relationship between intending and hedonic experience. To the contrary, acting on a decisional intention is perfectly compatible *both* with particularly intense affective experience and with hedonically neutral reactions on the realisation, or definitive non-realisation of an intention's content. On the one hand, where someone has committed herself to pursuing some project and has, as a result, formed further attitudes that either support or presuppose that intention, its frustration is likely to bring more intensive feelings of disappointment than in other cases. On the other hand, intention also leaves open the possibility of the converse pattern of emotional reaction – precisely because of the possibility of intending to do something one has no everyday desire to do. Someone may fully intend to go a conference, because he feels unable to disappoint the organiser, but be relieved when a train drivers' strike prevents him from doing so. In this respect, intending is, unsurprisingly, like acceptance or assent in many instrumental, moral or institutional cases: if you were reluctantly willing to go to the shops and, after forming a corresponding intention, suddenly realise that they are already closed, you might find yourself more relieved at not having to make the effort than frustrated at the prospect of a weekend without muffins.

7.1.2 Motivational Strength and Persistence

Let us now return to the question of what is meant by the locution "strongly intend". I shall suggest that there are three distinguishable kinds of strength that can be picked out by such expressions, two of which I have already discussed for wants* in general (Sects. 3.3.4 and 5.2.2).

The most obvious candidate can be labelled "motivational strength in the narrow sense", that is, the extent to which certain of the agent's physiological properties dispose him to realise, or attempt to realise a want's* content.[4] Ascriptions of motivational strength clearly contain a significant comparative component: the motivational strength of some specific want* is generally ascribed relative to that of competing wants* the agent is disposed to leave unrealised in realising, or attempting to realise the relevant content. However, because exhaustion, illness or

[3] On automaticity in action, see Sections 9.5.2 and 9.5.3 below.

[4] I am assuming that, should precise physical correlates of optative representations turn out to be identifiable, these would not include all those physiological processes whose occurrence determines motivational strength. In this assumption, I diverge from Mele (1998, 32f.). Cf. the discussion of the two dimensions of motivation in Section 2.5, particularly the characterisation of motivational force.

depression can lead to the general reduction of a person's motivation and because we can compare the overall motivational conditions of different people, the motivational strength of an agent's want* cannot be simply defined in terms of a relation to the dispositions to act mobilised by competing wants* of that person. If someone is more strongly motivated to φ than to do anything else, but is so exhausted that the slightest difficulty is likely to cause her to abandon the project of φ-ing, there is a clear sense in which her motivation to φ is weak. For this reason, any attempt to give an explicit account of "the *extent* to which an agent is disposed to realise or attempt to realise a want's* content" would not only need to take account of competing wants*, but would also have to include reference to external standards.

Now, if someone avows that she "strongly" or "fully" intends to φ, what she means may be, firstly, that it will take considerable obstacles or subjectively weighty counter-reasons to prevent her from φ-ing. "Motivational strength" is a functional concept that appears equally applicable to everyday wants, desires, longings and intentions. A second feature that an agent may be referring to in saying that she "strongly intends" to do something may also be picked out if she avows that she "firmly intends" to do it. She may be affirming that her attitude is to a significant degree *resistant to change*.

The difference between the two forms of strength is, with intentions as with "desires", discernible from the fact that the two properties can vary independently of each other. Someone might be the bearer of a desire or intention that is determined only to be of short duration – perhaps it is the effect of some passing situation or of the consumption of some drug – but which only the strongest of disincentives can prevent her from realising during that short period. Conversely, a person might for a long time have had a desire or intention to φ at some point in his life, in spite of the fact that there have, up to now, always been other, perhaps fairly minor motivational factors that have effectively stood in the way of his φ-ing.[5]

Intentions can thus, like other wants*, be qualified both by greater or lesser degrees of persistence, just as they can muster greater or lesser degrees of motivational force in the narrow sense. Nevertheless, intentions may appear to be characteristically located towards the upper end of both scales and thus to be wants* that are realised more often than others, both because they tend to stay around long enough for realisation opportunities to arise[6] and because they typically move their

[5] Note a linguistic phenomenon that can easily contribute to conceptual confusion here, eliding the distinction between these two types of strength: someone who is strongly motivated to bring about *p* may be said to be "persistent" in her attempts to do so. This "persistence" (perseverance) of the person is to be distinguished from the "persistence" (continued existence) of the motivating attitude, which may or may not lead its bearer to persevere in actions aimed at its realisation. The two dimensions are also tested for together by the standard operational criteria for motivational strength in social psychology: the tenacity with which someone sticks at a task and the resumption of work on the task after interruption (cf. Chartrand and Bargh 2002, 19ff.).

[6] In other words, intentions typically have what Harman terms "inertia" (Harman 1975/76, 446, 450).

bearers to grasp those opportunities.[7] Intentions may thus seem to be situated at the opposite end of both scales to "whims", optative attitudes that can just come and go and may even fail to muster a minimal level of motivational force. However, this need not be true of any particular individual intention: a spontaneous intention formed to look at my watch or to ask someone a question during a conversation may quickly be forgotten and therefore be short-lived and thus remain unrealised. For this reason, the fact that intentions tend to be among the optative attitudes situated towards the top end of both scales is no candidate for what makes an attitude token an intention and a fortiori no candidate for an explication of commitment.[8]

It is worth noting a second reason why persistence may appear to be a particularly, even criterially important feature of intentions. This impression may derive from a focus on distal, or "future-directed" intentions, rather than on their proximal, or "present-directed" cousins.[9] Focussing on the former variants is, like focussing on decisionally, rather than spontaneously or otherwise unspectacularly formed intentions, going to provide more material for an investigation to work with. But this epistemic advantage is clearly no guarantee that the characteristics thus uncovered are universal features of everything labelled by the term. Indeed, if it is true that intentions tend to be characterised by motivational strength in the narrow sense of the term, then proximal intentions are rarely going to persist at all. Thus, if proximal and distal intentions are indeed attitudes of the same kind, it would not be particularly informative to say that attitudes of this kind are generally possessed of both motivational strength in the narrow sense and persistence.

Still, once we turn to distal intentions, it is clear that, as Michael Bratman has repeatedly pointed out (1987, 16ff., 65; 1995, 36; 1999b, 4), they do tend to have a notable level of inertia or stability. What is particularly striking is that this tendency to persist involves two distinct mechanisms. Firstly, intentions are much less likely than other wants* (Sect. 5.1.2) simply to dissolve without having been realised. The most prominent way in which unrealised intentions come to an end is through a change of mind of their bearers; intentions tend not to dissolve simply as a result of their being forgotten, although this does happen. On a first level, then, intentions tend to persist because of a *resistance to nondecisional dissolution*. For a person to change their mind, it is necessary, as I shall argue in Chapter 9, for her to

[7] Bratman marks the fact that the bearer of an intention to φ is considerably more likely to φ than is the bearer of an otherwise unqualified want* to φ by means of the distinction between "conduct-controlling" and "potentially influencing" attitudes (1987, 16).

[8] Pace network functionalism (Sect. 3.3.2), I take it that conceptual analysis should help us to understand what it is in virtue of which individual attitudes are such as to belong to the relevant attitude type.

[9] On this distinction, see Bratman (1987), 4f., 108 and Mele (1992a), 159f. Bratman criticizes what he calls the "strategy of extension" that understands intention beginning with the paradigm of realised proximal intention. Such a strategy, he points out, suggest a model of belief-desire reduction that seems a lot less plausible when distal intentions are placed at centre stage. In Chapter 10, I shall suggest that the inverse strategy of extension from distal to proximal intentions threatens to be equally misleading.

re-engage in at least a minimal level of deliberation on the practical question at issue. The second persistence-conducive mechanism characteristic of intending is what Bratman has termed *resistance to reconsideration*: we tend only to reopen deliberation as to whether to rescind some earlier decision where we take it that we have a relatively weighty reason for doing so. Resistance to bringing about necessary conditions of decisional intention shedding obviously entails resistance to decisional or deliberative intention dissolution.

This dual barrier against dissolution certainly appears to be highly specific. It has appeared to some to speak strongly for the irreducibilty thesis. Once again, however, if we consider the whole breadth of the phenomena that a theory of intending has to cover, surely it is just *too specific*. There is nothing in the concept of intending that necessitates that a particular level of persistence characterise any individual case. I might intend to make a not-very-important telephone call later this morning, but easily be brought to abandon the intention by considerations of minimal weight, such as preferring to continue doing whatever it is I'm involved in at that moment. For this reason, it is a mistake to identify the "commitment" constitutive of intention with "resolve" or "determination" to perform the intended action (cf. McCann 1986a, 193; 1991, 197; Gollwitzer 1990, 57; 1996, 289). When people say they are "resolved" or "determined" to act, they are ascribing to themselves an intention with specific levels of strength in both the motivational and persistence dimensions.[10] But neither of these forms of strength can be constitutive of intending itself, or of the "commitment" component that is added to optativity. Rather, they presuppose it.

7.1.3 Pervasion

A final form of strength that can be predicated of intentions as of other wants* is the *extent of the attitude's influence* on a person's thought and perceptual processes. Like other wants* (Sect. 3.2), intentions can leave traces in more extended or restricted spheres of a person's action and non-agential behaviour.

Perceptual Salience

A non-agential feature of the optative syndrome that lends itself to comparison with corresponding symptoms of intending concerns perceptual salience and what psychologists call the "accessibility" of thought contents: wanting* some *p* generally

[10] Richard Holton's characterisation of "resolutions" as intentions formed with the specific purpose of overcoming anticipated temptation (Holton 2009, 9ff., 77ff.) is narrower than the everyday conception of being resolved. The tenaciousness someone might have in pursuing a goal may be a matter of resolve in the face of obstacles even if the relevant intention was not formed *with the purpose* of overcoming them.

increases the probability that one will have thoughts about or perceive *p*, or states of affairs conducive to bringing *p* about, be sensitive to indications that *p* has come to pass and to lexical items associated with *p*.

One way of approaching the specificity of intending's effects on perception and thought is via a comparison of the effects exerted by "needs", "desires" and "values",[11] as reported in the older studies of Bruner, Postman and associates (cf. Sect. 3.2.2), with corresponding features of the "implemental mind set" focussed on in the "Rubicon model" of action phases (cf. Sect. 6.4.1).[12] Of course, because the two groups of studies are not designed to permit comparison, there are limits to what conclusions can be drawn. Nevertheless, it is worth describing the main points of similarity and dissimilarity between their results. Placing the studies side by side provides support for the view that decisionally formed intentions share certain perceptually relevant causal characteristics with other optative attitudes, whilst also unfolding their own specific kinds of effects.

Both strongly "desiring", "needing" or "valuing" some item and decisionally intending some action tend to increase the *speed of recognition* or *extent of recall* of the relevant items (Bruner and Postman 1947/48, 75ff.; Postman et al. 1948, 150f.; Postman and Leytham 1950/51, 398ff.; Beckmann and Gollwitzer 1987, 276ff.). It seems, then, that both kinds of pro-attitude lead to forms of epistemic bias. This first rough parallel, which suggests treating the salience effects of deciding as specific variants of the corresponding feature of the optative syndrome, may seem to be in tension with the way the relationship between the first two phases of the "Rubicon model" is interpreted by its advocates. According to Heckhausen and Gollwitzer, the predecisional "mind set" is characterised by "impartiality", "open-mindedness" and "realism" relative to the options, a perspective they contrast with the postdecisional orientation, in which agents become "narrow-minded partisans of their plans of action" (Heckhausen and Gollwitzer 1987, 103; Gollwitzer 1990, 65; Gollwitzer 2003, 264). However, the "impartiality" in the "deliberative" phase concerns the way deliberation is carried on, not how salient the options are in the first place. Nevertheless, it is conceivable that entering into deliberation about some wanted* option could have effects on how that option is "seen", since beginning to deliberate about something involves a step back from the flow of action, suspending its immediate motivational effects. It would be interesting to know whether the salience effects of certain wants* – for money, for food – taken in isolation (Bruner and Goodman 1947, 49f.; Bruner and Postman 1948, 206f.; McLelland and Atkinson 1948, 218ff.) decrease when a choice has to be made between their objects and the object of some other want* of comparable strength.

[11] The authors of these studies were unconcerned with the differences between these concepts. "Value" is clearly meant to be the most general term, subsuming both "need" and "desire" (cf. Bruner and Postman 1948, 207). I take it further that "need" is being used in a way that excludes its non-attitudinal interpretation.

[12] At one point (1990, 84), Gollwitzer warns against "confusing" the approach to "cognitive tuning" within the framework of the "Rubicon model" with that of the "New Look" psychology of perception. My question is how the results of the two sets of studies might be related.

Of more direct relevance for the confrontation of wanting* and intending is the fact that having decided to do something tends to correlate with the *inhibition of thoughts about the alternative* thus rejected and with the recollection of negative properties of that alternative, where the agent is forced to think about it (Beckmann and Gollwitzer 1987, 276ff.). These tendencies seem to result from the unique practical status conferred on the attitude that results from decision. Gollwitzer expresses this consequence with a quote from Jones and Gerard, according to whom making a decision "stops the 'babble of competing inner voices'" (Gollwitzer 1990, 62). In the light of intending's special connection to action, the processing of information concerning the objects of competing wants* is dysfunctional and thus tends to be inhibited. This at least seems a plausible explanation.

Although the experiments in the Rubicon paradigm are silent on this point, the notion of an alternative at issue presumably only picks out those possible contents of intentions on which an agent doesn't decide to act, not those on which she decides not to act. Considering that someone is only likely to form an intention with negative content where she has some optative attitude with the performance of that action as its content, an intention to refrain from acting seems unlikely to mobilise less attitudinal and motivational resources than an intention with positive content. I am not aware of experiments designed to test the difference in effects on perception or thought processing of negative and positive intentions. The parallel question was looked at by the (old) "New Look" psychologists. Their findings indicate that aversions* can *reduce the recognition threshold* for items presented by a tachistoscope to the same extent as, if not even more markedly than wants* (Postman and Leytham 1950–51, 399). Were similar effects of negative intending to be discovered, that would show that wanting* and intending share further significant causal mechanisms.[13]

Finally, the most specific epistemic features of the "implemental mind set" are strikingly *formal* in character. That is, it seems that the subrational effects of making a decision can include phenomena that have nothing to do with the *content* of the particular intentions thus formed. Alongside global tendencies to self-overestimation and "illusionary optimism" (Sect. 6.4.1), having decided can apparently correlate with a *general* decrease in the span of one's short-term memory (Heckhausen and Gollwitzer 1986, 1077f.) and an increased tendency to perceive others as committed to realising *some goal or other* (Gollwitzer et al. 1990, 1122f.). It is implausible that these content-independent features could typically have analogues where other optative attitudes are concerned. As people generally want* something or other most of the time, there would appear to be no non-optative state with which content-independent effects of wanting* anything whatsoever

[13] In one much-publicised case, it has been shown that intending not to do something leads to a significant increase in the probability of having thoughts closely connected with what one intends not to do. The case is that in which the intention not to think of a white bear has the effect of leading by a process of "ironic control" to thoughts of white bears (Wegner 1994). Clearly, this is a very special case, as the object of the intention is the refraining from a thought.

could be contrasted. Thus, such effects certainly support the idea that something more than optativity is at work in intention.

However, there are two considerations that recommend caution. The first suggests that, although the phenomena in question are not typical of the optative sphere in general, they may nevertheless occur in connection with other attitudinal phenomena with an optative dimension. The second raises the question of whether these formal phenomena may not turn out to be characteristic of only certain subspecies of intention.

The idea that these content-independent effects may be more widely distributed suggests itself when we look at the emotions. It is a commonplace, and surely a plausible one, that people who are depressed, happy or angry are disposed to perceive the world differently than people in a fairly neutral mood. And joy, sadness or anger about some particular thing all seem to cause us to think less about other things we want*. Moreover, neither the modification of perceptual speed or short-term memory span nor the tendency to project one's own view of things onto others appear particularly unlikely as the effects of emoting.[14]

The converse worry – that the "implemental mind set" delineated in the Rubicon model may not be representative of intention in general – is raised by a number of features of the experimental designs employed. Most importantly, the intentions in these cases are all decisionally formed. However, intentions can be acquired in various other ways, for instance, spontaneously or gradually, as when someone's desire to pursue some goal imperceptibly "crystallizes" over the years, yielding a corresponding intention (cf. Sect. 9.2). It is therefore conceivable that the effects that have been catalogued are the effects of *decisions*, but not of intentions in general.

Moreover, perhaps these are only the effects of decisions that fulfil certain conditions. Many of our decisions are taken after only a minimal amount of deliberation, and certainly without requiring any great resolve. Is Sam's decision, after a mere moment's hesitation, to take a ham rather than a cheese sandwich, likely to be accompanied by "illusionary optimism", the belief in increased control, the decrease in his short-term memory span and the tendency to see others as "set on" acting? I don't know, but I would be not a little sceptical. Significantly, the relevant effects have been demonstrated above all in cases in which the *content* of the decision is some personal concern of the experimental subject and where the subject engages in the production of *subordinate intentions* detailing the means of the decision's realisation.[15]

[14] Ferguson (2000, 111ff., 116–121) reports studies which demonstrate the effects of emotions on perception, attention and memory. The "mood-congruent effects" reported here are also – unsurprisingly – not limited to features of the environment connected to the *contents* of the subjects' emotions, but largely concern apparent instances of, or evidence for, environmental presence of the same *types* of emotional states. That is, people who are sad or happy are prone to seeing others as sad or happy, independently of whatever they are themselves sad or happy about.

[15] Cf. Section 6.4.1, note 26. A more recent conceptualisation of the relation between the general effects of the "implemental mind set" and the specific effects of forming subordinate "implementation intentions" is set out in Achtziger and Gollwitzer (2008).

Finally, it would also be interesting to know whether the same effects can be registered if the course of action the agent decides against is either the object of a want* with particularly high *motivational strength* or represents something the agent particularly *values*. One would suspect that the smoker who has decided against smoking in order to accommodate his hosts might suffer precisely the opposite salience effects. And the same may be true of someone who guiltily decides to smoke in spite of believing that it would be best not to.

Further Intentions: Generation and Eschewal

In Section 3.2.1, I discussed the fact that wanting* *p* tends to generate further wants* in an agent. Secondary – epistemic, imaginative and expressive – wants* presumably also result from intending, although intention's more intimate relationship to action will frequently obviate an agent's need to acquire additional motivation to find out whether the attitude's content has been realised. It might also lessen the tendency to fantasize or talk about it, although there are likely to be cases in which the converse is true. Any such differences would, however, appear to be entirely contingent and likely to be dependent on contextual or individual specificities.

What does appear to mark intentions off from other optative attitudes is their distinct tendency to generate subordinate attitudinal conspecifics. An intender will be likely to support her intention with further intentions whose realisation would in the situation be necessary, sufficient or conducive to the realisation of the first intention. She will also be likely to avoid coming to intend actions the realisation of which would be incompatible with other intentions. The relevant mechanisms of generation and avoidance are in part automatic. They are also, importantly, in part reasoning-based. Bratman has argued that intentions "involve" dispositions to reason in ways such as to produce supporting intentions and avoid undermining intentions (1987, 108f.).

This seems broadly correct, but again, only broadly. The claim that intentions "involve" such dispositions cannot be understood – and, in line with his functionalism, Bratman doesn't understand it – as pinpointing a necessary or sufficient condition on individual intentions. That it is not necessary can be seen from cases in which undermining intentions are not eschewed, as where someone decides to catalogue her spice rack when she has reserved the day for work on a paper the deadline for which is uncomfortably near (Sabini and Silver 1982, 134ff.). The everyday phenomenon of procrastination testifies to the fact that there is no incompatibility between intending to φ and not forming further intentions conducive to one's φ-ing or even forming intentions subversive of the primary intention to φ.

Moreover, the tendency to form further attitudes that increase the likelihood of a primary optative attitude's realisation is clearly not sufficient to make that primary attitude an intention. Take the case of an agent with an organising desire to be respected in the community in which she lives. Such a person will, so it seems, not only be consciously or unconsciously on the look-out for information as to her standing, as well as for opportunities to behave in ways she deems likely to result in her gaining, keeping and increasing the respect she desires. She will also be disposed to eschew behaviour that will be detrimental to the satisfaction of her desire and to plan to act in ways that will contribute to her being respected. The fact that a desire has these dispositional consequences does not seem to entail that the desire is an intention. Think about how a desire with a comparable set of effects might be an integral feature of an emotional disposition such as fear, as where someone's life is significantly structured by fear of not being taken seriously. Again, in such a case, it seems fairly clear that the optative dimension of the person's basic fear won't be an intention.[16]

Nevertheless, Bratman is surely right that intentions typically exert a particularly pervasive influence on the further attitude formation of their bearers in a way that distinguishes them from everyday wants and desires. The question is whether this has anything to do with what it is that makes intentions intentions. The fact that we are only dealing with a typical consequence and not a feature in virtue of which individual attitudes belong to the class might seem to suggest that the answer is no. However, it is clearly significant that we don't only have a label – "procrastination" – for cases in which such typical consequences are missing, but that the label's applicability also involves a negative evaluation. The characteristic patterns of subordinate intention generation and further intention eschewal are bound up with clear, indeed with strict standards of evaluation. That is special.

7.2 Deontic Consequences: The Intention-Consequential Requirements

Procrastination and the *IC* Requirements

There is a whole body of empirical psychological literature that focuses on issues connected with procrastination. According to the literature, procrastination divides into two distinct kinds, one concerning the formation of a relevant intention and one concerning action to realise an intention.[17] When things go wrong as a result of procrastination, either an action or the formation of an intention is delayed so long that the action ends up being significantly suboptimal in its efficacy or quality or even not being performed at all. In such cases, something goes awry either with the motivation associated with the intention or with the mechanisms responsible for its pervasion of the agent's psychic economy. There are cases of

[16] This point holds whether one takes it that wants* can be components or merely accompaniments of emotions (cf. Sect. 4.2, note 12).

[17] The distinction is labelled by some authors as that between "behavioural-avoidant and decisional procrastination" (Orellana-Damacella et al. 2000, 226). Schouwenburg, in his survey of the particularly preeminent field of academic procrastination, distinguishes "an intention-behavior discrepancy and a lack of promptness in intending to perform" (Schouwenberg 1995, 72).

the latter kind in which the agent fails so spectacularly to eschew undermining intentions that empirical psychologists have felt the need to coin a special term, "self-handicapping". In such cases, agents appear to go out of their way to decrease the probability that they will attain their goal or do so at a satisfactory level, as when a student parties into the early hours of the morning before an exam. Cases of this ilk look so much like deliberate self-undermining that they have encouraged the search for psychodynamic explanations in terms of further unconscious goals (Higgins 1990).[18]

Both procrastination and self-handicapping appear to be highly paradoxical forms of behaviour. The reason why this is so takes us to the heart of the specificity of intentions relative to other optative attitudes. In the words of a team of empirical psychologists, the agents in question "seem to maintain an avoidance relationship with their goals" (Orellana-Damacella et al. 2000, 228). In as far as a person's goals are propositions she wants* to bring about (cf. Velleman 1989, 112), her avoiding doing so is not what one would expect. Still, we can construct cases in which someone desires, even strongly desires some end, but is also motivated to avoid attaining it. Something someone desires can be irrealisable together with something else he desires, perhaps more strongly, as a result of which he prioritises the second end, whilst continuing to desire the first. And there can be things someone desires so strongly that he is motivated to avoid coming under their spell, fearing a general loss of autonomy should he embark on attempts to satisfy the desire. Both of these last sentences describe what might be highly rational patterns of attitudinising and behaving. However, if we substitute "intends" and "intention" for "desires" and "desire" in those sentences, the patterns can no longer be rational.

The avoidance of those goals that are the objects of our intentions is paradoxical because such goals are, for conceptual reasons, necessarily aimed at by their bearers. In as far as the goals of procrastinators and self-handicappers are set by intentions, then, both aim at, and in some sense fail to aim at certain results of their actions. In some cases, they even appear to aim at some state of affairs and simultaneously aim not to attain it. We think of someone of whom this description is true as being attitudinally at odds with themselves in a way that goes not just gradually, but categorically beyond the tension involved in wanting some end and wanting not to attain it. An agent may be irrational if he doesn't abandon a want* when time and experience have shown that it is not realisable in conjunction with the realisation of some other want* that he would not drop under any circumstances. Whether that is the case will depend on what he takes the costs to be of upholding it (Sect. 4.2.2). In contrast, the irrationality of certain attitudinal patterns consequent

[18] The term "self-handicapping" is actually used far more broadly in empirical psychology, also covering cases in which an agent has not actively contributed to the increased probability of his failure, but merely attributes it to external factors. The psychodynamic explanations, which often refer to strivings to protect the agent's "self-esteem", could plausibly be made sense of in terms of either the Broad-Sartre strategy or the extended optative conception discussed in Sections 5.1.4 and 5.1.6.

to intention formation looks to be a priori. It is independent of what the experience-based expectations of the results of such patterns' realisation might be.

The subjection of intentions to a priori rational standards is easily the most significant feature of the intentional syndrome that has to be accounted for by the attitude's analysis. There is a clear parallel with belief, which, again unlike wanting*, also comes with a set of standards – for instance, the proscription of belief inconsistency and a principle of belief formation in accordance with modus ponens[19] – to which any rational attitudiniser is necessarily committed. The key question for the reductionist is thus *why* such standards attach to this particular kind of optative attitude. I shall offer my answer to this question in Chapter 10. Before developing that answer, it will be helpful to establish a level of clarity on the standards themselves. In what follows, I shall argue for specific formulations of five intention-consequential standards.[20]

Broome's Three Claims

Practical rationality is an area in which there has been intense and fruitful philosophical activity in the last decade, activity that has been given particularly strong impulses by the work of John Broome. In order to frame my discussion of the intention-consequential requirements (*IC* requirements), let me comment on three claims of Broome's that have played a significant role in structuring the debates here.

The first claim is that there is a categorial distinction between rational requirements and reasons.[21] This can be seen from the fact that an agent may be rationally required to be the bearer of some attitude, whilst having no reason, or at least no reason independent of rationality, to be its bearer. Bela, to whom the truth of q matters, believes that p and believes that if p then q. Bela is rationally required, under these circumstances, to believe that q – irrespective of whether p, if p then q or q are true. So it seems that whether someone is rationally required to take on attitude A cannot in general be dependent on whether she has a reason to do so (Broome 1999, 401ff.; 2001, 105ff.; 2004, 29).

Now, the situation here is somewhat more complicated than it first appears. If the explanation of Bela's belief that p is Bela's intense desire that p in the face of the

[19] As John Broome, following Harman, has pointed out, such a principle doesn't require that we believe everything that we believe would be immediately implied by the truth of our beliefs, only that we acquire those beliefs thus implied whose truth matters to us (Broome 2005, 322; 2009, 64). Setiya thinks we should see such a restriction of the requirement's application as only imposed by non-ideal rationality (Setiya 2007, 665f.).

[20] I will discuss the intention-antecedential norm that relates conclusive reasons judgements and intentions (what Broome has variously called "krasia" and "enkrasia") in Section 8.5.2.

[21] The claim that we should distinguish between rationality and reasonableness, where the former is a matter of attitudinal "coherence or consistency", is prominent in Davidson's discussions of irrationality (Davidson 1982b, 290; 1985, 345f.).

decisive evidence he believes he has that $\neg p$, then Bela may not only have no reason to believe that q; it is presumably also irrational for him, all in all, to have that belief. As John Brunero in particular has pointed out, we need to distinguish within the sphere of rationality between local and all-things-considered judgements (Brunero 2010, 33; 2012, 129f.; cf. Way 2011, 228f.). If Bela's belief formation conflicts with the modus ponens requirement, he is at least locally irrational in believing that p. If his first doxastic premise state (the belief that p) is the mere result of wishful thinking or self-deception, then his believing that p under these circumstances is also locally irrational. His belief that q is then presumably irrational all in all. This is all possible whether or not Bela's belief concerning his decisive doxastic evidence is true or not, that is, whether he has a sufficient reason to, or ought to believe that p.

The a priori character of the requirements of rationality thus has two distinct consequences. First, rational requirements make claims on agents even where the agents appear to have no reason to take on any of the individual attitudes in the requirement's scope. Second, each individual requirement makes a claim on agents even where conforming to it involves the agent maintaining an overall state of irrationality.

There appear, then, to be two kinks in the normativity that applies to individual rational requirements: it may not be the case that we ought to, or even have a pro tanto reason to do what rationality requires of us and it seems that it would sometimes not be overall rational to do what individual requirements require of us.

At this point, let me note an ambiguity in the term "normativity". There is a sense in which individual rational requirements are accurately described as "normative", in that they have a deontic form, make demands on, rather than simply describing persons and come with correctness standards. In another sense, one that John Broome has come to adopt, it is unclear whether rational requirements are per se normative, if "normative" is understood to mean reason-providing (Broome 2005, 325; 2007b, 162f.; 2010, 287; 2013a, 204). I will not be adopting this latter use, as the former has such wide currency beyond the specific debate on the requirements of rationality, a currency that is fairly natural considering the linguistic relation between "normativity" and "norm".[22] Where Broome and others use "normative", I will talk of "reason-giving" or "-providing", or some such.

I agree, then, with Broome that there is an important distinction to be drawn between reasons and requirements of rationality. However, I shall argue in Chapter 10 that rational requirements do turn out to provide reasons, just reasons of a special sort. Moreover, I will also be taking issue with both the second and the third debate-structuring claims he has made. According to the second claim, rational requirements generally take wide, rather than narrow scope, that is, they apply to internal relationships within packages of items, rather than to the individual items themselves (Broome 2004, 29f.; 2007c). Now, this basic idea does help to

[22] It should perhaps also be emphasized that "requirements" are not specific to the sphere of rationality. The concept of a – social or moral – requirement was central to the discussion in Section 4.4.2. I take it that such requirements can provide reasons.

distinguish rational requirements from (other) reasons. Take a simple doxastic case. It is irrational to hold the following three beliefs simultaneously: the belief that p, the belief that if p then q und the belief that $\neg q$. That irrationality can be avoided by altering any one of the beliefs in the package. The wide scope of such rational requirements is standardly articulated by placing the linguistic indicator of requirement ("It is rationally required that …") before the formulation of the conditional's antecedents ("if you believe that p, believe that if p then q"), rather than merely before its consequent ("believe that $\neg q$"). Broome claims that almost all rational requirements take wide scope.[23] While I think – and will argue in Section 7.2.3 – that a central argument against the wide scope view fails, I will also adduce a reason why not all *IC* requirements have *as* wide a scope as 'wide scopers' generally claim.

The reason why at least one of the premise states of certain *IC* requirements falls outside the requirement's scope is also one of the reasons why I take Broome's third important claim to be false. This is the claim that rationality supervenes on the mind (Broome 2007a, 352; 2010, 288; 2013a, 89), a claim whose formulation Broome takes from Ralph Wedgwood. According to this claim, the items between which rationality requires coherence are attitudes of the agent in question. If this is correct, there is no way of altering an agent's rationality without altering her mind; altering the rest of the world without altering features of agent's mind can make no difference. I will be challenging this claim. Everyday judgements of rationality involving our intentions don't seem to me to be quite as restricted as Broome suggests. There are, I shall be claiming in the course of the next three chapters, a number of considerations that count against the supervenience thesis.

7.2.1 The Requirement of Executive Consistency

Rational requirements are requirements of coherence. Coherence is a pretty vague concept. Indeed, one might suspect that the idea of "fitting together" central to the concept must itself be a normative notion. There are, however, central norms of rationality for which this problem doesn't arise. These are norms for which coherence can be specified as consistency, as in the requirement not to have contradictory beliefs. There are two *IC* requirements of this form. I will come to the second in Section 7.2.2. First, I turn to what one might see as the most fundamental intention-consequential norm. In fact, the fundamental character of the requirement has led a number of philosophers to deny that it could be a norm at all.

[23] He mentions one exception, the requirement not to believe p and $\neg p$ (Broome 2005, 323). I would hesitate to call this a requirement, as I think what it proscribes is impossible. Cf. Section 4.1.2.

The requirement in question seems fairly simple and can be given the following preliminary formulation:

EC′
It is rationally required of *X* that
if *X* intends to φ,
and believes that she will never φ if she doesn't φ at *t*,
X at a time she takes to be *t* φs.

A few brief comments on features of *EC′*: first, the requirement is of wide scope and can thus be complied with by not satisfying either the one or the other of the antecedents, as well as by satisfying the consequent.

Second, the personal pronoun in the content of the agent's belief once again refers directly, unmediated by the representation of properties.[24]

Third, the requirement works with a conception of necessity that is relative to a set of – epistemic, volitive and ethical – constraints that the agent takes as given, although they don't have the force of physical or metaphysical modality (cf. Wallace 2001, 109f). *X* doesn't need to believe that there is any strict necessity to her φ-ing by *t* in order for her to φ at all; she need only believe that *she*, for whatever reasons, won't φ otherwise.[25]

Fourth, the requirement is formulated in purely synchronic terms. A diachronic principle that replaced the phrases "at *t*" in the antecedent and "*t*" in the consequent with the phrases "by *t*" and "not later than *t*" respectively would perhaps appear more natural, as it would make explicit the fact that the requirement can be satisfied by φ-ing at any time up to and including *t*. The synchronic requirement, which doesn't make this explicit, can nevertheless be equally satisfied in this way. *EC′* simply picks out the point in time at which things go wrong if the conditions are not simultaneously satisfied. A diachronic version that worked with the replacement phrases would perhaps be closer to the thoughts that might go through the mind of an agent ("I have to make sure that I have φ-d by *t*"). It would, however, be derived from the more basic synchronic version.

Finally, the principle picks out the now-or-never situation in terms of a time *t*. This is perhaps misleading, as the agent may be thinking in terms of a specific situation that he cannot rationally let pass ("the last opportunity to speak to her"), rather than a time that can be specified on some objective scale ("at 3 o'clock"). Nevertheless, the decisive relation remains temporal, as can be seen from the fact that the situation is one whose passing means the agent is too late. For this reason, I will stick to the simple temporal representation.

[24] We have already encountered this representational format in the explication of motivational representations (Sect. 2.3.2) and the discussion of intention's doxastic condition (Sect. 6.3.2). Cf. Broome 2007b, 162; 2009, 64.

[25] Here we meet again the agent-relative modality at work in the rational requirement connecting intention and belief (Sect. 6.3.2).

Now, a number of philosophers are convinced that there cannot be any such requirement of executive consistency. There are two reasons one might have for rejecting it, both of which go to the heart of intention's relationship to its normative environment.

Conceptual, Not Rational?

This first reason grounds in the claim that a principle of this kind articulates a conceptual, rather than a rational requirement. It has seemed to some authors that anyone genuinely intending to φ and fulfilling the doxastic condition will necessarily φ, unless they change their mind, are prevented from φ-ing or have forgotten their intention (Williams 1985, 18). It is the tightness of the connections that can be spelled out here that confers a certain plausibility on the Logical Connection Claim (Sect. 4.4) as developed for intentions, rather than for mere wants* (cf. von Wright 1971, 107). Nevertheless, it is unclear where the compulsion to see the requirement as purely conceptual is supposed to come from.

Broome has offered the following fast argument for the conceptual connection: "An intention to get on a bus is particular sort of disposition to do so. If you are disposed to do some act, you do it unless something prevents you. Therefore, if you do not get on the bus, and nothing prevents you, you do not intend to" (Broome 2008, 105; cf. 2013a, 151).[26] Let me comment on both premises of this argument.

On the first premise: we have seen that intentions characteristically have various kinds of effects. One way of putting this would be to say that intentions dispose their bearers to various kinds of behaviour – to perceive certain kinds of things more readily, to reason in specific ways, to form subordinate intentions and eschew undermining intentions, and to perform or try to perform the action intended. But if we talk this way, we should be clear, firstly, that intentions are not identical with any of the individual dispositions[27] and, secondly, that talk of dispositions may suggest a level of precision that is simply not available. Although we have fairly clear ideas of the kinds of behaviour that would count as manifestations, we don't know anything very precise about the stimulus conditions.

This brings us to the second premise. An entity with a disposition to behave in a certain way will behave in that way if the stimulus conditions and no defeating conditions are instantiated. The advent of a now-or-never belief plausibly counts as a stimulus condition. Note that this condition appears to be exactly analogous to a condition on the rational formation of subordinate intentions – the belief that now is *t*, where one won't perform one's intended action unless one performs a subordinate action by *t* (see Sect. 7.2.2). Now, the conceptual connection theorist claims that,

[26] Broome seems to have changed his mind on this point. In an earlier article he claimed "Unrepudiated intentions normatively require to be acted on" (Broome 2001, 112). This change of mind is independent of his decision to restrict the reference of the term "normative".

[27] Compare my discussion of dispositional analyses of the attitudes in Sections 3.3.2 and 3.3.3.

whereas the disposition to form subordinate intentions requires the further condition that the agent be rational, the disposition to act doesn't. The conceptual connection theorist who argues with dispositions owes us an explanation of why rationality is required to complete the triggering conditions in one case, but not in the other. The appeal to dispositions, then, provides no fast argument for the conceptual connection claim.[28]

The following example seems to me not only to be coherent, but also to be an example of a kind of case that does occur empirically: Ewen has resolved to bring up an awkward topic with his boss. He knows that it is now or never, as he is now in the situation he has brought about to this end: he is sitting in his boss's office. Nevertheless he doesn't manage to broach the subject, allowing the opportunity to pass with eyes wide open. Reflecting on the missed opportunity after the event, he gradually comes to abandon the intention, as he forces himself to accept the fact that he has once again failed to overcome his fear of authority. The conceptual connection theorist insists that Ewen necessarily either abandons his intention at the moment at which he doesn't take what he believes is his last opportunity of realising it or never really thus intended in the first place. As von Wright argued in his early discussion of the Logical Connection Argument, this seems to be mere "dogmatism" (von Wright 1971, 117).

In contrast, I am impressed by the parallel between failures in executive consistency and subordinate intention formation. They seem to me to be equally possible and equally failures in rationality. Indeed, in the light of the guiding claim of part II of this study, that intention is distinguished from wanting* in being eminently practical (Sect. 6.1.1), failure in executive consistency looks like the more basic form.

Supervenience and Trying

I turn now to the second consideration that has motivated objections to a rational requirement of executive consistency. It grounds in the thesis that rationality necessarily supervenes on the mind. If the thesis is true, non-attitudinal features of the world could not make a difference to the rationality of an agent. Actions, although they involve an attitudinal component, go decisively beyond attitudinising. But this involvement in extra-attitudinal features of the world cannot, if the supervenience thesis is true, be covered by requirements of rationality.

In this particular case, the supervenience thesis seems to be supported by an obvious thought: an agent may be prevented from performing some action she

[28] Note further that the dispositionalist also needs to provide an explanation of why the instantiation of the stimulus conditions sometimes leads not to the agent acting, but to the loss of the disposition. The mere presence of defeating conditions – for instance a reverse-cycle finkish environment (Martin 1994, 3) – won't do, as that would be compatible with the disposition's continued existence.

genuinely intends to perform without that reflecting on her rationality (Broome 2005, 323). This would be explained by the supervenience thesis: rationality requires that an agent be psychologically coherent in certain decisive respects, not that she achieve alignment between her psychology and relevant ways the world is.

However, this particular problem with *EC′* can be solved without making as global a claim as the supervenience thesis. Most obviously, one could add an extensional clause outside the scope of the requirement that limits its applicability to cases in which *X* isn't prevented from φ-ing at *t*. The addition would not be ad hoc, as it would simply exclude those cases in which failure to achieve agential coherence has nothing to do with the agent.

Nevertheless, we should reject this solution for the following reason: there might be cases in which what prevents the agent from performing the intended action is psychological interference. In particular, the agent might be prevented from even trying to perform the action, for instance by means of some kind of neurosurgical or hypnotic procedure. And it seems that such cases ought *not* to be excluded from the purview of the requirement: an intervention of this kind would be an intervention in the rational capacities of the agent, i.e. capacities whose functioning is demanded by rationality.

One could try and repair the exclusionary condition by specifying that it only covers non-psychological prevention of the intended action. However, this move would run into a new objection. The requirement should, so it seems, cover both physical and mental action. But if we now exclude cases of psychological prevention from the exclusionary condition, failures to perform intended mental actions will be classified as irrational even where they result from failures in capacity that could not have been anticipated. Someone might, for instance, intend to relive the joys of her youth imaginatively, but on attempting to do so, find that her memory fails her. But such a person would be no more irrational than someone who attempts to move his arm and finds that it is paralysed.

The problem, then, will not be solved by an exclusionary condition placed before the requirement operator. We need a solution that involves the modification of the consequent and that explicitly names the psychological feature that the exclusionary condition attempted to pick out without naming. We need to mention *trying*. Doing so gives us the following formulation:

(EC)
It is rationally required of *X* that
if *X* intends to φ,
and believes that she will never φ if she doesn't φ at *t*,
X at a time she takes to be *t* φs or tries to φ.

Trying is compatible with the action being prevented and with the agent being organised in way that justifies us still seeing her as rational. The idea, then, is that the relevant form of executive consistency is upheld by an agent who either acts or at least tries to do so.

This solution may appear uncomfortable because of the disjunction in the consequent. Indeed, someone might attempt to rescue the supervenience thesis by deleting the first disjunct, claiming that executive consistency is actually at core a matter of consistency between intending and trying. This might seem to be an attractive way to go for those philosophers who claim that every action involves trying, where trying is the decisive psychological component of action (cf. McCann 1975, 97; O'Shaughnessy 1980II, 46, 88ff.; Hornsby 1980, 33ff.). Were this claim to be correct, we could indeed dispense with the reference to action in the consequent.

There are, however, two reasons why this is no solution. First, it is anything but clear that, even if trying were to be a distinct and unitary phenomenon, we should think of it as purely mental.[29] So even if there were to be no actions without tryings, so that the first disjunct of the consequent could be dropped, the reference to trying with which we'd be left may itself refer to more than the mental. As it seems unlikely that any physical component of trying will be entirely determined by its mental component, an exclusive reference to trying is unlikely to be compatible with the supervenience thesis.

Second, and more importantly, we don't have good reasons to believe that every action is or involves a trying. As Severin Schroeder has argued convincingly, to say that X is trying or tried to φ is to bring up the possibility or certainty of failure from the perspective of either X or the speaker (Schroeder 2001, 218ff.; cf. Sect. 3.2.1). In other words, to talk of trying is to import a perspective-relative doxastic component into the description of an action or possible action. So, actions cannot always truthfully be described as tryings because speakers cannot always coherently indicate that they take the outcome to be either a failure or undecided. Moreover and decisively, the question of whether actions are necessarily tryings is misconceived, as there is no such thing as an intrinsic feature of actions that we call "trying".

Garrett Cullity has proposed a rational requirement that would cover executive consistency and yet be compatible with the supervenience thesis. According to Cullity, the consequent of a comparable principle should be given a doxastic twist, specifying that the agent *believes* she is φ-ing or trying to φ (Cullity 2008, 74).[30] There is, however, little reason to think that we always, or even generally, have beliefs that we are performing the actions we are in the process of performing. There is, perhaps, some plausibility to the claim that acting has a specific phenomenology, a phenomenology missing in pathological behaviour, as in cases of anarchic hand syndrome – although it is controversial what that phenomenology precisely consists in and whether it is unified. However, it does seem fairly clear that, if there is some kind of cognitive relationship agents generally entertain to their own actions, it is likely to be closer to perception or feeling than to belief (cf. Bayne and Levy 2006). But it is difficult to see how any claim of a general nature might be made to stick

[29] O'Shaughnessy, who makes the most substantial use of the concept, believes that trying is both mental and physical (O'Shaughnessy 1997, 73f.). However, as he also believes that what is picked out by mental and physical predicates is identical, his "dual aspect" theory would presumably be compatible with a non-disjunctive formulation of the consequent of *EC* that is unproblematic for the supervenience thesis.

[30] Cullity's *4-T* is actually a hybrid principle, combining features of a requirement of executive consistency with features of a requirement of subordinate intending.

here. There is certainly little plausibility to the idea that feelings of alienation in pathological cases entail that non-pathological actions are necessarily accompanied by some low-level feeling of doing. So, as we have no clear evidence that acting is necessarily accompanied by any kind of cognitive state or even proto-cognitive state, a putative standard of rationality demanding any such state cannot count as a requirement of executive consistency.

I have thus far argued for the claim that the first important intention-consequential standard is a requirement of executive consistency. I have argued against the counter-claim that any such principle can be no more than a conceptual requirement. In the process, I have rejected both a version of the Logical Connection Claim for intentions and the contention that actions cannot be relata in the wide-scope package over which such a requirement would have to range, as rationality necessarily supervenes on the mind. I have therefore rejected the supervenience thesis.

There is a further reason why we should reject the supervenience thesis. It is a reason that will turn out to be relevant for the formulation of all the *IC* requirements. Taking it into account here will give us what could be understood as a second principle of executive consistency. It is, however, more accurately understood as a first principle of intention pervasion.

7.2.2 Requirements of Intention Pervasion

Intentions tend to pervade the mental lives of their bearers. So too do certain desires and emotions. What is particularly striking about intention pervasion is that rationality requires it in certain forms. The form that has been most frequently discussed is often referred to as "the principle of instrumental rationality", a variant of which Kant labelled "the hypothetical imperative". It is, however, worth remembering that rationality also *recommends* certain kinds of mental behaviour to intenders, for instance, that they don't leave decisions about how best to implement their intentions until the very last minute, where such dilatoriness risks vitiating, if not preventing the performance of the intended action. Herein often lies the irrationality of procrastinators and self-handicappers, who may perform the actions they intend, but do so in ways that are seriously sub-optimal from their own perspectives. As the rationality of risk-taking depends on further wants* and beliefs of the individuals in question, the standards at work here are individual and holistic. Moreover, they also appear to be gradual, as certain risk levels are, by the overall lights of the risk-taker, less rational than others, without necessarily counting as irrational in the sense of being proscribed by rationality.

There has thus far been relatively little discussion of the relationship between such holistic and gradual standards of rationality and the categorical requirements that rationality imposes locally. Practical rationality's recommendations, which provide gradual scales for the local evaluation of behaviour, depend both on agents' intentions and on their other optative attitudes. Practical rationality's requirements, which provide stringent schemata for the local evaluation of behaviour, depend on only one sort of optative attitude alongside belief, namely on intentions.[31] I turn now to the first of three rational requirements that intending pervade, that is, structure the mental life of intenders in specific ways.

Agent-Relative Perceptual Obviousness

If the first of these requirements exists, there is a second reason to reject the supervenience thesis. We can get at its rationale by varying the story of Ewen's executive difficulties. In this version, Ewen II has resolved to discuss the issue with his boss before the latter goes on holiday. Ewen II is sitting at his desk when his boss walks across his field of vision towards the office door carrying an oversize suitcase decorated with a large "Gran Canaria" sticker and whistling "We're all going on a summer holiday". His boss is clearly leaving and this is obviously the last chance for Ewen II to speak to her about the topic, but the fact doesn't register with him. He simply doesn't take in the obvious signs. Indeed, he suddenly develops a strong interest in certain baroque-like features of the coffee stains on the desk, and the details of his environment relevant for the realisation of his resolution make no impression on his perceptual apparatus. As in the first version of his story, he will later find himself self-reproachfully relinquishing his intention after the (non-)event.

There is surely something wrong with Ewen II's rational capacities. If this is true, then, alongside the referent of the consequent, there can be a second feature relevant to executive rationality which doesn't supervene on the agent's mind. Unlike Ewen, Ewen II isn't the bearer of a belief that the time to act is upon him. Instead, he is perceptually situated in a way that presents him with the corresponding state of affairs as obviously true. What is obviously true – from where he is sitting – is that he won't perform the action in question if he doesn't perform it now.

Importantly, it is not just this non-attitudinal fact that is decisive. It is the fact plus the way the agent is perceptually situated relative to that fact. Note, moreover, that it is no objection here to point out that facts such as the one just named are reasons and shouldn't be mistaken for constituents of rational requirements. First, it isn't merely the fact that is in play. Second, in spite of the importance of distinguishing rational requirements and reasons, there is little plausibility to the claim that they are entirely unrelated. So pointing out an overlap is not a counter-argument.

[31] The stringency of individual *IC* requirements might be difficult to see in those cases in which satisfying local intention-consequent standards of rationality turns out not to be rational overall, by the lights of the agent in question. Someone satisfying *EC* may have violated some other requirement in forming the intention thus realised. He might have formed the intention in spite of believing he has conclusive reasons not to do (cf. Sect. 8.5.2). Then he will not count as rational overall. Conversely, an agent may find he has insufficient courage to go through with an intention formed in the face of the belief in conclusive reasons not to form it (cf. Mele 1995a, 60ff.). Not acting in the relevant way may be rational for him in the light of the totality of his attitudes. That doesn't change the fact that he is still locally irrational on two counts.

'Obviousness' is a relation of immediate epistemic accessibility. I suggest we understand it in terms of default causation: that is, in terms of the beliefs that would be acquired by an agent functioning normally in the situation. In Ewan II's case, the relevant beliefs are perceptual beliefs that would normally be caused in an agent with Ewan's intention and in Ewan's perceptual situation. His irrationality, we could say, grounds in the blockage of normal epistemic service.

Ewan II's failure to talk to his boss when it is glaringly obvious that he is in a now-or-never situation is a failure to do what he intends to do and looks very close to Ewan's failure. In Ewan's case, it seems that some kind of psychological obstacle must be responsible for him not doing what he believes he is required to do when he believes he is required to do it. In the case of Ewan II, it seems plausible that an obstacle of pretty much the same kind interferes with the mechanisms of formation of his perceptual beliefs. In both cases he is shying away from what he is rationally required to do.[32]

If this is correct, we need a second principle that complements *EC*, but characterizes the agent as irrational even in cases in which he doesn't doxastically process the information immediately accessible to him that he is in a now-or-never situation. The principle gains its specific form through the insertion of a clause outside the requirement's scope, a clause that picks out the now-or-never opportunity referred to in the content of a doxastic "taking" in *EC*, now framing that opportunity as agent-relatively obvious. The principle, then, looks like this:

(EC2)
If relative to *X* the time is obviously *t*,
it is rationally required of *X* that
if *X* intends to φ,
and *X* believes that she will never φ if she doesn't φ at *t*,
X φs or tries to φ.

We might think of *EC2* as generated from *EC* through the application of a 'perceptual obviousness operator' to the doxastic attitude responsible for monitoring the development of the conditions conducive to and necessary for *X*'s φ-ing. As it now names a condition of the requirement's applicability, rather than a feature of the agent the alteration of which could lead to the requirement's fulfilment, we are bequeathed a requirement whose scope is no longer wide – at least not maximally wide.

The claim that we should recognize such a modified principle of executive consistency grounds in what I take to be a natural everyday understanding of practical irrationality. If the principle does articulate such an everyday understanding and this understanding doesn't generate incoherencies when added to other features of rationality, it is further evidence for the falsity of the supervenience thesis – as well as a counter-example to any claim that all rational requirements are of wide scope.

[32] Compare the suggested explanation of the hedonically induced withdrawal of attitude contents from consciousness in Section 5.1.6.

The supervenience thesis certainly has the advantage of cordoning off the sphere of rationality neatly. Perhaps a certain amount of stipulation is necessary here. However, I have claimed that there are two significant ways in which the everyday conception of practical rationality goes beyond what the supervenience thesis can countenance.[33] This indicates that the province of rationality may not have borders that can be circumscribed by simple criteria.

Although the label "EC2" picks out the principle's character as a modification of the principle of executive consistency, I have introduced it as a first principle of intention pervasion. Intending to perform some action commits an agent to performing the action under certain circumstances. It also commits her to playing host to certain doxastic movements of mind that may be requisite for the action's performance.

Someone who rejects the supervenience thesis as a result of cases like Ewan II's might think that what is required is not so much a second principle of executive consistency, but rather a theoretical principle that requires the formation of beliefs under circumstances in which their content is obviously true and it matters to the agent whether it is true or not (cf. Cullity 2008, 79). Perhaps rationality recommends some such patterns of belief formation. However, as mattering is a gradable feature, its strength will play a role in determining how strongly the formation of relevant beliefs is recommended. But even strong mattering would not generate a stringent requirement. What is decisive is that the agent intends something, not that something matters to him. It is only under these circumstances that there is a stringent or categorical demand on the agent to process obvious evidence for a now-or-never situation.[34]

Subordinate Intending

What is often called "the instrumental principle" – roughly, the requirement to intend actions one takes to be necessary for one's ends – is the intention-consequential requirement that has been most widely discussed. Its status as a requirement of rationality has been much less frequently challenged than that of the requirement of executive consistency (cf. von Wright 1972, 45). As I have indicated, I take it that the two principles are structurally analogous. This is sensibly reflected in a formulation of the former principle that is modelled on that of the latter. I will be claiming that the obviousness operator is equally applicable here to the

[33] I will mention a third, more local way in Section 7.2.3 and a fourth, more global one in Section 9.4.2.

[34] A further reason why the principle should remain strictly practical and, in its consequent, specify action, rather than belief is that it is an open question whether the registering of the now-or-never situation need be in terms of beliefs or whether some sub-doxastic mechanisms may not suffice. The discussions of "implementation intentions" in social psychology tend to be formulated in terms of the triggering of actions by environmental cues (cf. Sect. 9.5.2). While such complete avoidance of mentalistic vocabulary seems implausible, the "automaticity" at work in such cases may get by not just without conscious belief, but without belief of any kind.

content of the agent's now-or-never monitoring. Moreover, here again, as in all the *IC* requirements, the personal pronouns that appear within the scope of the agent's belief pick out their bearer without predicative mediation and the modal character of the belief's content is again agent-relative: just as the agent subject to *EC* merely needs to believe she won't ever perform the action in question if she doesn't perform it now, the agent subject to the requirement on subordinate intending only needs to believe that she won't perform her intended action unless she performs some other action. Neither agent need have any beliefs concerning the question of whether a further action is, in Kant's words, "indispensably necessary" (GMS 417).

This last point is one of several that would need modifying if one were to take Kant's canonical formulation of the hypothetical imperative as the model here.[35] A second would concern the relationship between intending and what Kant calls "willing".[36] Whereas this need not concern us here, a third point is important if we are trying to be precise about the requirement. Kant's principle specifies an analytic connection in rational beings between "willing the end" and "willing the means". We should note that the focus on means, like the specification that they be taken to be[37] strictly necessary, involves an unnecessary restriction of the principle's applicability. As I argued in Section 3.2.1, means are best understood as antecedent causal conditions deemed at least conducive to the satisfaction of some want*. As such, they are just one way of instantiating the in-order-to relation. I suggested labelling attitudes taken on in order to facilitate the satisfaction of the content of a farther-reaching optative attitude "subordinate". Intending the means to some end is one sort of subordinate intention. An example of a different sort of subordinate intention would be the intention to go to St. James' Park, held by someone who intends to have been at least once to each of the Premier League football stadia. Going to the ground of Newcastle United is not a means, but part of the only *way* of realising this rather special intention.

As we should be attempting to formulate a principle with the maximum extension compatible with the maintenance of the strict character of a requirement, I suggest we think of the principle as one of subordinate intending covering both means and ways. These terms function equally to avoid a problem that arises if the principle is only formulated in terms of what the agent has to bring about in order to realise her intention. Side effects – or what we best label "collateral consequences", as these also need not be causally produced – may also be unavoidably brought about in the course of an agent realising her intention. But an agent who recognises the

[35] Cf. Korsgaard 1997, 234ff; Wallace 2001, 108ff.

[36] Both Audi and Searle suggest that the concept of "willing" in Kant's principle should be seen as narrower than intending. According to Audi (Audi 1991, 374), the former should be understood as an occurrence involving a level of exertion, i.e. one feature often subsumed under "trying hard". Searle thinks willing comprises, in his terminology, both a prior intention and an intention-in-action (Searle 2001, 264f.). In his early writings on these matters, Broome simply took willing and intending as equivalent (Broome 1999, 418; 2001, 103; 2002, 97).

[37] Actually, Kant's hypothetical imperative describes the means not as being taken to be, but as actually being necessary.

unavoidability of such by-products isn't therefore under a rational requirement to intend to bring them about.[38]

One might think that a precise description of the contents of the relevant doxastic attitude should be able to do without the terms "ways" and "means". There appears to be a simple way to do this, namely to work with two levels of belief: we specify that the agent believe she will not perform the superordinate action if she doesn't bring about some state of affairs, which she in turn believes she will not do unless she intends to do so. As collateral consequences obviously don't require an intention in order to be brought about, picking out those unavoidable bringings-about which themselves need to be intended successfully distinguishes ways and means from collateral consequences.

John Broome has proposed a formulation along these lines (Broome 2009, 64, 2010, 289).[39] Moreover, he adduces an excellent reason for doing so which goes beyond the desire for a clearer or more reductive analysis. His point is that we often don't need subordinate intentions in order to perform the relevant subordinate actions. A great deal of action subordination is automatically controlled. This is particularly obvious for a lot of conventional actions: an experienced cyclist doesn't need a separate intention to stick his right arm out if he intends to signal a right turn. Perhaps this is a reason why the traditional focus has been on means. But the point can be just as true of instrumental actions. Someone with a morning tea-making routine probably doesn't need a separate intention to pour boiling water into the tea pot. In these cases, the requirement we are after should not convict the cyclist and the tea-maker of irrationality.

Broome's requirement is, I think, one that indeed applies to rational agents. However, building into the antecedent a belief on the part of the agent that concerns the functioning of an attitude of hers, that is, a belief that she won't perform the subordinate action without a corresponding intention, imposes a massive restriction on the cases it covers. I am pretty sure that the second-order belief in question is one that deliberating agents very rarely form at all.[40] Remember that we have adjusted the conception of necessity and the relevant forms of instantiation of the in-order-to relation in order to expand the purview of the requirement. We should similarly be interested in avoiding a massive restriction of the requirement's scope through the addition of such a clause. We don't need to believe we need to intend a subordinate action for it to be rational for us to intend that action in order for us to realise

[38] He is, however, subject to a rational requirement to *accept* that he will also be bringing about the relevant collateral consequences, a point easily obscured by talk of "merely foreseeing". Unlike foreseeing, the acceptance rationally required is also optative. On this point and the distinction between acceptance and subordinate intending, see Roughley 2007c, 95f.

[39] The two beliefs of Broome's agent are "that, if m were not so, because of that e would not be so" and "that, if she herself were not then to intend m, because of that m would not be so". For a similar formulation see Bratman 2009a, 29; 2009c, 413.

[40] Cullity also makes this point (Cullity 2008, 70). I am using "second-order" here to designate an attitude towards an attitude (which itself does not have a further attitude as its content), irrespective of whether the attitudinal relation is mode-congruent (as in beliefs about beliefs) or mode-incongruent, as in beliefs about intentions.

the content of a further intention. More to the point, where you believe that you won't perform an intended action φ unless you perform some subordinate action ψ, it is certainly not true that the only circumstances under which it is irrational not to intend to ψ are those under which you believe you wouldn't ψ if you didn't intend to.

The solution is, I think, fairly simple. We need to specify that the agent is not disposed to automatically perform the relevant subordinate action in the situation in question. In other words, we need an extensional, situation-specific restriction on the agents to whom the principle applies. Once this is in place, we need merely stipulate the content of the agent's belief in terms of the agent-relative necessity of an action she takes to be a way or a means. It has the form:

(SI)
Where *X* is not the bearer of a disposition to automatically φ in situation *s* by ψ-ing,
it is rationally required of *X* that
if she intends to φ in *s,*
and believes that her ψ-ing at *t* is a way or means of her φ-ing in *s* such that she won't φ in *s* if she doesn't ψ at *t*,
she at a time she takes to be *t* intends to ψ.

As with executive consistency, the application of the perceptual obviousness operator to the contents of *X*'s taking of the time to be now or never generates a companion principle with an agent-relative obviousness clause outside the requirement's scope:

(SI2)
Where *X* is not the bearer of a disposition to automatically φ in situation *s* by ψ-ing,
and relative to *X* the time is obviously *t*,
it is rationally required of *X* that
if she intends to φ in *s,*
and believes that her ψ-ing at *t* is a way or means of her φ-ing in *s* such that she won't φ in *s* if she doesn't ψ at *t*,
she intends to ψ.

A final point on subordinate intending concerns a second way in which precisely formulated requirements are only the tip of what appears to be an iceberg of intention-related rational standards. Alongside the fact that there are modally lax, individual, holistic recommendations of rationality, it is worth noting that there are also further modally strict requirements which nevertheless remain vague because the conditions under which they are triggered are not precisely specifiable. An example is a requirement that has broadly the structure of *SI*, but in which the doxastic premise is vague or contains significant leeway. As Bratman points out, many of our intentions are such that they need "filling in" at some point (Bratman 1987, 31), that is, they will only be carried out if several, perhaps a whole series of subordinate intentions are formed. In these cases, agents are rationally required to form subordinate intentions to perform one of any number of candidate actions that may be individually no more than sufficient or conducive. They are required to do so no later than a point in time at which they believe they need to carry out *some* contributory action.

Eschewing Intention-Undermining Intentions

Since *Intentions, Plans and Practical Reason*, Bratman has repeatedly emphasized the fact that, alongside a constraint of "means-ends coherence", intentions are also subject to what he calls "consistency constraints" (Bratman 1987, 31ff.; 2009a, 29, 49ff.; 2009c, 413). The rational requirement on intentions that is most naturally described as a consistency constraint is, I think, best formulated as follows:

> (IIC)
> It is rationally required of *X* that
> If she intends to φ at *t*,
> she doesn't intend not to φ at *t*.

Note that *IIC* ('intention-intention consistency') strictly parallels the requirement of theoretical rationality that if an agent believes that *p*, she not believe that ¬*p*. *IIC* doesn't require a mediating belief of the agent that the two intentions are not co-realisable for her – just as the requirement not to believe *p* and believe that ¬*p* is not dependent on the believer having a third belief that the first two beliefs are incompatible. On the contrary, an agent who intended to φ at *t*, intended not to φ at *t* and also believed that φ-ing at *t* and not φ-ing at *t* were compatible would be doubly irrational. On top of contravening *IIC*, he would also be believing something he is required not to believe.

Interestingly, Bratman doesn't mention *IIC* in his discussion of consistency. Instead, he focuses on a requirement that certainly covers a great deal more cases, a consistency requirement that is doxastically mediated, as the incompatibilities it proscribes are contingent on the way the world is taken to be. Bratman originally spoke here of "strong consistency relative to [the agent's] beliefs" (Bratman 1987, 31). The requirement in question is a close cousin of *SI*. Whereas *SI* deals with a (believed) relation of (agent-relative) necessity of a second action for the performance of a first, its cousin deals with a (believed) relation of (agent-relative) sufficiency of a second action for the non-performance of a first. Correspondingly, where the package required by *SI* contains a second intention, the package required here contains the eschewal of a second intention. One way to put it is this:

> (IBC)
> It is rationally required of *X* that
> if she intends to φ,
> and she believes that she won't φ if she ψ-s,
> she doesn't intend to ψ.

The requirement is considerably less complicated than *SI*. For one thing, there is no need for the exclusion of automatic processes: *IBC* is neutral between deliberate and automatic refraining. For another, there is no need for differentiation between grounds of (agent-relative) sufficiency as we needed to differentiate between causally and conventionally constituted grounds of necessity in *SI*. Like *IIC*, *IBC* is a requirement that aims at eschewing intention-undermining intentions. However, where intention consistency in the strict sense guarantees that the intentions' bearer avoids undermining her own intentions in the crassest manner, intention-belief consistency merely guarantees the appearance that her intentions don't undermine each other. Whether the appearance corresponds to reality will depend on the truth of the agent's beliefs.[41]

Bratman has argued that *IBC* is analysable as consisting of two norms, one exerting "rational pressure" in the direction of intention agglomeration and one proscribing inconsistent intention contents (Bratman 2009a, 50f.; 2009c, 412, note 5). This is, I think, a mistake. Compare, first, *IIC*. Here, we clearly don't need a separate norm proscribing the incoherence that arises when someone forms a single intention whose content conjoins two contradictory contents. The proscription of $p\&\neg p$, i.e. the law of non-contradiction, is so basic that there is a sense in which we don't need it, as we can't make any sense of what it is that it forbids. It formulates a limit on the contents of any kind of attitude, including wants* (Sects. 4.1.2 and 4.2.2).

IBC forbids combinations of intentions such as "to take the bus home at 6 o'clock today" and "to watch the news when I get home" in the light of the belief that, because the news starts at 5.45, taking the bus at 6 will make it impossible to watch the news. But there appears to be no particular reason why an agglomerated intention that mentions both bus-taking and news-watching should, together with the belief just mentioned, more substantially contribute to, or detract from the agent's rationality than the two individual intentions out of which it is constructed. The inconsistency only becomes clear when beliefs concerning the contents of *each* of the intentions conjoined in the bigger intention are brought to bear. Connectedly, the inconsistency at issue is not internal to the product of agglomeration, but requires the relation of the intention to the relevant belief. There appears to be no reason why the inconsistency between a belief and one conjunctive intention should be thought to be in some way more substantial than that between the belief and two simple intentions. The individual intentions will do the job just as well.

Bratman first claimed that there is rational pressure to agglomerate intentions in his discussion of the video games example, in which an agent attempts to hit target 1 and attempts to hit target 2 in the knowledge that success in the one excludes success in the other. Bratman argues that the rational pressure to agglomerate intentions rationally excludes the agent's guiding attitudes from being intentions, rather than what he calls "endeavourings". Were the agent to be the bearer of two intentions with those contents, he would rationally end up intending to hit both. But that would

[41] *IBC* is equivalent to the consistency principle Bratman formulates at 2009c, 413. My formulation maintains the structure of the formulations of the other requirements in this chapter. It is also explicit about the agent-relative character of the modality.

be problematic (Bratman 1987, 134). What the problem would precisely be is left somewhat unclear in the discussion of the video games example. If the problem is the non-satisfaction of the consistency constraints, it is also unclear why the conjunctive intention should be thought to fare any worse than the conjunction of intentions taken by the agent to be equally unrealisable.[42]

I think that agglomeration is much less important than Bratman seems to assume. It should, I suggest, be seen simply as an epistemic tool that can help to make clear what one is committed to by one's attitudes (cf. Sect. 4.1.2). Where two attitudes of the same type are agglomerated, producing one attitude with a conjunctive object, this can be conveniently represented in English by the determiner "both", a device that helps to make the problem more vivid. There is no comparable device where the number of attitudes put together exceeds two.

A further reason why Bratman's two-step analysis of the consistency constraint doesn't work is that "rational pressure" (2009a, 52) is in a specific sense too weak to do the job Bratman assigns it. In order to count as providing an analysis of the requirement, both steps would need to have the same modal status as the requirement itself. But *IBC* like *IIC* is a stringent requirement, whereas pressure, once again, comes in varying strengths. It is surely plausible that there is no such strict requirement on intention agglomeration. This certainly follows if agglomeration is no more than a useful epistemic tool. The "pressure" then simply consists in the advisability of agglomerating certain intentions in order to get clear on what playing host to them simultaneously amounts to. If *IBC* is a genuine a priori requirement, such pressure cannot be its foundation.

7.2.3 The Requirement of Deliberative Intention Persistence

The claim that requirements of practical rationality take wide scope has seemed to several authors to be flawed for a reason that has nothing to do with the supervenience claim. The objection grounds in the intuition that practical reason has a natural direction, namely from premise states to conclusion states (Schroeder 2004a, 346; 2009, 227; Kolodny 2005, 527ff.). According to the wide scope reading of the intention-consequential requirements, they can be equally satisfied by abandoning the primary intention, abandoning the relevant belief or by forming (or not forming) the intention or performing the action specified in the consequent. Although it is plausible that abandoning one of the premise states is rational in some cases, it seems equally plausible that there is at least a presumption in favour of

[42] In this early discussion, Bratman actually considers first the consistency constraints and then the pressure to agglomerativity, and doesn't use the former to explain the problem resulting from the latter. Rather, he claims that consideration of the second speaks to "a related point", which makes it sound as though the difficulty raised by agglomeration can be clarified independently of the application of the consistency constraints.

satisfying the consequent. This view seems particularly plausible once we focus on the fact that agents with dominant tendencies to abandon either one of the premise states specified by *SI* look to be suffering from characteristic forms of practical irrationality.[43]

Belinda, Willoughby and Subhi

Take Belinda. Whenever an action she believes she needs to perform in the service of some intention threatens to be strenuous, she simply abandons her belief. Belinda is a wishful thinker, a rationaliser, probably a self-deceiver: what she is willing to do functions for her as a reason for belief. That is fairly serious irrationality.[44] Now take Willoughby. Whenever an action he believes he needs to perform in the service of some intention threatens to be strenuous, he simply abandons the intention. Willoughby plausibly suffers from a variant of procrastination or of weakness of will: he is continually changing his choice of career, of leisure activity and of relationships and, as a result, never gets anywhere with any of them.[45] Now, Belinda and Willoughby are certainly unusual characters. But compare Subhi, who inevitably forms intentions to perform actions without which he thinks he won't be able realise the aims he brings to deliberation. Like Belinda and Willoughby, he satisfies *SI*. Unlike them, he seems, from the limited description we have of him, to be perfectly in order rationality-wise.

There is, then, some sort of asymmetry at work here. Nevertheless, as has been variously argued (Brunero 2010; 2012; Way 2011), it would be over-hasty to conclude that this demonstrates the falsity of the wide scope view. The first point to notice is that there is an important difference between the first and the second type of premise state, a difference concerning their rational susceptibility to voluntary revision. Someone who alters her beliefs simply because they don't fit her wants* is already behaving irrationally, whereas revising your intentions so that, as you believe, they better fit other of your wants* is in itself a perfectly rational thing to do. This may suggest that it is above all the doxastic state that should be rationally held stable. The suggestion would, however, be wrong. The wide scope interpretation says nothing about the *reasons* why an agent might abandon either of the premise states. In particular, it doesn't entail that the reason why the agent abandons either primary intention or belief be a lack of willingness to form the

[43] I will suggest at the end of this section that some cases are more accurately classified as examples of practical unreasonableness.

[44] Setiya and Schroeder have argued that wide-scope interpretations of rational requirements risk endorsing wishful thinking or rationalisation (Setiya 2007, 667; Schroeder 2009, 227).

[45] Compare Thomas E. Hill's important and insufficiently discussed Amy (Hill 1986, 120ff.). More local variants of Willoughby are my example of Slim (Roughley 2008a, 145) and Brunero's Candice (Brunero 2012, 136f.; Bratman 2012, 83).

subordinate intention.[46] Clearly, if an agent finds himself with a primary intention and belief that fit the bill of *SI*, it is likely to be unwillingness to form the subordinate intention that is going to interfere with a direct move to complete the prescribed package of attitudes. However, that unwillingness doesn't therefore need to be the reason for the revision of either of the premise attitudes. It may simply be the feature that persuades the agent to think through the grounds for his belief in order to check whether he really would need to overcome that optative resistance in order to realise his primary intention. The answer might turn out negative for reasons relating to features of the grounds for the belief that the agent had thus far not felt it worthwhile looking at in detail.

Two points of importance emerge from this first clarification. First, it can be rational to retract the doxastic premise state. If this is the case, that is likely to be because the move is covered by some further requirement of theoretical rationality relating belief and beliefs in grounds for that belief, a requirement that is independent of *SI*. Conversely, where doxastic revision is irrational, that is because the move contravenes such a requirement. In this way, further principles of rationality narrow down the options an agent has of being rational overall. However, the fact that someone might fail to be rational overall by failing to meet a requirement on belief doesn't change the fact that she may be perfectly in order as far as *SI* is concerned. Second, the conditions under which the belief has been incorrectly formed by the agent's own lights are going to be much less frequently satisfied than the conditions under which the belief seems to her to hold up in the light of her grounds. This, as John Brunero has convincingly argued (Brunero 2012, 130), would explain why abandonment of the belief doesn't appear to be rationally on a par with adoption of the subordinate intention. All in all and for empirical reasons concerning the way we tend to form our beliefs, it isn't, but locally – as regards *SI* – it is.

Belinda, then, is seriously irrational from a theoretical point of view, even as she satisfies the practical requirement *SI*. But what about Willoughby? An explanation parallel to the one we have given for Belinda would have to advert to a further principle, or further principles of rationality that would clarify why we can also see him as being locally rational yet irrational all-things-considered. The task looks to be somewhat less tractable for a number of reasons.

[46] Kolodny's "reasoning test" incorrectly presupposes this. According to the proposed test, one is subject to a wide scope requirement "only if, from a state in which one has conflicting attitudes *A* and *B*, (i) one can reason from the content of *A* to dropping *B* and (ii) one can reason from the content of *B* to dropping *A*" (Kolodny 2005, 520f.). But between the case of reasoning from the contents of the attitudes covered by *SI* and the case of unreasoned attitude change – for instance, as a result of an electric shock – lies the case of reasoning from other beliefs acquired or reassessed in view of the costs of forming the subordinate intention.

Is Intention Persistence Ever Rationally Required?

The most obvious reason is that it is unclear whether there really is such a principle and, if so, what form it takes. This contrasts with the question of belief's stability, which is uncontroversially regulated by a requirement relating beliefs and beliefs in their grounds. A principle of intention persistence would also contrast in the same way with requirements such as *SI*, *IIC* and *IBC*, which – independently of their precise wording – set up an explanatory challenge. That is, some such requirement seems intuitively in place and calls out for an explanation. The existence of a requirement of intention persistence may, in contrast, have to be argued for in tandem with arguments for its justification.

This need not count as an argument against there being such a principle. T.M. Scanlon has plausibly claimed that acting according to requirements of rationality doesn't require beliefs that there is a rational requirement to act in the relevant ways (Scanlon 2007, 85f.). If this is correct, we may be subject to requirements in spite of having no strong intuitive access to their contents. This might be the case with a requirement of intention persistence. Alternatively, there might be forms of pressure to maintain our intentions that ground in characteristic constellations of reasons. Such pressure would not amount to a requirement, but might still be datum that needs explaining by a theory of intention. In what follows, I shall be arguing that the normative dimension of intention persistence is a somewhat messy affair, as intention retention is subject both to a strict rational requirement and to standards of reasonableness that specify distinctive pro tanto reasons.

Let us begin with a suggestion of Richard Holton's. Holton has claimed that one of the striking parallels between intention and belief lies precisely in their characteristic persistence. Intentions and beliefs, he thinks, have a lower threshold for formation than for revision – both psychologically and rationally (Holton 2009, 30f.). For both kinds of attitudes, we both naturally and rationally require more substantial relevant input in order to abandon some token than we do in order to acquire it in the first place. The suggestion seems to be that we have a natural disposition here which it makes sense to have all in all. This is, however, not to say that there is a corresponding requirement on token intentions.

There is, moreover, a worry about this suggestion. If it were meant to pick out a feature that intentions and beliefs share and that distinguishes them from other attitudes (Holton doesn't say this), it would, I think, be false. Certainly, many of our desires and emotions are not easily jettisoned once we have acquired them. Some whims or spontaneous affective reactions may dissolve very quickly, independently of further inputs. But some low-level intentions – to make some remark in the course of a conversation, to try one of those canapés before one leaves – may quickly dissolve through forgetting or getting lost in the whirr of things (Sect. 7.1.2). Where our desires or emotions are deeply entrenched, there are going to be reasons of psychic economy for sticking with them that will generally carry the day unless there are other prudential or moral reasons to be rid of them. Reasons of the same kind will be in play in relation to intentions that have become broadly entrenched or pervasive in their influence on the agent's psychology. This is certainly not enough to help us explain Willoughby's problem. Indeed, part of the problem may be precisely that his intentions never are sufficiently entrenched. Moreover, there would be little plausibility to the claim that, although the dissolution of low-level intentions is psychologically not unusual, it is nevertheless irrational. Willoughby clearly has a serious problem in the rationality department, and one that requires an explanation, whereas someone who forgets spontaneous and inconsequential intentions is hardly worthy of rational criticism.

Deliberation and Rational Intention Persistence

I wish to suggest that the focus on deliberatively generated intentions will allow us both to restrict the application of a requirement of intention persistence to cases in which some such principle is plausibly in play and, ultimately, to explain its existence. In what follows, I will be concerned with arguing for a specific formulation of the requirement, a formulation that grounds in the assumption of deliberation's significance.

The suggestion, then, is that it is those intentions that result from an active process of formation that are candidates for rational persistence. Recall that the empirical evidence for doxastic mechanisms that shield intentions against destabilising input (Sects. 6.4.1 and 7.1.3) is drawn exclusively from decisional cases. This plausibly ties in with the fact that such processes of active formation distinguish the resulting intentions from entrenched emotions and desires. In the next chapter, I will come back in some detail to the sense in which deliberative intention formation is active (Sects. 8.3, 8.4 and 8.5). There is a linguistic feature of the debates on intention persistence that points to an inexplicit recognition of the significance of prior deliberation, namely the use of the term "*re*consideration" when discussing reflexion on whether to uphold an intention (Bratman 1987, 16f, 60ff; 1995, 53ff.; Holton 2009, 3, 71ff., 121ff.). Obviously, you can only reconsider something you have already considered in the first place. If we equate "consideration" with "deliberation", this suggests that a principle of persistence may only apply to deliberatively formed intentions.[47] This is, I think, indeed the case. The intuition here is that there is something about the products of deliberation that makes their preservation particularly requirement-worthy. This would explain why there is a rational presumption that they should not be dropped.

If it is correct that the rational default for deliberatively formed intentions is their preservation, we are going to need clarity on two points. One is the doxastic condition under which the default mechanisms are rationally overridden. The other

[47] A number of the discussions of rational requirements don't focus on intentions in general, but on intentions with a decisional aetiology. For example, although Wallace, Cullity and Scanlon all formulate requirements in terms of intentions, their discussions focus on "choice" (Wallace 2001, 16) or "decisions" (Cullity 2008, 58; Scanlon 2007, 92ff.). Even Bratman emphasises the importance of decision in his discussion of rational requirements (Bratman 2009b, 230).

is the question of what is precisely meant by “dropping” an intention. This is considerably more important than may at first appear because of the fact that reconsideration, i.e. the redeployment of the procedure through which the intention was formed, is frequently, but not necessarily, involved in its dissolution.

Let us begin with the latter point. There are four ways in which an agent can cease to intend.[48] She can (i) perform the intended action; she can (ii) permanently forget her intention; she can (iii) put the intention on hold by re-entering deliberation as to whether to realise it and she can (iv) abandon it as a result of such deliberation.

Obviously, a requirement on persistence has to allow (i). Moreover, the principle should cover cases of sudden, permanent forgetting (ii). Imagine an agent who finds himself in a situation covered by either *EC* or *SI*, but who, at the very moment at which he is required to bring about coherence suffers a sudden bout of permanent intention forgetting. Although he thus satisfies the requirement, there is something clearly wrong with the way he is organised from the point of view of rationality.[49] Such an agent falls foul of a requirement of intention persistence.[50]

It might be objected that the requirement cannot cover such cases, as these will be cases in which the agent has no ability to do otherwise. And “required”, it may appear, implies “can”. However, it is unclear whether this last assumption holds. If to register irrationality is to say that there is something wrong with the way a person is, that does not entail that there is anything the agent could have done to prevent that state coming about. Irrationality may in some cases simply be a matter of being psychologically organised in a problematic way. Whether an irrational person might, through strength of will or therapy, be able to find a way of rectifying that seems to be a further question.[51]

Importantly, the requirement should concern (iii) as much as (iv), even though the suspension of an intention may be followed by its reinstatement. This is suggested by an empirical point and necessitated by a conceptual one. The empirical point is that, as Holton has argued for the specific cases of intentions he calls “resolutions”

[48] Compare the five explanations of want* non-persistence in Section 5.1.2. Note that in contrast to cases of want* dissolution, talk of intention non-persistence naturally makes the intention’s bearer the subject of the description. This is a symptom of the fact that we think of the possession of intentions as being in some sense minimally active. This is not the case with wants* in general, the dissolution of which can be explained by purely sub-personal mechanisms. Because of the relevance of such mechanisms, there are more ways to lose wants* than to lose intentions. However, those wants* that are not intentions do not dissolve merely as a result of their bearer’s deliberation on whether to realise them.

[49] Scanlon disagrees on the basis of the supervenience thesis for what he calls “structural rationality” (Scanlon 2007, 93). Scanlon’s modus ponens is my modus tollens here: the role that forgetting can play in irrationality provides a third reason for not restricting the concept of rationality to interattitudinal coherence.

[50] This seems not to be the case if the forgetting has taken place over a longer period (cf. Broome 2001, 112f).

[51] The different reactive attitudes that appear appropriate where someone is irrational, as opposed to where someone is immoral, may be indicative of a difference in the relation between “rationally required” and “can” on the one hand and “morally obligated” and “can” on the other.

(Holton 2009, 140ff.), it seems that the mechanics of non-revision ground primarily in non-reconsideration. Now, rational requirements don't necessarily follow psychology in individual cases. Still, it is plausible that the psychological constellations rational requirements pick out as unproblematic don't involve forms of mental behaviour that human agents characteristically find unnatural.[52]

It is, however, the conceptual point that is decisive: once an agent re-enters deliberation, she is confronting the reasons for and against the action in question, independently of the fact that she had up to a moment ago intended to perform it. If her having intended has led to her making changes to the world which have altered the balance of reasons relative to the time when she formed the original intention (Cullity 2008, 64ff.), these simply line up with the other reasons that are to be considered once deliberation has begun again. The suspended intention now only has the status of a content that could be intended. A rational requirement on intention persistence should thus proscribe reconsideration alongside other ways of relinquishing intentions, where the default mechanisms are cancelled.

Note an important consequence of this: A requirement that covers re-entry into deliberation cannot separately cover intention revision within deliberation. Once an agent has entered into deliberation, she has ceased to intend and so cannot be required to uphold an intention of which she is no longer the bearer.[53]

Sudden permanent forgetting, putting on deliberative hold and actively abandoning are, then, all cases that should be covered by a requirement of intention persistence. If, as I have been suggesting, the principle only applies to intentions acquired through deliberation, deliberation has an asymmetric role to play in such a requirement: the package it covers concerns intentions acquired, but not necessarily relinquished deliberatively.

The Doxastic Premise state

What, now, is the content of the doxastic premise state? The key here lies in the deliberative features of the intention's genesis. As deliberation primarily involves the weighing of reasons, it seems clear that the agent's beliefs about her reasons are going to be part of the package. However, the reason why a requirement should also cover reconsideration also in part determines the contents of the requirement's doxastic conditions: if the requirement were only to concern the upholding of an intention where the agent believes herself to have sufficient reasons to thus intend, persistence would not be the rational default, but at most the rational result

[52] Indeed, a normative functionalist such as Bratman (cf. Sect. 7.3) is committed to the claim that the dispositional and normative dimensions of intending run parallel.

[53] In contrast, the scope of the principle that has been proposed by Bratman includes the "abandonment" of intentions during deliberation (Bratman 2012, 87, note 24).

of renewed weighing. The fact that the agent had thus intended up to that point would be irrelevant and the requirement would merely concern rational intention formation.[54]

If, then, the threshold to renewed deliberation is the primary barrier to intention dissolution, the principle of intention persistence should cover not only an agent's beliefs about reasons that may be entertained in deliberation, but also her beliefs about reasons for re-entering deliberation. Indeed, it seems that the recourse to such beliefs may be the decisive way in which a principle of intention persistence goes beyond a general requirement on rational intending. The basic idea, then, is that an agent who deliberatively intends some action should not cease to intend it if she doesn't believe that the conditions for deliberation have improved or that the reasons on the basis of which she decided have deteriorated. I suggest, then, that the requirement on deliberative intention persistence looks like this:

(DIP)
It is rationally required of *X* that
if at t_2 *X* intends to φ at t_3,
where *X*'s intention to φ was deliberatively formed at t_1 and has persisted up to t_2,
and if at t_2 *X* doesn't believe that, relative to what she believed at t_1,
conditions for deliberation have improved
or that her balance of reasons for φ-ing has deteriorated,
X at t_2 does not cease to intend to φ.

A clarification and a remark: first, the "deterioration" of an agent's balance of reasons can involve one of a number of developments. It may be a matter of the devaluation of the intended goal, the improvement of reasons for an action an agent takes to be incompatible with the action intended or an increase in the costs taken to attach to performing the action. The specification that the deterioration need to have taken place "relative to what *X* believed at t_1" entails that the agent may come to see that her reasons have got worse or else simply become aware that she was wrong in her original judgement about how good her reasons were.

Second, *DIP* differs in a second, important way from the persistence principle proposed by Michael Bratman.[55] According to Bratman, the doxastic component concerns the presence of a supporting attitude, "confidently taking one's relevant grounds to support this very intention", whereas *DIP* specifies the absence of beliefs that would undermine the intention. Bratman's formulation covers far fewer cases and is thus perhaps in this respect easier to defend. However, it seems to me to be far more restrictive than we need to be. People frequently intend to do things over a period of time without having very many thoughts about the grounds for

[54] This point is emphasized by Luca Ferrero (2012, 148). On the requirement of rational intention formation, see Section 8.5.2.

[55] Here is Bratman's principle *D*: "The following is locally irrational: Intending at t_1 to *X* at t_2; throughout $t_1 - t_2$ confidently taking one's relevant grounds adequately to support this very intention; and yet at t_2 newly abandoning this intention to *X* at t_2" (Bratman 2012, 79).

thus intending. This seems to be at least partly because there is a thin line between thinking about such grounds and re-entering deliberation. The requirement we are after should cover cases devoid of such thoughts. Of course, how strong Bratman's doxastic requirement is depends on how exactly we read "taking" and "confidently". If the reading requires no conscious thoughts at all on the matter, then it is unclear how the condition is to be distinguished from a negative condition of the kind I am suggesting. Moreover, the qualifier "confidently" is clearly intended to go beyond mere belief and it would be fairly natural to understand the additional feature as affective and thus necessarily conscious. If we stick to negative conditions, we can avoid such unnecessary restrictions.[56]

Reasonable Intention Persistence

There is, then, a rational requirement on intention persistence, at least where intentions are formed as a result of deliberation. *DIP*, although it is irreducibly diachronic, is both of wide scope and fulfils Broome's supervenience stricture. It enables us to explain our sense that something is wrong with someone such as Willoughby, if he revises his considered career choices without believing that anything substantial has changed as far as either his deliberative situation or his reasons go. However, further consideration of Willoughby's case makes it clear that *DIP* only helps up to a certain point. The instability of Willoughby's intentions might correlate with an instability in his beliefs about his reasons. If that is the case, then *DIP* doesn't tell us what is wrong with him.[57] Rather, it seems that his problem can only be covered by external standards we bring to bear here that don't concern coherence between attitudes. Those external standards plausibly pick out specific kinds of pro tanto reasons, which broadly correspond to the kinds of reasons mentioned in *DIP*'s doxastic condition. First, there is a pro tanto reason for an agent to abandon an intention to φ where there has been a significant rise in the costs of his φ-ing or fall in the probability of his being able to φ relative to the moment of the intention's formation. Second, there is a pro tanto reason for an agent to suspend and reconsider his deliberatively formed intention where there has been a significant increase in relevant information or his capacity for clear thought since the intention's formation.

If Willoughby is not irrational in the light of his relevant beliefs, he will be unreasonable in revising his significant choices if there are no pro tanto reasons such as those mentioned which would justify such revision. The original description

[56] Like Bratman, I believe that the purview of the requirement is restricted by a condition that the agent think about her grounds. My claim, however, is that the consideration of grounds should predate the formation of the intention, not accompany its persistence. Bratman may feel pressure to strengthen the doxastic condition because of not specifying that the requirement only applies to deliberatively formed intentions.

[57] By the same token, Bratman's principle doesn't, contrary to what he suggests (Bratman 2012, 83f.), fully explain our reactions to Brunero's example of Candice.

I gave of Willoughby stated that he abandons career, leisure activity and relationship choices as soon as they seem to him to require slightly strenuous subordinate actions. It is, however, already obviously true at the time of Willoughby's decisions that these kinds of choices can only be realised by adopting courses of action with some strenuous components. So, ceteris paribus, the costs of sticking to his intentions haven't changed and thus don't deliver even a pro tanto reason for revision. If that is all there is to his reasons, then, he is behaving *unreasonably* in dropping the intention, independently of how he judges, that is, independently of how rational he is.

It could be pointed out that, in this case, Willoughby's problem lies not in the formation of his intention, but in his judgement. There is certainly a version of his story according to which he rationally forms his intention on the basis of an unreasonably formed, perhaps false reasons judgement. Two points are worth making on this. First, this version of the story cannot get by without going beyond problems of coherence within packages of – real or counterfactual – attitudes. It also needs to look to reasons that might well not be acknowledged by the agent. Second, as there is empirical evidence that sometimes our reasons judgements follow the formation of our intentions rather than the other way round (Holton 2009, 64ff.), we shouldn't make Willoughby's problem depend on his unreasonable judgements. There is a problem with his instantiating this pattern of intention revision, however he comes to do so.

We are now in a position to advance a diagnosis of the asymmetry between the primary intention and the subordinate intention in the scope of *SI*. Remember that, in the case of the doxastic premise state, the asymmetry results from the fact that our everyday beliefs are generally well aligned with our beliefs about the grounds for those beliefs: we tend to satisfy the epistemic requirement that makes our beliefs rational. In the case of the deliberatively formed intentions we bring to a situation covered by *SI*, things are somewhat more complicated. This is because rational intention persistence depends not only on the agent's beliefs about the reasons for her intentions, but also on the reasons she takes herself to have to re-enter deliberation. *DIP* is a principle that explicitly characterises persistence of deliberative intentions as the rational default. Absent beliefs undermining either the justificatory status of the agent's reasons for intending or the trust that agents normally have in their own deliberation, the retention of a deliberatively formed intention is rational. Moreover, in as far as the absence of such beliefs reflects the absence of the kind of reasons in question, retaining the intention will not only be rational, but also reasonable.

SI situations may trigger thoughts about how well founded one's primary intentions are, as they may concerning the justification of one's beliefs. Note an important difference, though: bringing together an intention with – perhaps newly acquired – beliefs about what one needs to do in order achieve one's goal may directly engage *DIP*, involving as it might revision of the agent's beliefs concerning the balance of reasons for her primary intention. This explains why the asymmetry between the primary and the subordinate intention is plausibly less marked than that between the doxastic premise state and the subordinate intention. Nevertheless, because our intentions are generally responsive to both our beliefs about reasons and to our beliefs about the reliability of deliberative conditions, *SI* cases will usually begin with intentions that are fairly well shored up by such beliefs. This is why there is still an asymmetry between the intentions at issue in *SI* situations as regards the rationality of abandoning one or forming the other.

The normativity of intention persistence, then, is rather a messy business. There is, I have argued, a rational requirement stipulating a form of coherence among an agent's attitudes that holds over time, not just at particular points in time. However, the relevant rational requirement doesn't cover certain clearly problematic kinds of case characterised by dilatoriness or a general inability to commit oneself to projects. Such cases are covered instead by what we might think of as principles of reasonableness, principles that specify when there is a pro tanto reason for intention abandonment or reconsideration. The important upshot for our purposes is that intentions differ from other wants* on *both* counts.

7.3 The Intentional Syndrome: Taking Stock

The most striking hallmarks of intending are its normative roles. Intention's causal environment differs somewhat from that of other optative attitudes. Those differences seem, however, to be quantitative rather than qualitative. This is plausible both for intending's salience effects and for the general tendency of intentions to be motivationally stronger, to be more pervasive and to be more persistent in the face of obstacles than many other kinds of optative attitude. The fact that we don't have a linguistic means of affectively qualifying intentions can plausibly be accounted for in terms of the interest that structures the concept's focus.

The strongest challenge to the reductionist lies in making sense of the role that intentions play in various standards of practical rationality. I have claimed that there are five basic intention-consequent rational requirements: a requirement of executive consistency, of subordinate intention formation, of intention-intention consistency, of intention-belief consistency and of deliberative intention persistence. They are supplemented by variants of the first two requirements generated by the application of the perceptual obviousness operator, which removes one conditional component from the requirement's scope. The components of the packages covered need not be simultaneously instantiated, as the requirement on persistence shows. I have also argued that they may not be only mental states. This is true if there are indeed rational requirements that include reference to perceptually obvious truths in their conditions. It is also true if there is a requirement of executive consistency. I have argued for both claims. Finally, because of the special status of deliberative intentions as attitudes that are actively acquired, deliberative intention persistence is also subject to reasons of a specific sort. In all these respects, intentions seem categorically different from other sorts of wants*.

I have been looking at these phenomena in detail because a theory of intention, particularly a theory that sees intentions as types of wants*, needs to offer an explanation for them. Before proceeding, I want to note an alternative strategy that has been prominent in the discussion of intention and say why I shall not be adopting it. According to this proposal, what intending is can only be understood in terms of the norms to which an intender is subject. Thus understood, intention is itself a normative concept: intentions are whatever it is that plays the role marked by the term “intention” in standards of the kinds we have been discussing. Conceptions of this ilk are variants of *normative functionalism*.

There are, broadly, two versions of this proposal. The first version is Michael Bratman’s and is a normative extension of conceptual functionalism. According to Bratman, intending, in particular being the bearer of intending’s commitment component, is constituted by two sets of “roles”, one set being descriptive, the other normative. It is a matter, on the one hand, of being the bearer of dispositions to act and to reason, and of having the higher-order disposition to retain these first two sorts of dispositions. On the other hand, it equally involves being subject to norms of the kind we have been discussing (Bratman 1987, 9ff., 107ff.). The second, more radical version of the proposal abandons conceptual functionalism’s criterial use of dispositions, claiming instead that all conceptual work is done by the norms to which bearers of attitudes – beliefs, desires, intentions, emotions – are subject. This proposal could be labelled “pure normative functionalism” (cf. Zangwill 1998, 190ff.; cf. Sect. 4.5.1, notes 39 and 43), to be distinguished from Bratman’s dispositional-normative version.

According to the pure normative functionalist, “rationality” is a property that is essential to the explanation of action. “Rationality” here means the disposition to respond appropriately to the “normative essences” of the mental states one finds oneself in. This move presents us with an uncomfortable dilemma with respect to an understanding of the explanatorily primary property of rationality. Either rationality is some sort of substantive property or else it is simply a term by means of which the diverse tendencies to react to our attitudes are grouped together. If the first is the case, explanations of action will be seriously incomplete, indeed mysterious, until we know what constitutes this special supplementary property. If, on the other hand, we take the second option, then talk of “rationality” is not doing any explanatory work of its own.

Further, it is unclear how such a conception, if generalised to apply to all the attitudes, can deal with the obvious fact that agents behave in all sorts of ways as a result of their various forms of attitudinising, forms of behaviour that are often not to be understood as particularly rational. Take people’s tendency to talk about the objects of their desires (Sect. 3.2.2) or to blush, stutter and look at the ground when they are embarrassed. The most important point for our discussion, however, is that normative functionalism’s all-encompassing conception ignores the fact that certain attitudes, in particular intention and belief, entertain a *special* relationship to a priori standards of rationality. If the explanation of this difference relative to other attitudes is one of the central challenges of a theory of intention, a conception that sees all attitudes as essentially normative is going to be unhelpful.

In contrast, Bratman develops his position as a specific analysis of intention. Moreover, the hybrid conception accommodates the everyday understanding that our attitudes are essentially features of our psychology. His characterisation of the relationship between the norms and the corresponding dispositions has been primarily in terms of the non-explanatory relation of “association” (Bratman 1987, 9, 109; 2007b, 5). However, it is surely plausible that we are dealing with a form of directed “association”, that is, a relation that is, at least in part, explanatory and thus runs from one of the relata to the other. That certainly seems to be what the pure normative functionalist believes. He claims that the dispositions are associated with the conceptually constitutive norms because people are in general disposed to be rational. In this view, the rationality of possessing certain dispositions explains their possession.

Recently, Bratman has also become explicit about the direction of the association, claiming that the normal causal roles of intentions result from the norms’ “acceptance” (Bratman 2014, 16). But it is unclear how a conception according to which acceptance of the norms is explanatorily primary is compatible with a conception according to which the attitude that is subject to the norms is defined in terms of its causal and normative environment. The idea of acceptance of a requirement on one’s φ-ing presupposes that one is able to identify what it is that is subject to the requirement one is accepting. But if the acceptance of the norms is supposed to explain behaviour according to them, the identification that makes acceptance possible cannot depend on the behaviour that is explained by the acceptance. Moreover, for the same reason, the identification cannot depend on the subjection to the norms of an attitude only identifiable as the attitude subject to those norms.

What I think we should, instead, be saying is that the association between the causal and the normative environment of intending is mediated by intending itself. According to the proposal I shall be advancing, there is a non-normative, psychological essence of intending and an explanation of why this descriptive feature, or bundle of features, is the ground of certain normative standards and also, perhaps in part as a result of acceptance of those standards, explains intention’s characteristic effects. There will turn out to be a kink in the explanation that derives from the fact that the structure we identify with intending is itself, to a certain extent, shaped by a specific kind of normative consideration. This important detail, however, doesn’t impugn the descriptive definition of intending.

One general difficulty with defining intention in normative terms concerns the point already made that the strict requirements of rationality to which we can give clear-cut formulations are no more than the tip of the iceberg. *EC*, *SI* and *IBC* only pick out cases in which the agent believes she won’t realise her primary intention unless she instantiates some further condition. Most cases of our acting or deliberating on the basis of intentions don’t involve subjectively necessary, but rather sufficient, conducive or opportune conditions. In those cases in which we act or form further intentions in spite of it not seeming to have become necessary for us to do so, knowing that we intend something and what the content of our intention is does not in general appear particularly problematic. That this is a problem doesn’t depend on an untenable view that conceptual criteria are necessarily employed by the bearers of the mental properties picked out by the relevant concepts. The fact that intentions are in general accessible to their bearers just makes it plausible that there are features of their own intending they are in touch with whose accessibility doesn't require counterfactual normative thinking.

Imagine for a moment something that I think is false: that the doctrine of "double effect" is true. Should this lead us to incorporate intending's role in the doctrine into the concept of intention? I don't think so. What I think we should be doing instead is asking what it is about intention that enables it to support this particular moral norm. An answer to the question would leave us wiser not only as regards the concept of intention, but also as regards the justification of the norm. The same strategy is, I suggest, also appropriate for standards of rationality.

According to this traditional strategy, what we want to know first and foremost is *in virtue of what* it is that the relevant norms attach to intentions. This is going to be a matter of understanding intention's special status in a person's optative economy in terms that don't themselves involve the rational norms to which intending is subject, but allow us to make sense of them.

The answer that I shall be arguing for grounds in the significance of deliberation. Intending is essentially, I shall be claiming, a way of preserving and realising the results of practical deliberation. For such a strategy to be successful, it must, first, paradoxically, be compatible with the fact that not all intentions are products of deliberation. Second, the explanation cannot be consequentialist in structure. As the requirements have an a priori status and cover each individual case, the answer must say more than that a general tendency to uphold the norms will, all in all, lead to more results of our deliberation being realised than not. It must explain the fact that the requirements are strictly applicable to individual cases, even where they are overridden by other requirements or reasons. Finally, the explanation will need to make clear what it is about deliberation that gives its results the character of to-be-realisedness or to-be-preservedness, even where deliberation begins from false premises. Most views on this question are versions of cognitivism, according to which the constraints on rational intention derive from the constraints on rational belief (Wallace 2001; 2006b; Setiya 2007; Schroeder 2009; Broome 2009).[58] In contrast, I shall be arguing for a noncognitivist answer.

According to the theory I shall be developing over the next two chapters, there is a conceptual connection between intention and the capacity for deliberation: intenders are necessarily practical deliberators. The central features of intentions result from their paradigmatic genesis through deliberation. Intentions are wants* with a particular kind of aetiology. Both intentions' characteristic effects and their normative environment can be explained by their aetiology. However, as the aetiology of

[58] The "cognitivism"/"noncognitivism" terminology was first applied to practical deliberation, that is, to intention formation through reasoning, by Bratman in his discussion of Velleman, where "cognitivism" entailed the view that intentions are themselves cognitive states (1991, 250f.). Bratman's later usage is wider, "cognitivism" coming to cover any conception for which intending entails believing. The decisive cognitivist feature becomes the explanation of the rational requirements on intention as derivative of the rational requirements on belief (Bratman 2009a, 30; 2009b, 229). Cf. below, Section 10.1.

the attitudes we call intentions is not unitary, intention is not, according to the theory, a unitary category. Non-deliberative intentions are deliberative intention surrogates; they derive from processes that are themselves surrogates for what I call "minimal deliberation". They share most of the causal and deontic environment of deliberative intentions. In discussing the intention-consequential requirements of rationality, we have seen that the applicability of the requirement of intention persistence is restricted to intentions acquired through deliberation. In Section 6.3.3, I argued that there is a weak doxastic condition on conscious intentions if the intentions were deliberatively generated. In Section 9.3, I will support the conjecture that there is no comparable condition on intentions with no deliberative aetiology. Otherwise, deliberatively and non-deliberatively formed intentions share enough for them to play the same role in the requirements of executive consistency, subordinate intending, intention consistency and intention-belief consistency.

After we have looked in detail at the deliberative genesis of paradigmatic intentions (Chap. 8) and the decisive features of their non-deliberatively acquired relatives (Chap. 9), I will argue that the normative predominance of commonalities reflects the sharing of a decisive descriptive feature. The fact that sharing this feature is decisive will itself turn out to have a normative explanation (Chap. 10). It results from a central feature of the personal life form, a demand that humans impose on their progeny, as the latter edge towards full agential status. This is a demand for what I call "personal responsibility". It is in this basic demand that mature agents *take* responsibility, a demand that pervades our life form, that the importance of deliberation grounds. The intention-consequential requirements of rationality name, I shall be claiming, conditions whose satisfaction is essential to the fulfilment of that demand. The basic pressure is thus both a priori for full agents and noncognitive. Thus, although the concept of intention doesn't hinge on specific norms or requirements, intentions, as opposed to mere goals, are mental states that could only be possessed by creatures whose life form is normatively structured. This, at least, is the line of argument I will be developing in the rest of this study.

Chapter 8
Deciding

8.1 Towards a Genetic Disjunctive Theory of Intention: The Itinerary of the Next Three Chapters

In the course of Chapters 6 and 7 – in the discussions of intention's doxastic features (Sect. 6.3.3), of its salience effects (Sect. 7.1.3), its characteristic stability (Sect. 7.1.2) and rational persistence (Sect. 7.2.3) – I have had occasion to remark on the importance that particular features of the aetiology of paradigmatic intentions – deliberation and decision – may have. When someone takes a decision to do something, she "commits herself" or "settles on" doing that thing as opposed to other candidates in her optative purview. There is thus an intimate connection between deciding and intending. Moreover, the intimacy of the connection looks very much like a conceptual matter. Leaving aside cases of deciding to bring about conjunctions of states of affairs,[1] it seems that necessarily if at t you decide to φ, you at t come to intend to φ. If intention were to be essentially a normative concept, that is, a matter of being subject to the standards of rational or reasonable intending, then deciding would be essentially a matter of taking whatever step is necessary and sufficient to put oneself in whatever state it is that subjects one to these norms. If, on the other hand, one is on the lookout for a descriptive specification of what it is to intend, a specification that helps explain why someone in that state is subject to norms of rationality and reasonableness, then deciding is an excellent candidate for a process describable in non-normative terms that could confer that status.

[1] These are special cases because deciding to bring about a conjunctive state of affairs doesn't seem to entail intending to bring about each of the conjuncts individually. On the questions this raises for my view and my answers to them, see my article "The Double Failure of Double Effect" (Roughley 2007c).

N. Roughley, *Wanting and Intending*, Philosophical Studies Series 123,
DOI 10.1007/978-94-017-7387-4_8

The former, normative reading of what it is to intend makes the process of coming to a decision look strangely mysterious from the point of view of the decider. It makes it appear a matter of lucky coincidence that the agent does whatever it is that triggers the normative machinery. However, deciding is a movement of mind that agents generally make in full consciousness of taking that step. An explanation of this point is, I think, something a theory of intention owes us. For this reason, there is considerable pressure on such a theory not only to look *downstream* – to the normative and characteristic dispositional consequences of intending – but also to look *upstream* to intention's genesis. Might it not be the case that the normative and characteristic motivational environment of intending is a consequence of the mental step involved in "committing oneself to" or "settling on" the relevant dynamic proposition? I shall be arguing that this question should be answered in the affirmative.

If that can be shown to be correct, we will have come a long way towards an understanding of the mental states we call "intentions". However, we will still not have achieved the goal of a comprehensive theory. The reason lies in a consideration that no doubt explains why the route I am proposing has frequently been shunned. This is the fact, on which I have repeatedly remarked in the preceding chapters, that not all those attitudes we label "intentions" or express by the appropriate first-person phrases are generated by decisions. This fact would be a decisive objection to locating the central descriptive feature of intending in deciding were the thesis of the unity of intention (Sects. 6.1.2 and 6.3.3) to be true. I believe that it isn't and that therefore the force of this objection can be shown to evaporate. I shall be arguing that the unity of intention thesis is false because those non-decisionally generated attitudes we tend to see as intentions are in fact mental states that are sufficiently similar to paradigmatic intentions to be picked out by the same linguistic mechanisms and to fall under the scope of *most* of the same requirements.

That, then, is a preview of the argumentative strategy I shall be developing in the next three chapters. My itinerary is the following: I shall begin in this chapter by taking a step back from intention to focus exclusively on the phenomenon of decision. In Chapter 9, I shall attempt to show that an aetiology involving the process thus outlined is what bequeaths us paradigmatic intentions, before going on to offer an analysis of non-paradigmatic, that is, non-decisional intending. In the last chapter, I return to the question of how the disjunctive upstream conception that emerges can make sense of the key features of the intentional syndrome, the intention-consequential requirements.

My account of decision in this chapter is developed in the following steps. I begin by inquiring whether two theories in which decision is writ large, decision theory and existentialism, can provide leads as to how the concept is best understood (Sect. 8.2). Although the answer turns out to be negative, a key question emerges from the confrontation of the two positions, namely whether making a decision is a type of action. In order to answer this question, I ask how we are to understand the notion of "making" at work here. The apparently parallel locution "making a judgement" suggests a comparison of these two forms of ostensibly active mental behaviour. Both are, I argue, necessarily preceded by mental processes that can be termed "minimal deliberation" and "minimal inquiry" respectively (Sect. 8.3). After an analysis of the way the latter leads to judgement (Sect. 8.4), I make use of the results of the analysis in order to clarify deciding's specific relation to deliberation, in particular to "minimal" deliberation. In doing so, I will discuss a further – intention-antecedent – requirement of rationality, which relates decisions to conclusive reasons judgements (Sect. 8.5). It turns out that neither judgements nor decisions are actions, although for slightly different reasons (Sect. 8.6). These steps clear the way for an understanding of deciding as a particular species of occurrent wanting* (Sect. 8.7).

8.2 Decision: Two Not Particularly Helpful Theories

It seems sensible for an investigation into the concept of decision to inquire what there is to be learnt from the two most publicised theories in which the term "decision" plays a prominent role. Consider, first, what is known as *decision theory*. What the theory offers is basically the spelling out of the consequences of one grounding premise, namely that rational action is action carried out in order to realise maximum estimated desirability. In order to answer the question of which actions under which circumstances meet this criterion, the assumption has generally been that it is irrelevant whether the actions up for consideration are really preceded by any particular kind of mental occurrence that might be labelled a "decision". Thus, in decision theory, "decisions" are in a sense the locus of human rationality whilst paradoxically remaining completely untheorised.[2]

If decision theory is prepared to talk as if all human action were caused by decisions, *existentialism* makes the paradigm of decision mental action under exceptional circumstances. The relevant circumstances are those under which an agent is faced with a choice between incommensurable, usually moral values and therefore needs to take a rationally unjustifiable "leap" in opting for one rather than for another course of action (Luebbe 1965, 19). In contrast to decision theory, existentialism makes it definitive of decisions that they are indeterminable by rational considerations; they are mental steps that have to be taken blindly when the grounds give out for doing one thing rather than another.

Now, although the notion of decision at work here is in a number of respects opposed to that of decision theory, there is a sense in which the two are, surprisingly, not so far removed from one another. Where the two theories agree is that

[2] This is explicit in Philip Pettit's classification of Bayesianism as providing what he calls a "normalizing" explanation of behaviour that need have no recourse to "interpretation", a form of explanation that makes essential reference to the way subjects of explanation see things. According to Pettit, decision theory doesn't even require that the bearers of utilities and subjective probabilities are conscious. Some such subject of explanation could be an automaton whose "decisions about what to do might just happen without anything that approximates an interpretation of the situation" (Pettit 1996a, 187).

there is nothing of philosophical importance to be said about what decisions *are*. Whereas decision theory leaves tends to leave decisions untheorised because there appears to be nothing more to be said about them after the relevant calculations have been carried out, existentialism leaves them untheorised because they appear inexplicable. Either way, decision remains obscure.

Nevertheless, confronting the two perspectives does raise a question an informative response to which promises to provide a way to get a structured hold on the phenomenon. This is the question as to what kind of occurrence a decision is. From the perspective of decision theory, decisions generally appear to be theoretical constructs, hypothetical mental events that register the result of a calculation of desirability, in the standard version: of subjectively expected utility. In contrast, existentialism explicitly conceives deciding as a special kind of mental action.

For all its merits, decision theory is at the very least misleadingly named. This is true whether the theory is understood normatively or descriptively. Perhaps understandably, normative decision theory has not seen it as part of its remit to say anything about what deciding is. The conceptual apparatus of desirabilities and probabilities is designed to be applied to what agents do. As such, it has been generally assumed that the content of putative decisions can be read off from the behaviour of the theory's subjects. Descriptive versions of decision theory tend to equate deciding with the acquisition of either an *axiological optimality belief* as a result of the relevant calculations or of the *unrivalled motivation* that such a belief generates.[3] Such more or less explicit identifications raise the question of their adequacy to everyday understanding, a question that is, I think, clearly to be answered in the negative. We can broadly distinguish three kinds of case in which a gap can open up between a decision to φ and either a belief that φ-ing is the best option in the circumstances or the unrivalled motivation to φ.

8.2.1 Noncommittal Motivational or Evaluative States

In cases of the first kind, an agent may be in a noncommittal state relative to some prospective future action in spite of being the bearer of either the unrivalled motivation so to act or the belief that thus acting would be her best option.

In cases of noncommittal unrivalled motivation, a person has at t_1 not decided to φ at t_2, although she is, relative to the options she believes will be available to her at t_2, most strongly motivated to φ. Someone thinking about what to do next Saturday night might be most motivated to go to see a certain play, but might nevertheless still

[3] The first variant is advanced by Christoph Lumer (2005, 245; 2007, 159ff.), the second is suggested by Philip Pettit, for whom deciding is forming a "preference" for a prospect (Pettit 1991, 156), where that preference is a dispositional state explicable by the strengths of the beliefs and desires codified in terms of subjective probabilities and utilities. Pettit is critical of the assumption, which is "natural if only because of the name given to the theory", that decision theory is supposed to be a complete account of preference formation (1991, 160).

be undecided about what to do. She may want to think about it a bit longer or to wait and see whether any further options crop up before taking the step of "committing herself to", or "settling on" the theatre option (cf. Bratman 1987, 18f.; Mele 1992a, 154ff.; 1995a, 60ff.).

Moreover, the agent's unrivalled motivation may well be the result of her believing her φ-ing to be her best option at t_2 without this involving that she have committed herself to φ-ing. This is particularly obvious where the agent's belief is unreflectively acquired, perhaps the effect of direct inculcation by her parents or educators. Deciding involves taking a step that appears in some sense active.[4] But even where an "active" mental step has been taken in coming to an axiological optimality judgement, that still leaves open whether the agent has decided to do what she judges to be best. As Jay Wallace has pointed out (2001, 94f.), putative Moore-paradoxical sentences can again be of help in clarifying the separability of the mental steps at issue here: "I have become convinced that φ-ing is my best available option in situation *s*, but I haven't decided yet whether to φ" is not only a coherent conceptual possibility, but surely the sort of thing that people on the street think not infrequently.[5] Take Stu, who believes that the best thing for him to do in response to his worrying chest pains would be to consult a doctor. Being the stubborn person that he is, he has not, at least not yet, decided to go to the doctor's. Like unrivalled motivation, optimising value judgements can be noncommittal.

8.2.2 Deadlock-Terminative Deciding

A second phenomenon that demonstrates the lack of equivalence between decision and the acquisition of an optimising evaluative belief has been given considerable publicity under the rubric "Buridan cases" (Ullmann-Margalit and Morgenbesser 1977; Elster 1985, 65ff.; Bratman 1987, 11f.; Mele 1992a, 67ff.). In such cases, an agent comes to a decision to φ rather than to ψ in spite of believing that his reasons for φ-ing and for ψ-ing are equally good. Buridan cases are only one of a set of different types of example in which decision is possible in the face of inconclusive evaluation of the options between which decision is required.

[4] Daniel Dennett rejects this, or at least rejects the claim that we should take the appearance seriously. His grounds are causal. If I seem to decide to get out of bed but then proceed not to get up, then, Dennett suggests, I cannot really have decided. Conversely, decisions can take place in me without my being aware of the relevant step having occurred, except perhaps retrospectively (Dennett 1984, 80). Of course, philosophers are free to use the term "decision" to mark only causally effective occurrences. As the everyday concept of decision involves no such causal necessitation, such a terminological proposal rests on the belief that the everyday concept is philosophically irrelevant.

[5] It is no coincidence that the examples by means of which I illustrated the gap between wanting* and value judgement in Section 4.3.2 are also relevant here.

In cases of this first kind, option evaluation is insufficient to determine the content of decision because the upshot of weighing is precisely a judgement of evaluative indifference or equidesirability. In such cases of what Ullmann-Margalit and Morgenbesser call "picking" rather than choosing, the agent often sees a type of option as preferable, but has no reason to go for one token rather than another of the same type. Mass production and consumption confronts us continually, for instance in supermarkets, with the requirement to form an intention in the face of the equidesirability of the option tokens before us.

In related cases, an agent may believe that it would be undesirable for him to incur the costs that would be necessary in order for him to come to an accurate judgement. Such second-order evaluative judgements are also commonplace on shopping expeditions, where someone may be convinced that detailed examination of all the apples in a box would permit an optimality judgement, but where discerning and weighing the exact differences does not seem worth the effort. In such cases, the assumption that there are evaluatively relevant differences combines with the lack of any will to discover them. Agents frequently develop no such will because they are well aware that they are able to come to a decision in the absence of any determinate value judgement – as they are in the presence of a value judgement with no decisive content.

In a third group of cases, an agent again assumes, or even knows, that there are significant evaluative differences between options facing her. As in the second group, she doesn't take the step to the discovery of the differential location of those differences. Whereas in the second group, the reason for the omission is a second-order evaluative belief, here the reason is epistemic impossibility.[6] An example is provided by a person standing at a road junction from which two forks lead to places he assigns very different values, but where the situation provides no indication of which road leads where. We tend not to be comfortable with such situations.[7] Nevertheless, we are generally able to plump for one fork rather than the other.

It is arguably the capacity of decision to outrun evaluation that gives existentialism an initial plausibility. Typical existentialist cases are, of course, not concerned with road junctions and supermarkets, but with matters that appear to call for heroic choices. Such a characterisation seems appropriate where the values associated with the options are incommensurable, or are seen by the agent as such. The young man in Sartre's famous example is unable to find a common standard against which to measure the desirability of loyalty to, and love for an individual family member on the one hand and the duty to fight for the freedom of a collective on the other (Sartre

[6] Ullmann-Margalit and Morgenbesser (1977, 763ff.) label these cases of "picking by default".

[7] Christoph Lumer claims that discomfort in these kinds of case "confirms" the contention that intentions are optimising value judgements as the contention explains the difficulties we have in such cases. He sees in the resort to random devices – whether deliberate, in coin tossing, or automatic and internal to the workings of the agent – as compatible with the claim of identity between optimal evaluation and intention (Lumer 2005, 257; 2007, 161f.). But as the intention in question has a content that picks out one option token, whereas the evaluation does not, there can be no question of conceptual identity.

1946, 35f.). Again, the phenomenology of decision seems to tell us, and certainly existentialists have claimed, that he can overcome such a situation of evaluative and motivational deadlock *by simply deciding*.

Note that the capacity to simply decide is presupposed by anyone who claims that potential Buridanian asininity can be cured by the deliberate employment of some random device such as tossing a coin. Clearly, no evaluation can provide conclusive reasons for coin tossing rather than applying some other device, for tossing coin *A* rather than coin *B* or for assigning one particular option to heads and the other to tails. Random devices, then, can only help us come to a decision if we already capable of "simply deciding" in our selection and setting up of the random device itself.

Evaluation can, then, for various reasons come up short of providing decisive considerations for one action rather than another: it might take a number of options to be equidesirable; it may be unable to find a measure that allows precise comparison; it may find that the world deprives it of the information that would be decisive or it may itself reject the effort required to acquire that information. If the motivation of rational agents aligns with the results of their evaluation and evaluation is the only source of motivation, then it's clear that rational agents are in trouble. However, it's equally clear that we are not the rational agents in question. Can decision theory make sense of our ability to escape from asininity? J.H. Sobel thinks that it can certainly *allow for* it, as Bayesianism, he argues, "says that [rational agents'] actions are maximizing, not that they are uniquely maximizing" (Sobel 1994, 243). But even if it is correct that Bayesian decision theory isn't incompatible with deadlock-terminative decision, the important point for our purposes is that it leaves a large explanatory hole at one of the places in our conceptual fabric where we most readily think of ourselves as deciding.

8.2.3 Counterevaluative and Countermotivational Decision

It seems, then, that a decision to φ cannot be equivalent to the acquisition of either the unrivalled motivation to φ or of a belief that φ-ing is the best course of action in the circumstances. That neither phenomenon is sufficient is shown by the compatibility of either with a lack of commitment concerning the action in question. That an optimising evaluative belief is unnecessary is shown by the deadlock-terminative role that decision can, and surely does frequently play. The phenomenology suggests that a prior motivationally unrivalled want* is also unnecessary. That this is so for rational motivation follows from decision's capacity for evaluative tie-breaking. Nevertheless, the possibility remains that some non-conscious mechanism may generate prior motivation and assign it to options, thus at least enabling, if not determining a corresponding decision.

A brief look at the field of weakness and strength of will suggests that there are also cases which, taken individually, speak against both the necessity and the sufficiency of either evaluative or motivational conditions on decision.

Let us begin by returning to Stu and fleshing out the details of his story slightly. In Section 8.2.1, Stu believed that it would be best for him to consult a doctor about his persistent chest pains, but nevertheless remained undecided about whether to go to the doctor's. It certainly seems to the person on the street that Stu can go a step further and, in the face of his axiological optimality judgement, decide not to go. Or take Reg, who believes that the satisfaction derivable from taking revenge on someone who has harmed him in some small way is negligible compared to the trouble it is bound to cause him. Nevertheless, Reg resolves to exact vengeance. Note that these cases of vengefulness and stubbornness – cases that again fit Stocker's bill of "desiring the bad" (cf. Sect. 4.3.2) – are examples of akrasia that are it would be peculiar to describe as manifesting weakness of will.[8] Rather, there is a sense in which the agent's will is too strong. Such "wilful" agents have, so it seems, taken counterevaluative decisions.

Someone wedded to the claim that decisions necessarily either are or reflect evaluations will claim that Stu and Reg must have, perhaps inexplicitly, reevaluated their situation, as the only attitude that could, for conceptual reasons, override an explicit evaluative belief is another evaluative belief. This is a view backed by the weight of tradition from Socrates to Davidson. Nevertheless, the aprioristic reasoning at work requires backing of a systematic order if it is to displace the everyday phenomenology. Its primary motivation lies in the conviction that human agents are in some sense behaving rationally whenever they are genuinely acting. But as cases of deadlock termination are sufficient argument against the claim that genuine actions are necessarily guided by evaluation, such an a priori view is already undermined. Absent commitment to such an a priori view, we have every reason to stick to appearances. These involve Stu and Reg taking counterevaluative decisions.

People on the street, then, seem to believe that decisions to act that conflict with optimising value judgements can manifest either wilfulness or weakness of will.[9] Particularly with regard to weakness of will, decision theorists have been prominent in holding onto Socratic apriorism according to which any apparently counterevaluative decision must be or reflect a short-term evaluative reversal.[10] In

[8] For the claim that weakness of will is a phenomenon clearly distinct from akrasia, see Richard Holton's important article (Holton 1999; more or less reproduced as chapter 4 of Holton 2009, 70ff.). I suggest some modifications to Holton's proposal in Roughley 2008a and 2008b.

[9] In line with the articles mentioned in the previous note, I see the identification of akrasia and weakness of will as at the very least unhelpful, as it hides from view an important kind of case for the study of intention. Nevertheless, it is plausible that the majority of cases of weak-willed action do also involve akrasia. As to whether persons on the street identify weakness of will with what the tradition has called 'akrasia', see the discussion between Mele and Holton (Mele 2010; May and Holton 2012).

[10] There are various decision theoretic proposals on offer as to when evaluative (or "preference") reversals are to be seen as problematic: if they come about without any change in subjective probabilities (Jackson 1984, 13f.); if they result from our unhelpful disposition to "hyperbolic" discounting (Ainslie 2001, 28ff.); or if they result from strong emotions and conflict with stable, long-term evalulations (Lumer 2005, 257).

as far as an explanatory psychological theory is committed to reconstructing "folk psychological" processes,[11] it ought to account for the everyday conviction that an agent can decide to eat a large piece of cake or smoke a cigarette in the face of her unshaken belief that doing so is the very thing that her values forbid in the situation.

One way of holding onto the conception of decision as evaluation in the face of weakness of will involves interpreting weak-willed decisions as re-evaluations of some state of affairs – and of the associated behaviour – that fail to take into account the change in, or stability of the agent's subjective probabilities relative to some earlier evaluation (Jackson 1984). Take a smoker such as Nosmo, who at breakfast evaluates his continuing smoking behaviour as significantly suboptimal in the light of his recently acquired conviction that smoking drastically increases the probability of his health deteriorating. Later in the day he finds himself evaluating having a cigarette as the best action in the circumstances – in spite of the stability of his opinion concerning the health risks of so acting. According to the decision theoretic view under consideration, Nosmo's irrationality consists in an evaluative change that is independent of any subjective probability change.

Certainly, coming to see a form of behaviour as more worthwhile than one had previously taken it to be without any change in non-evaluative beliefs will often be an irrational movement of mind.[12] However, where the attitude that ends up controlling an agent's action is arrived at independently of the consideration of reasons, it is anything but obvious that the attitude in question is appropriately described as an evaluation. This point is easily obscured by the decision-theoretic terminology, where "preference" is the catch-all for potentially motivating attitudes that are, as the etymology says, "placed before" rivals in one way or another. The evaluative interpretation of the relevant form of "placing before" is smuggled in under cover of the plausibility that when someone φ-s, this will always be because they prefer φ-ing to not φ-ing – in a broad or vague sense of "prefer". Strip away the evaluative reading and we are left with the indeed plausible claim that Nosmo forms the counterevaluative "preference" – or privileged want* – to have a cigarette. Understanding decision seems to involve the challenge of making sense of want* formation in as far as it can be counterevaluative.

Finally, one might wonder whether there is such a thing as countermotivational decision: cases in which an agent decides to φ, in spite of being most strongly motivated to behave in some way incompatible with his φ-ing. Return to Nosmo, who decides in the spring not to smoke at his cousin's wedding that is due to take place in the autumn. Common sense tells us that Nosmo's thus deciding does

[11] Lumer argues that a purely "hydraulic" interpretation of the decision theoretic apparatus, one independent of the subjective processes the folk associate with decisions, precludes an understanding of intentions (Lumer 2005, 245; 2007, 164). According to Pettit, any interpretation of decision theory as rejecting explanation in terms other than degrees of belief and strengths of desire is at odds with the our everyday deliberative perspective (Pettit 1996b, 243ff.).

[12] Pace Jackson, this is not sufficient for irrational evaluation revision. Someone might simply find herself in a better, perhaps more relaxed mood and, as a result, come to appreciate better the value of some activity the thought of which had previously left her cold.

not require that his strongest want* concerning his smoking or not smoking at the wedding be that he not smoke. On the contrary, his decision can co-exist with the counterfactual truth that, were he to be teletransported to the wedding, the first thing he would do would be to light up. Perhaps he has a plan to transform the balance of his motivation, involving nicotine chewing gum and jogging. If his plan works, then Nosmo's desire to smoke will weaken and perhaps even become weaker than his want* to jog. But whether or not he succeeds, it is perfectly clear both that Nosmo can decide to do so and that he has a long way to go motivationally in order to be able to realise that decision.[13]

Cases such as these, in which a countermotivational decision is taken with the aim of realising an optimising value judgement, confer some plausibility on evaluative conceptions of decision. These are cases in which, unlike in the cases of Stu and Reg, decision aligns with evaluation against motivation. Although it is plausibly a conceptual truth that our actions necessarily express our strongest motivation to act,[14] the claim that decisions concerning future actions are necessarily the result of prior motivation is a strong claim that flies in the face of the phenomenology.

Of course, if a decision doesn't represent prior motivation, but the agent ends up intentionally realising the decision's content, some sort of motivational transformation must take place after the decision. This might happen as a result of Nosmo engaging in forms of self-manipulation such as chewing nicotine gum or engaging in sport. However, we tend to believe that an agent's deciding to φ can also have the direct effect of increasing her motivation to φ. It seems perfectly possible that Nosmo's decision not to smoke might suffice to generate the surplus motivation requisite for him to give up smoking. In as far as he is rational, the appeal to requirements such as *DIP* and *SI* can help to explain this: if he doesn't come to believe he has reasons to re-enter deliberation or to abandon his intention, he is subject to the presumption to uphold it and as long as he does so, he is subject to the requirement to adopt the means he believes necessary for realising his decision. Rational agents are disposed to act in accordance with the principles of rationality. In as far Nosmo is rational, then, his consciously deciding to give up smoking will tend to lead to an increase in his motivation so to do.

[13] Audi (1986, 23) simply denies this. According to Audi, the claim that Nosmo has decided countermotivationally is simply incoherent. However, he offers no argument as to why this should be so. We clearly don't have the impression that someone who claims to have made such a decision cannot, for a priori reasons, know what he is talking about. One strategy open to the motivational determinist here is to see countermotivational decisions as generated by higher-order wants* concerning the rearrangement of first-order motivational patterns. A problem with such a strategy is that it makes the contents of our decisions our own motivational patterns, rather than our own future actions. Nosmo doesn't decide to strengthen his want* not to smoke, but simply decides not to smoke.

[14] Somewhat more precisely: An agent's intentional performance of φ at *t* entails that she is at *t* more motivated to φ than to perform any other action or inaction of which she consciously believes that its performance would be incompatible with the performance of φ. This retrospectively applicable principle expresses what I take to be the uncontroversial truth in Davidson's predictive principle *P1* (Davidson 1970, 23).

8.2.4 Folk Psychological Seeds of Existentialism

In contrast to both wants* and evaluative beliefs, decisions are necessarily events with a phenomenal aspect, that is conscious occurrences. Moreover, the phenomenology of decisions presents them not merely as conscious attitudinal occurrences, but as occurrences that are in some sense active. The relevant sense of activity involves their being under their bearers' control – again in some specific sense. The kinds of case discussed in the last two sections – in which decisions terminate evaluative deadlock or run counter to either prior motivation or evaluation – present particularly salient examples of such activity or exercise of control.

The coherence and apparent empirical occurrence of such cases makes it plausible that there is at least *something* right about the existentialist perspective. However, that perspective is as a radical as it is unsatisfactory. A decision is, according to Sartre (1943, 450), "the very act by which I project myself towards my ends". Sartre thinks of decisions as actions, indeed as the paradigmatic actions. All Sartrian actions are, for conceptual reasons, characterised by the exercise of indeterministic freedom (1943, 436). What Sartre calls "reflexive consciousness" is supposed to suspend every previously given motivationally relevant fact about an agent, so that momentary consciousness bears complete responsibility for everything a person does. Indeed, Sartre goes so far as to deny the importance of practical deliberation, which he sees as a person's attempt to "discover" (1943, 451) facts about herself, none of which, however, can be decisive in the way that the momentary self-projection of the agent has to be.

Applied to cases such of countermotivational or deadlock-terminative decision, this position is not without a certain phenomenological adequacy. Standing at the crossroads, struggling against the desire to smoke or torn between mother and the Free French, people tend to feel that it is in some sense entirely "up to them" how they end up deciding. And deciding in these contexts is, often at least, accompanied by a strong sense of one's own activity.

A theory of decision has to make sense of these cases. It also has to clarify whether what is to be said here is, as Sartre clearly thought, extendable to other mental events we might describe as "decisions", but which are accompanied by no such active phenomenology. There can clearly be no question of a coherent causal theory of action postulating indeterministic freedom at the heart of every action[15] or of it relegating practical deliberation to irrelevance. Nevertheless, the up-to-usness

[15] Contemporary incompatibilists typically attempt to circumscribe certain kinds of situations in which indeterminist freedom is at work. See, for instance, van Inwagen 1989, 415–8 and Kane 1996, 125ff. Even O'Connor, who criticises van Inwagen's "restrictivism", is clear that there remain a range of cases in which libertarian freedom is absent (O'Connor 2000, 107).

that appears to characterise our decisions needs explaining. In particular, we require an explanation of the relationship between this up-to-usness and the sense in which ordinary actions are due to their agents.

The most obvious, and perhaps most elegant explanation would simply classify decisions as actions of a particular kind.[16] Indeed, a number of authors have taken that route (McCann 1986b, 133ff.; 1998; Mele 1995a, 17; 2000; 2003a, 209ff.; Pink 1996, 3ff., 17ff., 90ff.). I turn to this claim in the next two sections. Before in Section 8.6 I explain why I think it should be rejected, it is necessary to take a slight detour, which paves the way both for this negative claim and for the constructive proposal that I will then go on to advance.

8.3 Deciding and Judging

Deciding is clearly a form of mental behaviour in the sense introduced in Section 2.2.2: it is one of those mental things that we 'do' like getting annoyed, worrying and daydreaming. Our question is whether deciding is a form of intentional action, belonging to those mental behavings that are "due to us" in a stronger sense. Certainly, we talk of "making" or "taking" decisions, active linguistic idioms. And when we make decisions, so it seems, we do so because we "want" or "intend" to do so. In both points, however, deciding is special. Vindicating a claim as to whether decisions are actions or not will require clarity on both the *kind of "making"* involved and on the precise *way* in which a want* or intention to decide can *lead to* such "making".[17]

There is, among the forms of our mental behaviour, another case that seems analogous. This is that of judging. Like decisions, judgements are "made". And they also appear to be made because a person wants to make them. The parallel is given further linguistic support by talk of "deciding *that*" something *is* the case, which parallels that of "deciding to" do something (Kaufman 1966, 25; O'Shaughnessy 1980 II, 297ff.; Mele 2000, 81f.).

The first thing to note about both types of mental occurrence is that, in their most salient form, they are preceded by reflection of some kind. Judging involves forming a belief and typically involves doing so on the basis of prior mental, and perhaps

[16] Decision theory has tended not to be interested in the rationality of mental events that precede action if they are not manifested in the action specified in their content. This question did find its way into the literature in response to Kavka's toxin puzzle, in which intention formation and corresponding action are conceptually forced apart by a thought experiment in which their rationality conditions clearly diverge. In response to the puzzle, some decision theorists have conceptualised decisions as further things that agents do that, like fully fledged actions, can be represented as nodes in decision trees (Soble 1994, 242ff.; van Hees and Roy 2009, 67ff.).

[17] Harry Frankfurt wonders relatedly about the character of the "making up" involved in "making up one's mind" (1987, 173). My response to Frankfurt's answer, which I think implausibly generalises from a restricted kind of case, is given below in Section 8.7.2.

other activity. A paleoanthropologist who is judging the age of fossilised bones will, before reaching his conclusion on the matter, take measurements, make comparisons and weigh reasons for competing answers. Similarly, someone faced with a practical decision will normally think through the consequences of the alternative courses of action, attempting to see whether there are stronger reasons for opting one rather than the other.

It will be helpful to have terms for these two species of reflection. I shall refer to pre-judgemental "theoretical" reflection as *inquiry* and to pre-decisional "practical" reflection as *deliberation*. Neither term is important in itself, as the "seeking" etymologically contained in the former and the "weighing" that is the root of the latter are equally appropriate for both in their developed forms. In both cases, the bearer of the process is seeking a content whose tokening in the appropriate attitudinal mode will dissolve some form of uncertainty he is faced with.[18] And in both cases, this is characteristically done by "weighing" the reasons for the alternatives. I shall argue in a moment that the core structure of either type of reflection can be instantiated in the absence of the latter feature.

Now, the reflective process that precedes the formation of either type of attitude is itself a complex intentional action. Wanting* to make a judgement or to take a decision, we engage in reflection *on the basis of which* the movement of mind of the kind we are aiming at should take place. The key question for understanding both notions is how "on the basis of" is to be interpreted. Clearly, judgements and decisions, unlike perceptual beliefs or homeostatically caused desires, don't just suddenly arise in us unbidden. Some form of genuine action is necessary for their production. What is not phenomenologically obvious, however, is whether the inquiry or deliberation that precedes the relevant mental move is all there is to the mental activity or whether judging and deciding are themselves additional mental actions.

The question takes on a certain urgency when we consider the fact that the cases of judging and deciding that are preceded by extensive reflection are only the most salient examples of the phenomena. In both kinds of case, so it seems, we can move very quickly to take the mental step. A driver may glance briefly at a parking space and then judge that his car will fit in; a shopper may hesitate for a second in front of a shop before deciding to go in.

How, then, are these split-second judgements and decisions to be distinguished from the mere non-active acquisition of beliefs or wants*? The answer, I believe, grounds in the fact that both mental moves must be responses of their bearers to some form of uncertainty. Both forms of uncertainty are expressed in a question that the agent puts to herself: the theoretical question "What is the case?" or the

[18] Various authors (Hampshire and Hart 1958, 2ff.; Ginet 1970, 122; Grice 1971, 65f.; O'Shaughnessy 1980II, 301; Magill 1997, 87ff.; Watson 2003, 125) have seen both deciding that and deciding to as essentially matters of resolving uncertainty, an insight that has not infrequently been clouded by the belief that uncertainty is necessarily cognitive. That this is not the case is recognised by O'Shaughnessy, Magill and Watson.

practical question “What am I to do?” Note that either question can be posed in the blink of an eyelid and be phenomenologically unobtrusive: unlike small children, we have a lot of practice in these matters.

Putting either of these questions to oneself involves taking a – perhaps barely noticeable – step back from the immediate flow of one’s action, much of which is carried along by the constant interplay of beliefs and wants* backed by the requisite motivational force. That step may only involve a moment’s hesitation, as when the driver wonders whether the car will fit into the gap or not, before immediately concluding one way or the other. If the conclusion is that it will only just fit in, that will in turn generally raise the practical question as to whether he should try to manoeuvre it into the space, which again will generally receive a fairly immediate answer.

In spite of the speed with which we often proceed from the question to the movement of mind that resolves the relevant form of uncertainty, I think that the core phenomenon of either form of reflection is at already at work in such cases. To see this, let us dwell for a moment on what happens in the theoretical case, that is, in what I shall label “minimal inquiry”.

8.4 Minimal Inquiry and Judgement

A minimal inquirer, firstly, takes a step back from the immediate flow of behaviour. This is generally because, for one reason or another, he sees himself confronted with an epistemic gap: there is something he doesn’t know. Moreover, this must be an epistemic gap that he is concerned to overcome. There are many reasons why he might have such a concern: the gap might prevent him from achieving some practical aim, it might constitute a hole in his otherwise complete knowledge of some area or he might just be irritated by the fact that here is something that he doesn’t know.

Whatever the precise background motivation that generates the concern to overcome the epistemic gap, some such concern is a necessary condition of any minimal inquiry. Note, though, that it is possible for someone to realise that they are ignorant in some respect, be bothered by the fact, but then push the matter to one side and either consciously or non-consciously live with the relevant form of ignorance. For minimal inquiry to be set in motion, the relevant concern must be backed by sufficient motivational strength to move the person to take mental, and perhaps other steps that he takes to be conditions necessary or sufficient for, or conducive to the gap being closed.

Of central importance here is the fact that not any kind of mental step can be coherently taken as apt to close an epistemic gap if the consequent movement of mind is to count as a judgement. For instance, a paleoanthropologist who is having difficulty explaining certain discoveries might be convinced that taking L.S.D. would cause beliefs about these matters to pop into his consciousness. But, so long as he believes that the beliefs thus induced would be mere products of epistemic distortion, he will not take the drug in the conviction that doing so is a part of his inquiry.[19]

This is because both the concepts of inquiry and of judgement tie the relevant kind of preparatory activity to a *restricted form of motivation*. This conceptual restriction grounds in the fact that beliefs "aim at truth" (Sect. 4.1.1). Both inquiry and judgement require that the movements of mind resulting in the acquisition of the relevant belief be made in order that the belief attain believing's internal "aim". In other words, the motivation to close the epistemic gap must be motivation to do so by forming a belief whose content is true. In the simplest form of such inquiry, the person has simply asked himself whether p or $\neg p$ is the case. Movements of the mind that carry him to a belief that can qualify as a judgement must therefore be motivated by the want* to believe that p if p is indeed the case or to believe that $\neg p$ if $\neg p$ is the case. To be more precise: the inquirer must want* to believe p if p is the case and *because p* is the case. Inquiry must, in order to be inquiry, be powered by wants* with contents of this structure. That is, such motivationally efficacious *inquisitive wants**, as I shall call them, are internal to or constitutive of what it is to engage in inquiry. An inquirer, then, must be the bearer of a want* whose content consists of the disjunction of two (or more) conditionals, which, in the simplest case, can be rendered thus[20]:

$$\mathrm{W}^*(\mathrm{p} \rightarrow_{\mathrm{c}} \mathrm{B(p)} \ \mathrm{v} \ \neg\mathrm{p} \rightarrow_{\mathrm{c}} \mathrm{B}(\neg\mathrm{p}))$$

These, then, are the materials I am suggesting are sufficient for the inauguration of minimal inquiry: the registering of an epistemic gap, i.e. the genesis of theoretical uncertainty; the concern to acquire a belief that closes the gap, a belief that will do so because of the truth of its content; and sufficient motivational force behind that inquisitive concern to move its bearer to take what he believes are appropriate steps to realise it. These steps will usually consist in the weighing of reasons and counter-reasons. Nevertheless, they are not necessary for an agent to be engaged in *minimal* inquiry.

Restricting ourselves to these minimal conditions makes sense here because their satisfaction is all that is necessary for us to move on to clarify the concept of judgement. I propose that we understand a judgement as that occurrence of belief formation whose product appears to its bearer to satisfy the inquisitive want* that had initiated the episode of (minimal) inquiry responsible for that doxastic occurrence. That is, a judgement resolves the theoretical uncertainty or answers the epistemic question that had triggered the episode of (minimal) inquiry.

[19] On the other hand, a psychedelic paleoanthropologist, who is convinced that L.S.D. increases its taker's cognitive capacities, might well take the drug as part of his inquiry.

[20] "$\rightarrow_{\mathrm{c}}$" is to be read as "if and because of".

This conception takes account of the fact that a judgement need not be "considered", but may be "rash" – that is, passed with insufficient consideration of relevant reasons or counter-reasons – whilst still enabling us to make sense of the active part played by a judgement's bearer in its production. The proposal also helps to clarify what it means to judge "on the basis of" inquiry and what constitutes the "making" of a judgement. The judgement is, I suggest, based on the antecedent inquiry in a sense that is both causal and semantic. That is, the final movement of mind is, firstly, the *effect* of the agent's previous mental activity and, secondly, is seen by the inquirer as providing an *answer* to the question that initiated the inquiry. From the first point it follows that the notion of "making" at work here does not strictly apply to the event of the belief's formation, but only to the motivated activity of its preparation. Of course, in order to qualify as a judgement, an event of belief acquisition cannot simply happen to its bearer – it cannot be an undergoing in the sense opposed to behaviour (Sect. 2.2.2). It must come about as a result of things he does to satisfy his inquisitive motivation. In maximal forms of inquiry, for instance in a laboratory experiment, an inquirer will take all sorts of measures to ensure that the belief that ensues will be true. But in the last resort he still has to wait and see what his belief will turn out to be.

Unlike many other "makings" – of paper aeroplanes, of cups of tea, of love – making a judgement is not performing an action. We do not – we logically cannot – judge that *p* merely because we want* or intend to judge that *p*. There are wants* involved, indeed necessarily involved in our judging. However, these are inquisitive wants* with *disjunctive content*. It is true that cases can be constructed in which wanting* to judge that *p* contributes causally to the judgement that *p*: if someone is fairly confident that *p* will turn out to be true, his desire to acquire the true belief that *p* might motivate his inquiry into whether *p* is true or not. Nevertheless, if the agent really is engaging in inquiry, then its result must still be brought about by his attempting to satisfy the disjunctive inquiry-constitutive want*. In this somewhat unusual case, the inquisitive want* is itself acquired because of the agent's belief that satisfying it will satisfy his want* to judge that *p*.

The case against judgements being actions can also be made fairly simply by considering the question of judging's *intentionality*. On the one hand, the want* to judge that *p* cannot directly cause us to judge that *p*. Further, the judgement that *p* is not an action part that might derive its intentionality from some more complex sequential whole, as a dance step might be thought to do from a dance. Thus, specific judgements are not intentional actions. On the other hand, judgings are clearly not unintentional actions. If someone intends to judge that p v $\neg p$ and ends up judging that *p*, he can neither have done so by mistake nor as a result of doing something else that he accepted might lead to his so judging.[21] Finally, there is little to recommend the idea that judging might be a kind of subintentional action, caused by non-conscious wants* (Sect. 5.1.3) to make a judgement with that particular content. This would amount to the claim that judging is necessarily a

[21] In 8.6, I go through these arguments relative to deciding is more detail.

form of wishful thinking. No doubt some judgements are precisely that. However, the universal truth of the claim would render inquisitive wants* unsatisfiable and make the frequent cognitive success of our inquiries a mystery.

The applicability of the idioms of intentionality and control, whether positively or negatively predicated of our involvement in some event, is an indication of its agential character (cf. Thalberg 1972, pp. 51 ff.). These idioms, however, are inapplicable to specific judgings. This speaks strongly against their being actions.

8.5 Minimal Deliberation and Decision

Is the same to be said about the "making" of practical decisions? In the end, I think the answer is yes. Nevertheless, decisions are more complicated and in a sense more puzzling than judgements. The first thing to notice, however, is that much of what I have said about judgement's relation to minimal inquiry *can* be said about the relation of decision to minimal deliberation.

Parallel to the initiation of inquiry, deliberation is inaugurated as a result of someone registering what can be labelled an *optative uncertainty*, which she has a sufficiently strong concern to overcome. Where an epistemic uncertainty concerns the question of what *p* to insert in an attitudinal stand of the form "It is the case that *p*", optative uncertainty concerns the same question relative to a stand of the form "Let it be the case that *p*". It would be somewhat misleading to talk here of an optative "gap", as we often begin deliberation where we are already bearers of attitudes specifying the candidates for an answer. Deliberation is frequently inaugurated as a result of someone registering an *incompatibility* between wants* of which she is already the bearer. This will generally be because of certain changes in her beliefs, which bring the wants* into a contingent conflict. However, this need not be the case. Someone might, for example, want to work out where to go on holiday without having previously been the bearer of desires to go to two incompatible destinations. Conversely, there are also cases in which inquiry is inaugurated as a result of registering an epistemic incompatibility, that is, of coming to realise that one believes two propositions that cannot both be true. Again, this is the less usual case.

So far, the difference between the two forms of minimal reflection concerns only characteristic features, not necessary conditions. However, once we attempt to give a specification of the precise character of the concern that constitutively powers minimal deliberation, we hit on a significant disanalogy. As I argued above, it is an essential, internal feature of inquiry that it be powered by an inquisitive want*, that is, by a want* which specifies the conditions under which some belief content can count as an answer to the question that inaugurated that inquisitive episode. In contrast, *no such criterial specification* of the motivation internal to minimal deliberation can be given.

The essentially practical question that inaugurates deliberation can be formulated as "what shall I do?", "what am I to do?" (Taylor 1964, 74; Williams 1985, 18;

Watson 2003, 134) or, if one keeps in mind the subjective nature of the standard setting involved in taking on optative attitudes (Sect. 4.3.4), “what should I do?”. Because optative attitudinising is constitutively a matter of subjective standard setting, the attitude in which reflection is to issue provides, in contrast to the doxastic case, no formal object whose attainment is the constitutive aim of deliberation. There are quite simply no criteria that the content of an optative attitude has to be thought to fulfil in order to be able to count as an answer to the question “what shall I do?” It is sufficient that the attitude’s content be in some sense the minimal deliberator’s *own answer* to the question in the situation. Saying this does not involve naming anything that could be thought of as a criterion. Deliberative wants*, then, contain no criterial reference that can be given a formal rendering comparable to that I gave for inquisitive wants*. Rather, the content of a deliberative want* is simply that its bearer resolve – settle on *her* answer to – the optative uncertainty that has triggered it.

I shall return to this point in Section 8.7.2, where I propose an explication of the idea of an attitude’s content counting as the agent’s own answer. First, we need to look in a little detail at why two other kinds of strategies on offer should be rejected. The first attempts to give the demand that the deliberator’s answer be her *own* a reading that is *strong* enough to provide a substantive criterion to be met in deciding. The second strategy ties deciding criterially to *value or reasons judgement*.

8.5.1 Strong Ownership

Frankfurt

Harry Frankfurt has seen decision as essentially a matter of a person’s “identification” with the want* to do what he decides on. Through such identification, Frankfurt claims, the person “constitutes himself”, seeking to make himself into “an integrated whole” (Frankfurt 1987, 170). The idea is that by deciding to φ we aim to integrate the desire to φ into the system of those attitudes that are really ours, thus distinguishing them from those attitudes that happen to occupy our mind-body, but which we reject. According to the holistic conception of self at work here, someone decides to φ by optatively endorsing a “desire”, apparently on the basis of the belief that doing so will allow it to take up a place within a coherent system of attitudes.[22] This may however not mean that the system was already such as to have left an open space that the newly endorsed desire could occupy. Rather, so it seems, the adoption of the new desire may bring about a shift in the structure of the self. That would justify Frankfurt’s use of the emphatic term “self-constitution”.

[22] Whether such attempted integration is successful is, for Frankfurt, a further question, that of “wholeheartedness”. Where this turns out to be lacking, the person will have difficulties realising her decision.

This conception is, however, far too strong as an analysis of much of our everyday, low-grade deciding. Certainly, it seems true that someone deciding to take a ham, rather than a cheese sandwich or to watch soap *A* rather than soap *B* is unlikely to assume that doing so poses a threat to the coherence of her self. But it seems manifestly false that coherence – or identification understood in any other way – plays any kind of criterial role in these extremely local kinds of decision. Moreover, we can also decide – in the everyday sense of the term – to perform actions, in spite of seeing the motivation to perform them as "external" in the sense Frankfurt is attempting to reconstruct. An unwilling kleptomaniac can wonder which department store to rob, assess the pros and cons and then opt for one of the two without thereby coming to see her desire to go on a stealing spree once every 2 days as part of who she "really" is.

In fact, it is fairly clear that Frankfurt is not claiming that everything that falls under the everyday term "decision" is to be analysed in terms of "identification". In an earlier article, he had stated that identification was a matter of "a particular kind of decision" (Frankfurt 1976, 68). And in the article under consideration here, he distinguishes terminologically between "decisions", which he sees as about the agent's own attitudes, and "choices", which concern an agent's potential actions (Frankfurt 1987, 172).[23] Frankfurt had been looking for a criterion for the distinction between those optative attitudes for which their bearer actively opts and those she rejects, and had hoped that decisions of a certain kind might do the trick – a hope he later abandoned (Frankfurt 1992, 99ff.). It ought therefore to be clear that there could be little chance of fulfilling the converse hope of defining decision in terms of the strong idea of ownership brought about by "identification".

Korsgaard

For more or less the same reason, Christine Korsgaard's adaptation of Frankfurt's idea is not going to help either. Like Frankfurt, Korsgaard takes it that, prior to reflective action, persons need to "endorse" the form of motivation that leads them to act.[24] Now, talk of "reflective endorsement" can be understood purely formally, effectively telling us no more than that some action want* is given a privileged status by its bearer at the end of deliberation. And at a minimum that is surely what decision is. However, like Frankfurt, Korsgaard also takes it that the relevant form of endorsement can be explicated in terms of the concept of identification. Unlike Frankfurt, she thinks that the object with which a deliberative agent identifies

[23] For a different attempt to draw a substantial distinction between "decision" and "choice" see Meiland 1970, 61ff. Bratman employs "choice" terminologically in a way that severs any necessary connection with intention (Bratman 1987, 152ff.). I respond to this in Roughley 2007c.

[24] It is unclear whether Korsgaard believes that the "reflective consciousness" which characterises the personal life-form forces this structure on every human action, and not merely on those preceded by optative uncertainty.

is not the attitude in question, but a principle under which he takes it that the attitude adopted falls: “The reflective structure of human consciousness requires that you identify yourself with some law or principle which will govern your choices” (Korsgaard 1996, 103f.).

Korsgaard thinks that the only way an agent can see himself as an active deliberator and decider is by conducting deliberation and decision with a conception of his “practical identity” in mind. And a practical identity is a principled way of understanding oneself, that is, a way of thinking about oneself as deciding as a result of considerations that meet some kind of supervenience requirement: if certain features of situation s_1 are seen as calling for an action of type φ, then sufficiently similar features of situation s_2 must also be seen as calling for action of type φ. The relevant principles, that is, the arrays of conditional requirements to which the agent is committed, constitute his identity as a husband, a health inspector, a colleague of *X* and *Y*, a friend to *Z*, a gardener, a supporter of Chelsea Football Club, a conservative, a lover of opera, etc (cf. Korsgaard 1996, 101). To have these partial identities is precisely to see particular kinds of considerations as providing reasons for particular kinds of action: not to take certain features of one’s garden as requiring specific types of action would disqualify an agent from partaking of the partial practical identity of “gardener”.

Now, it is one thing to argue that particular features of our self-understanding impose demands that we see specific kinds of considerations as reasons for action when we practically deliberate. It is another to claim that all decisions have to either draw on or involve committing oneself to such supervenience requirements. But this, it seems, is what Korsgaard is claiming: “The claim to generality, to universality”, she says, “is essential to an act’s being an act of the will” (Korsgaard 1996, 232).

The natural response to this claim is that we simply don’t think of all of our decisions as principled. If, as I have argued, there such things as akratic decisions, it would seem bizarre if even they had to be taken on the basis of commitments to take whatever consideration guided the action as providing an equally strong reason for action in a sufficiently similar situation. The whole point of the characterisation of akratic actions as “weak” is to bring out the fact that they appear to their agent as unjustified even as she is deciding on them. To pick up again the arguments of Section 4.3.2, self-destructive or spiteful actions may be carried out without seeming to their agents to embody the realisation of any feature to whose realisation in similar situations they want to bind themselves.[25] Moreover, the low-grade decisions mentioned before to buy a ham sandwich or watch TV programme *A* surely don’t commit their bearer to going for the same option a week later when the same options present themselves and there are no significant changes in the choice situations.

[25] The point is argued by Jay Wallace at somewhat greater length (Wallace 2001, 88ff.).

Velleman

Finally, David Velleman's conception, the cognitivism of which I discussed in Section 6.3.1, can be seen as a further variant on the idea of strong ownership as the substantial criterion of decision. For Velleman, my being "decided upon" φ-ing (Velleman 1989, 112) involves having accepted a belief that I will φ because φ-ing is an action which, in the situation, it would be intelligible for me to perform.[26] He correspondingly takes the constitutive "motive that drives practical thought" to be "a desire to do what makes sense" (Velleman 1992b, 139ff.). That in virtue of which it makes sense to me to perform that action may be a feature of my present mood, for instance, my despair (Velleman 1992a, 121), or of my personality in general. Velleman thus cashes in the idea of the agent's *own* decision in terms of the intelligibility to the agent of what the agent opts – or, as Velleman would put it: expects – to do.

As I already claimed in Section 6.3.1, we have good reason to be sceptical about the pervasiveness and unremitting causal efficacy in agents of the want* to make sense to themselves. This is not only so for phenomenological reasons, but also for reasons internal to Velleman's theory. As Velleman explicates the idea of making sense in terms of explicability (Velleman 1992b, 141), a person's being thus motivated would be a matter of her wanting to know which mechanisms in her are responsible for her doing what she does. Again according to Velleman, it is this very desire for self-understanding that frequently contributes significantly to us acting the way we do, as its functioning can be equivalent to the agent herself "throwing her weight behind" a motive that would otherwise have been too weak to move an agent to act. But if an action's explanation frequently or even necessarily involves reference to the motivation to make sense of oneself, then the agent must, in the course of her deliberation, be aware of that motivation and its potential effects. However, Velleman is quite explicit that this motivation is "largely subliminal" and "troubles the agent's consciousness only when it is thwarted" (Velleman 1989, 38). This is an attempt to counter the phenonomenologically grounded scepticism as to the pervasiveness of any such desire. But conceding this makes deliberation, in as far as its object is to settle on an action intelligible to its bearer, hopelessly opaque. However strongly non-conscious wants* may influence our action, perception and feeling, I take it that deliberation is a paradigmatically conscious affair. Thus, even if we were to accept that the want* to settle on an action that one could explain might power deliberation "though not under that description" – but rather under the description of wanting to act out of reasons (Velleman 1992b, 141) – our deliberative thought processes would be seeking an action whose performance we could explain although we are unaware of the feature of action explanation that is supposed to be decisive. The unfortunate result is that, on Velleman's picture, we really don't know what we are doing when we practically deliberate.

[26] Velleman sees his picking out this mental move as an analysis, indeed as a "reduction", of Frankfurt's talk of "identification" (Velleman 1992b, 136).

Velleman offers his analysis as a corrective to the idea that the aim of practical reasoning might be characterised as answering the question “what shall I do?”. He quite rightly argues that the question thus formulated is too formal to provide anything like a criterion for deliberative success. However, he also believes that, if no such substantial criterion can be provided, we will not be able to make sense of the idea of practical deliberation. Because deliberation can be successful or fail, there must be some substantial criterion against which these can be measured, he thinks (Velleman 1996, 176f.).

I think, however, that it is far from clear that we do have one clear criterion of deliberative success. When we decide to perform a certain action, our decision will in one sense be “successful” if, as Frankfurt says, it turns out to be “wholehearted”, i.e. if it turns out that we sufficiently motivated to prepare it, to ward off other motivational sources and to realise it in the relevant situations. There are also, second, certainly cases in which we achieve deliberative success by choosing the option specified by our values or, more generally, our reasons. Sometimes, though, a decision can count as successful in a third sense if it enables us to extricate ourselves from a situation in which our reasons appear inconclusive or in which an interrogation of our reasons doesn’t seem worth the bother. The only plausible candidate for a general characterisation of deliberative success, however, seems to me to be the agent’s determination of “where she stands” on the relevant issue at that moment. This metaphor is minimally informative, doing barely more than to repeat the formulation above that the optative stand on the matter is, in a weak sense, the agent’s own. A specification of why that particular stand is to be taken is no part of the want* that necessarily powers deliberation.

8.5.2 Conclusive Reasons Judgements

I have been claiming that minimal deliberation, the movement of mind that aims to resolve a consciously occurrent optative uncertainty, is the preparatory behaviour required for the resultant attitudinal event to count as a decision. Deliberation that goes beyond the minimal – what we normally mean by the term – involves the consideration of reasons, norms or values with the aim of coming to a conclusion about what we have most reason to do or what would be the best course of action in the circumstances. If I am right, the question naturally arises as to the precise relationship between what I am calling “minimal deliberation” and the more extensive mental activity that is normally of interest when deliberation or practical reasoning is the topic.

In practical deliberation that goes beyond the minimal, agents consider reasons in order to reach a judgement about what they have most reason to do. I will call such judgements *conclusive reasons judgements*. If practical deliberation were essentially the attempt to form such judgements, we would have a criterion for deliberative success for intentions. The meeting of that standard would then be represented in the content of deliberative wants* and we would have a strict parallel with theoretical inquiry after all.

Certainly, an acceptable account of deliberation has to explain the close connection that seems obviously to obtain between decision and the formation of beliefs concerning our conclusive reasons. Articulating a worry that has exercised authors from Plato to Davidson, Gary Watson has argued that there is a serious problem with any attempt to prise intention and judgement apart conceptually, because a sentence such as "I never intend to do what I decide is best" would be incoherent (Watson 2003, 133). Before discussing the problem, it will be helpful to modify Watson's formulation in two ways: firstly, I will replace "decide", meant by Watson as an instance of cognitive "deciding that", with the unambiguous "judge". Secondly, I prefer to talk of "what I have conclusive reason to do" rather than "what is best", as the former generally encompasses the latter, but also leaves room for the consideration of deontic facts that may not be evaluatively reducible. Doing so moves us away from the decision theoretic and Platonic axiological diction I employed in Section 8.2. Because talk of "conclusive reasons" is naturally taken to be wider than talk of "what is best", I think this move strengthens the challenge. This gives us the sentence "I never intend to do what I judge I have conclusive reasons to do".

Watson's challenge is important. We are indeed owed an explanation of why such a global claim would be so bizarre. However, we can provide such an explanation without taking the problem to derive from a conceptual relation. For this reason, although it hardly seems conceivable that a human agent might truthfully make the adapted Watsonian claim, the claim is itself not incoherent. This is because there are various gaps between the two kinds of mental move.

In Section 8.2, we have already seen why it is false that deciding to φ entails judging that we have conclusive reasons to φ: decisions can be taken in the face of judgements that we have no conclusive reasons to take the option decided on, in the face of judgements that we don't know what reasons are conclusive and in the face of judgements that it's too much trouble to find out where the strongest relevant reasons lie. I am claiming that this is because deciding brings minimal deliberation to a close and the features that make deliberation more than minimal – primarily the consideration of reasons – are inessential for decision.

Nevertheless, there is clearly a close relationship between conclusive reasons judgment and decision. The connections are both empirical and rational. The latter, it seems, explain the former. Rationality requires a significant level of coherence between the contents of conclusive reasons judgements and decisions. Because we recognise that we are subject to some such demand, we tend to align our decisions with our reasons judgements. Moreover, because we are also aware that our decisions are generally better when they are taken on the basis of reasons judgements, we frequently engage in more than merely minimal deliberation. Agents' reasonable disposition to consider reasons and rational tendency to decide on the basis of considered reasons judgements are in turn assumed to be at work by other agents concerned to understand and interact with them. But how circumspect in thought and how considered in their decisions people are varies significantly. Unreasonable behaviour and rash decision cannot be the norm, but they obviously occur in individual cases.

What, though, is the precise nature of the coherence demanded by rationality between conclusive reasons judgements and decisions? There has been considerable discussion of John Broome's formulation of a rational requirement relating "ought" beliefs and intentions.[27] Although the requirement I am interested in here relates conclusive reasons judgements and decisions, the question that Broome and I are posing is more or less the same, but there are important differences.

I will come back to advantages of talk of conclusive reasons judgements as opposed to "ought" beliefs. I am interested in the relation of the former to decisions rather than to intentions because I am at the moment developing an account of decision, which will only provide the basis for my theory of intention in a further step. Because not all intentions are generated by deciding, Broome's requirement is at least in one respect wider than the requirement I am seeking. However, I think that a requirement linking conclusive reasons judgements and intentions is itself derived from a still broader requirement that concerns being disposed to realise one's conclusive reasons judgements. I will propose such a requirement in Section 9.4.2, when I discuss the non-decisional generation of intentions on the basis of conclusive reasons judgements.

Now, it is certainly not true that there is a rational requirement on an agent who forms a conclusive reasons judgment about her prospective behaviour to decide to realise the content of that judgement. There are two kinds of reason that block any such unmediated requirement. The first is that there are many cases in which an agent will realise the content of her judgement without needing to form an intention to do so (a). The second is that there can an unproblematic temporal gap between such a judgement and a corresponding intention (b).

(a) Cases of the first kind subdivide into two groups: automatic actions and non-actions. An agent for whom getting up at six in the morning has become second nature, may one day sit down to review her lifestyle and conclude that getting up at six in the morning, as she does, is an unbeatable way to start the day.[28] Obviously, she doesn't need to take any decision to continue as was. Had she developed serious doubts about the wisdom of continuing in this way and therefore entered into genuine practical deliberation in order to dissolve her optative uncertainty, we could accurately describe the termination of that uncertainty as a decision. But she may have other motivation for thinking through the reasons for her present lifestyle. Perhaps she has a policy of reviewing features of her life on a regular basis and of only considering practical consequences where she finds the reasons to be strongly stacked against her present way of doing things.

[27] Broome has repeatedly reformulated the requirement under slightly different labels (Broome 2005, 322; 2007a, 360f.; 2007b, 161). Since 2010, it runs under the title "enkrasia" (Broome 2010, 290; 2013a, 170; 2013b, 425). Its most recent formulation goes: "Necessarily, if *N* is within the domain of rationality, rationality requires of *N* that, if (1) *N* believes at *t* that she herself ought that *p*, and if (2) *N* believes at *t* that, if she herself were to intend that *p*, because of that, *p* would be so, and if (3) *N* believes at *t* that, if she herself were not to intend that *p*, because of that, *p* would not be so, then (4) *N* intends at *t* that *p*." (Broome 2013a, 425). See the special issue of *Organon F* dedicated to the discussion of Broome's enkratic requirement (Fink 2013).

[28] Compare the justification of the restriction of the scope of *SI* in Section 7.2.2.

Just as obviously, it would be absurd to require that agents make decisions not to do all the things they judge they have conclusive reasons not to do. That is because we generally don't need to decide not to perform actions in order not to perform them. Exceptions are cases in which someone is strongly tempted to perform an action disqualified by her conclusive reasons judgement. In such cases, the conjunction of the agent's motivational condition and her judgement may rationally require a corresponding decision. Otherwise, what the judgement rationally requires is that the agent refrain from deciding to φ. Rationality certainly requires of an agent that, if she judges that she has conclusive reasons not to φ, she doesn't decide to φ.

What about cases in which the content of the judgement is that one does not have conclusive reasons to perform some action? Scanlon has argued that this judgement equally requires avoiding a decision to φ (2007, 95) and Kolodny has made the stronger claim that a belief that one lacks sufficient reason to φ imposes the requirement not to intend to φ (Kolodny 2005, 521). Doubts about these claims are raised by Buridan cases, in which an agent believes she has no better reasons for doing anything other than φ-ing or ψ-ing, but has no conclusive reason for either φ-ing or ψ-ing (Sect. 8.2.2). In spite of the lack of a conclusive reason for performing a particular one of the two actions, it is clearly rational for the agent to form the intention to do one thing or the other. It can therefore not be irrational for her to decide to φ in spite of judging that she has no conclusive reason to φ. A defining feature of these kinds of cases is that the agent is forced to act, whatever progress he may have been able to make in determining the reasons he has. If there are incommensurable values, as perhaps in Sartre's mother-versus-résistance example, there may be cases in which, for non-contingent reasons, an agent needs to decide in spite of judging that he doesn't have conclusive reasons either way. A requirement not to decide where no conclusive reasons are forthcoming would condemn these agents to irrationality. That would be highly implausible.

Judgements concerning the lack of conclusive reasons can safely be put to one side, but a general principle connecting decisions with judgements that there are conclusive reasons will need to account for the rationally unobjectionable character of non-decision where the agent's practical dispositions guarantee behavioural conformity with the judgement.

(b) Alongside these, there is a further set of cases in which rationality requires no corresponding decision because there may be a period of time after a judgement during which a lack of decision may not only be unproblematic, but may even be eminently reasonable. Take a person who one autumn is wondering where to go on holiday next summer. He judges that, all things considered, the balance of reasons clearly speaks for his going to some place *Q*. Still, he has plenty of time so he doesn't take the step of deciding. There seems to be nothing irrational about that (cf. Scanlon 2007, 95).

The reason why this seems perfectly sensible is the defeasibility of reasoning in practical matters: the addition of further considerations can invalidate a conclusion about what one has most reason to do in a manner that has no equivalent in deductive theoretical reasoning (cf. Geach 1966, 77). Note, however, that the rationally unobjectionable character of an agent's not taking the step to decision need not depend on the agent's belief that further relevant considerations might crop up. He might simply have turned his mind to other matters, but be disposed to return to the question in good time and, if no other relevant considerations have come into view, to take the corresponding decision.

The possibility of decision-independent behavioural dispositions and the defeasibility of practical reasoning can both be catered for by means of a doxastic premise concerning the dependency of the judgement's realisation on a decision taken by a certain point in time.[29] Comparable doxastic premises with a temporally indexed content appeared in all the *IC* requirements that concern intending, as opposed to the eschewal of further intentions (Sects. 7.2.2 and 7.2.3). The requirement should, I think, be given the following formulation:

(CRD)
For person *X* and behavioural option φ of *X*,
(where "φ" picks out a non-disjunctive option):
it is rationally required of *X* that
if *X* judges that she has conclusive reasons to φ,
X believes that she will not φ unless by *t* she decides to φ
and *X* doesn't believe that she won't φ,
X at a time she takes to be no later than *t* decides to φ.

CRD – the requirement relating conclusive reasons judgements and decisions – solves both kinds of problem that block an unmediated requirement to intend what you judge you have conclusive reasons to do. The introductory clause obviously needs some explanation, which I will provide in a moment.

A shorter word is in order first about the third, negative doxastic premise. Unlike intentions, conclusive reasons judgements don't carry with them any doxastic condition. For this reason, *CRD*, unlike *SI*, requires a doxastic premise that specifies that the content of the relevant judgement does not look to the agent to be practically irrelevant. The premise is thus identical to the negative doxastic condition on decisional intending (Sect. 6.3.2). In contrast to Broome's premise (2) (cf. note 27 above), the condition is negative, as the presence of a positive belief is simply unnecessary in order to avoid the obvious irrationality of deciding to φ whilst believing that one won't φ even if one decides to φ.[30]

Now to the introductory clause, the necessity of which has to do with *CRD*'s solution to the problem of judgements which look like they will be automatically realised (a). *CRD* deals with these by excluding them from the purview of the requirement by means of the second doxastic premise. Cases in which the judgement concerns omissions are also dealt with unproblematically by the same means. At least this is so if we assume that φ can be satisfied by ¬ψ. The requirement then covers omissions that the agent believes she won't realise unless she decides not to

[29] Miranda del Corral has argued that 'ought' judgements with negative contents and 'ought' judgements made significantly earlier than the action they concern should lead us to conceive of the enkratic requirement as a mere prohibition (del Corral 2013, 577ff.). As I go on to argue, a positive requirement of the form of *CRD* can deal with these cases.

[30] Compare my criticism of the positive doxastic condition in Bratman's principle of intention persistence (Sect. 7.2.3).

perform the relevant action. Absent any such belief, which will presumably draw on the agent's beliefs about her relevant motivation or habitual action tendencies, no decision is specified by the requirement.

This, however, is where the problem arises that necessitates the somewhat strange introductory clause. The problem concerns the forms of complexity that should be taken to satisfy φ. In particular, if one allows for disjunctive content, *CRD* may be in trouble. If an agent judges she has conclusive reasons either to ψ or to χ, but doesn't know which of the two possible actions is conclusively favoured and also fears that the non-favoured action would be disastrous, it may be rational for her to intend a third action θ, which she takes to be a safer option. Ralph Wedgwood responds to this worry by excluding cases of "relevant uncertainty" – here the uncertainty as to which of the disjuncts it is one has conclusive reasons to realise (Wedgwood 2007, 30f.). This is, I think, correct. For our purposes, it is helpful to see why.

Intending disjunctive contents seems to be a much less widespread phenomenon than forming judgements with disjunctive contents. This is because the *IC* requirements can make intending disjunctive contents an exceedingly demanding matter. It is rationally required of an intender that she take on a whole set of attitudes and practical dispositions. If what is intended is disjunctive, that can get very complicated so that it may be unclear to the agent how she should go about complying with the demands. But someone might sit back and think through a whole set of different scenarios and come to the conclusion that she certainly has conclusive reasons to realise one of them, without having any thoughts about an ordering of the options thus picked out. She might do so in order to exclude an option that is not in this set. There need be no uncertainty, in as far as she needn't form a judgement that she doesn't know which of the options is best supported by reasons. That wasn't her question. Nevertheless, she forms a conclusive reasons judgement at this juncture, one to which she may or may not return. The possibility of conclusive reasons judgements formed at a stage in the proceedings at which the agent doesn't want or need to decide on a course of action is central to the second kind of problem solved by *CRD* and to which I will come in a moment. At any rate, the specification that "θ" does not stand for disjunctive contents avoids this problem. The exclusion of disjunctive contents but the permission of negated contents may seem problematic. However, as there is a rationale to each of these specifications, I don't think they undermine *CRD*'s plausibility.

Return now to the second problem with the requirement of an unmediated transition from judgement to decision (b): *CRD* allows for defeasibility in practical reasoning by assigning the decision a temporal index and tying it to the belief that a decision with that content needs to be taken by the specified point in time, if the judgement's content is to be realised.[31] It has been denied that rationality allows any such temporal gap. Wedgwood, for instance, claims that, if I judge in year *n* that I

[31] As with the *IC* requirements that involve an 'at *t* or never' belief (*EC* and *SI*), *CRD* should be supplemented by parallel requirement whose applicability is conditional on the now-or-never point in time taken to be given in the conclusion of *CRD* being, relative to *X*, obviously reached (cf. Sect. 7.2.2).

ought in year $n + 1$ to file my tax returns before a certain date, it would be irrational for me not to already intend to do so. According to Wedgwood, the impression one might have that the intention is not yet required derives from the conflation of intending and taking practical measures to realise an intention. Such measures, of course, need not be taken in advance (Wedgwood 2007, 29).

Two points should be made in reply. The first concerns the status of the specific example. There may be cases in which there are clearly fixed necessary actions relative to agents' conceptions of the good. In such cases, in which what is desirable is clear long in advance and only likely to change should a social revolution occur, it may be plausible that there is simply no rationale for not immediately deciding to realise the content of one's judgement. But it doesn't follow that all cases are structured in this way. Indeed, they are not.

Secondly, perhaps Wedgwood's example should not be considered a counterexample to *CRD* at all, as the judgement with which he, like Broome, takes coherence to be required is not a reasons judgement, but an 'ought' judgement. Obviously, the extent to which this makes a difference depends on how one understands the relationship between ought and reasons. This is not the place to discuss Broome's and Wedgwood's understanding of that relation.[32] Instead, I want to point out a use of the expression "conclusive" according to which *CRD* may indeed look doubtful for the kind of reasons worrying Wedgwood.

It might be suggested that conclusiveness should be understood as entailing the supplementary claims that there are now no further factors whose consideration might conceivably justify revising the judgement and that no such further factors will crop up prior to the latest point at which the judgement could be realised. A conclusive reasons judgement would then involve forming two beliefs: one belief concerning the strength or overridingness relations between the reasons one has considered and a further belief about one's epistemic situation relative to relevant reasons. If we were to give talk of conclusive reasons judgements this strong, two-level reading, then we could delete the reference to a specific time in the doxastic premise and in the conclusion. Someone might also formulate the first premise in terms of an 'ought' judgement, assuming 'ought' judgements to entail such a strong epistemic condition.[33] Whatever terminological variant is preferred, such a strong reading of the first premise would allow us to formulate a requirement of rationality without any temporal indexing.

Such a requirement would, however, be relatively uninteresting, because the first premise is so rarely fulfilled. On the weaker reading, it is fulfilled much more

[32] See, for example, Wedgwood 2007, 126ff.; Broome 2013a, 46ff.

[33] According to Broome, "the central ought is the all-things-considered one" (Broome 2013a, 26). The question is whether judgements working with this concept of "ought" are simply made in the light of all the things that have been considered or, additionally, in the light of the view that the things one has considered are all that need, or will need to be, considered.

frequently. If we want to cover cases governed by the weaker reading, then we need the temporal index in the doxastic premise and in the conclusion.[34]

It is worth mentioning a further feature of the relationship between 'ought' beliefs and conclusive reasons judgements which speaks for the use of the latter rather than the former concept in the first premise of the requirement. John Brunero has argued, I think convincingly, that someone might believe he ought to φ, whilst believing that there isn't sufficient evidence for that belief. In such a case of an irrational 'ought' belief, it doesn't seem that the agent is required to decide correspondingly (Brunero 2013, 553ff.). In contrast, it is inconceivable that a conclusive reasons judgement could be made in the face of the belief that one has insufficient evidence for that belief. As evidential considerations are reasons, a judgement which fails to take the agent's beliefs concerning evidence into account simply cannot be a conclusive reasons judgement.

I conclude that, where an agent makes conclusive reasons judgements, rational requirements on her decisions come into play. For this reason, there can be no question of our decisions becoming "unhinged from reasons in a general way", as Watson worries that they would be if deciding is not "internally" bound up with evaluation. There are clear connections. They are, however, rational, not conceptual in character.

The rational necessity of decision in the light of certain conclusive reasons judgements establishes a tight tie between minimal and more-than-minimal deliberation. The slack, however, is loosened by the fact that decisions don't require conclusive reasons judgements in order to be rationally in order. We saw this in the discussion of the cases described in Section 8.2.

Tie-breaking decisions in Buridan cases are rationally required, in spite of the decision maker neither having, nor judging that she has, conclusive reasons to go for the specific option she decides on. Moreover, a decision may be rationally unproblematic where an agent judges that the matter doesn't justify her interrogating the pros and cons in any detail before deciding. In such cases, an agent may decide on the basis of what one might call a *quick conclusive reasons judgement*, that is, a judgement that the reasons quickly considered speak most strongly for φ-ing, supplemented by the awareness that further consideration would not improbably lead to the judgement's revision.

Finally, we also seem to make some decisions after what one might call *pure minimal deliberation*. Just as we are able to bring optative uncertainty to a close where reflection on reasons has proven inconclusive, it appears that we are also able to do so without engaging in any weighing of reasons in the first place. This seems to

[34] There is one way in which a requirement that does without any temporal indexing might appear preferable. This concerns gradual dimensions of rationality. As pointed out in Section 7.2.2, postponing a decision may risk reducing the efficacy or quality of the action that one takes in order to realise one's reasons judgement. Someone who satisfies *CRD* may thus still be irrational in a weaker sense: although she does realise the content of her judgement, she might do so in a way that is, by her own lights, suboptimal relative to how she would have done had she decided immediately.

be precisely what happens where people decide to take some course of action after only a moment's hesitation. It looks as though we can go spontaneously for some option without having made any comparison with competing courses of action. It might be objected that in such cases we always implicitly compare φ-ing with not φ-ing. But this talk of "implicitness" smacks of an a priori theory that rides rough-shod over the phenomenology. In the cases I have in mind, someone has a general idea of something they want to do – say eat a sandwich – and is confronted with a possible course of action – say eat a ham sandwich – which they, after hesitating for just a moment, simply plump for. There is no weighing up of the plus points attaching to a ham sandwich against those attaching to eating something else instead. Rather, the agent waits a moment to see whether the idea catches his fancy and if so, goes for it. In such a case no thoughts are spared to what may be gained from refraining. In contrast to cases of quick conclusive reasons judgements, these latter cases don't involve anything that might appropriately be described as reasoning. Nevertheless, they will often be rationally irreproachable.

8.5.3 The Basic Structure of Practical Thought

My main claim in Section 8.5 is that minimal deliberation – raising the optative question as to what to do with respect to some situation, and then moving to take a decisive optative stand on the matter – constitutes the basic structure of practical thought. It does so in two senses of the term "basic". Firstly, it is so *rudimentary* that it need not conform to any rational requirements other than that the content of the decisive optative attitude be able to count as an answer to the question that initiated deliberation. Secondly, it is the structure that is *presupposed* by all practical deliberation that exceeds the minimal. Put pictorially, it is the attitudinal stem onto which practical reasoning is naturally and rationally grafted.

Certainly, what is frequently called "practical reasoning", i.e. reflection on what there are, were or would have been conclusive reasons to do in some situation, can be detached from the attempt to answer a – generally first-person[35] – optative question. Such hypothetical reflection, as when someone attempts to clarify how it would have been best for European political leaders to have acted in the 1930s, is only "practical" in a secondary sense: its sense depends on the intelligibility of the first-person optative question-and-answer sequence for some real or hypothetical agent in the situation under investigation.[36]

[35] The optative question can also be asked derivatively in the second or third person, for instance, where one is giving advice.

[36] Because reflection on what we have, had or would have (conclusive) reasons to do seeks a truth-apt answer, Harman has suggested it be classified as "theoretical reasoning" (cf. Harman 1975/76, 431; 1986a, 77f.). My less drastic proposal is that we think of such deliberation as practical where it is grafted onto an attempt to answer the optative question. Where this is not the case, reflection on someone's reasons to act will be a "purely theoretical" exercise.

A full sequence of practical deliberation is initiated by optative uncertainty about some possible action or omission and ends in an optative stand on the matter after passing through a phase of considering reasons and counter-reasons. In cases of what one might think of as pure practical deliberation, such consideration leads to a conclusive reasons judgement, the content of which is made the content of a corresponding optative stand.

In contrast to minimal inquiry, then, minimal deliberation is not constitutively motivated by a want* whose content specifies criteria for answering the deliberation-inaugurative question. It follows that the relation designated by the expression “on the basis of” differs in the two cases where a judgement is made on the basis of inquiry and where a decision is taken on the basis of deliberation. The consideration of reasons and evidence that takes place in the two forms of reflection stands in a different logical relation to the final product in each case. Inquiry is a person’s attempt to bring features of reality or true propositions before his mind in such a way as to *cause* in himself the formation of the belief that answers the inquiry’s inaugurative question. The relation between the consideration of practical reasons and the resolution of optative uncertainty can, in contrast, not be thought of in purely causal terms. Rather, the optative attitude with which the reflective process concludes possesses a level of autonomy that is lacking in the case of judgement.

8.6 Why Decisions Are Not Actions

On the one hand, then, decisions are prepared for by deliberation, as judgements are prepared for by inquiry. On the other hand, the evidence is that decisions are causally, and conceptually detachable from their preparation in a way that judgements are not.[37] This result is ambivalent for the theorist who wishes to claim decisions as actions. In contrast to judgements, their detachability from their active preparation confers on them a level of autonomy that prevents them being adequately conceptualised as mere attitudinal happenings caused by the foregoing process of reflection. Moreover, this autonomy seems to tie in with the active phenomenology that is experienced in certain cases of deciding, particularly where those decisions are deadlock-terminative or countermotivational. However, the apparent causal detachability of decision from its attitudinal preconditions endows it with a status that is action-theoretically anything but clear. If decisions are not caused by the prior deliberative activity, what are they caused by?

Recall the argument in Section 8.4 that particular judgements cannot simply be explained by wants* whose contents specify the making of those particular judgements. That argument focussed on the fact that judgements with singular propositional contents cannot be caused by inquisitive wants*, the contents of

[37] This is the kernel of truth in Sartre’s radical rejection of the importance of deliberation for decision.

which are necessarily disjunctive. The argument may appear to be inapplicable in relation to decision if, as I have been arguing, no such determinate content can be provided for deliberative wants*. However, that appearance would be deceptive, as the disjunctive content of inquisitive wants* is merely the specification of a more general kind of content that is common to both inquisitive and deliberative motivation. Both forms of reflection are necessarily inaugurated and sustained by *the want* to resolve uncertainty*. If there is no uncertainty, then there can be no genuine reflection of either kind. As uncertainty precludes someone already being clear on what will resolve it, the claim that a person's deciding to φ is generally caused by a prior optative attitude with the content that she decide to φ would be as false as the claim that a judgement that *p* is generally caused by a want* to judge that *p*.

If this is correct, it confronts us with an alternative that is somewhat uncomfortable within the framework I have been developing: either we abandon the causal theory of action, removing the necessity of being able to name an attitudinal occurrence that is causally sufficient for decision, or else we reject the claim that decisions are a species of action. However, it seems that the alternative might be avoided if it could be shown that an optative explanation of specific decisions can be provided after all.

Al Mele has advanced two arguments that might help us to see how this could be done. It is at least conceivable, Mele claims, that someone can decide to φ because he intends to decide to φ. His main argument for this claim grounds in the rejection of what has often appeared to be a natural assumption (Frankfurt 1987, 172; Kane 1996, 138f.): that intending to decide to φ entails having already decided to φ or at least some form of intending to φ. Mele (2000, 90f.; 2003a, 203f.) argues by example against this assumption. It is conceivable, he claims, that someone might have been neurologically manipulated in such a way as to be incapable of deciding either to φ or to ψ, but wired up to two buttons, pressing one of which will remove his incapacity to decide to φ, while pressing the other will remove his incapacity to decide to ψ. Such a person could, according to Mele, decide – and thus intend – to decide to φ without having yet decided to φ.

Mele's story sounds coherent because of the assumption that the inability to decide to φ does not automatically involve the inability to decide to decide to φ. But how might it be shown that that assumption is justified? What Mele's protagonist can clearly decide to do is to remove his inability to decide to φ. But that does not amount to deciding to decide to φ. Whether he can do that depends on whether there is an implication from deciding to decide to φ to deciding to φ. If there is such an implication, then we simply don't understand what is being said when it is claimed that the two can be separated. If that is the case, then we presumably make sense of Mele's story by tacitly assimilating talk of deciding to decide to φ to the idea of deciding to remove one's inability to decide to φ.

Compare an example from the field of inquiry. An inquirer might be uncertain as to whether *p* or ¬*p* is the case and also believe that he will judge that *p* is the case, when he gets round to thinking the matter through. The latter belief does not entail that he has already judged that *p*. The idea that the person still has some inquiry to do intervenes between the two. For him to be able to engage in genuine inquiry, there has to be some level of uncertainty in his mind as to what the outcome will be. Should there be no uncertainty in his mind as to whether p or $\neg p$ is the case, he will not be able to conduct genuine inquiry, but will be just going through the motions. The situation is then analogous to a rigged trial. But a belief as to where genuine inquiry is going to lead – presumably based on evidence, although perhaps there are a couple of details that still need checking – need not remove all uncertainty. In fact, in order that inquiry can take place, it *cannot* do so. Thus, the belief about the prospective outcome of the inquiry cannot explain the inquiry's outcome. Were it to do so, there would be no inquiry to be embarked on.

The (minimally) deliberative case is structurally the same: either the agent is, prior to putative deliberation, the bearer of some attitude that already resolves any practical uncertainty, or he isn't. If he is, then there is no practical uncertainty to be resolved; if he isn't, then whatever state he is in prior to (minimal) deliberation cannot be primarily responsible for the deliberation's outcome. As long as someone's intending to decide to φ leaves room for her uncertainty as to whether to φ or not, it does not entail that she has decided to φ. But precisely that uncertainty which it is necessarily the job of decision to resolve prevents the intention to decide to φ from explaining the decision.[38]

Certain sorts of everyday examples that may at first glance appear to speak against this conceptual claim actually turn out to support it. For instance, it may be thought that a young man who longs to ask a young lady to dance and is too shy to do so might therefore decide to decide to ask her after he has had a drink or two – without yet having decided to ask her.[39] As far as I can see, there are two kinds of sequence that could be misleadingly characterised by this description. In the first, the decision that the agent still has to take has as its content the proximal performance of the action, for instance, "now" or "when she returns from the bar". Here, the decision to ask her may have been taken, but taken in the belief that the realisation of its content will only be possible after the realisation of facilitating conditions. There is no reason why the belief that those facilitating conditions have not yet been brought about should prevent an agent from deciding to make the request[40]: countermotivational decision is possible. In the second version, the agent wants*, perhaps hopes or yearns, to make the request, but has not yet decided to do so. What he decides to do is have a drink or two in order to be able to fulfil his want*, hoping that alcohol-induced courage will enable him to decide to ask her. As in Mele's thought experiment, the agent in this version decides to take action that he believes will facilitate his making a decision. In the first version, the agent has attained optative certainty, in the second case, he remains uncertain.

[38] As Watson says, "Deliberation ... cannot target its specific terminus" (Watson 2003, 141).

[39] The example is Christian Piller's (Piller 2001, 210).

[40] Compare the example of someone intending to have a conversation in a language they don't yet speak in Section 6.3.2.

We should, then, firstly distinguish a bare decision to φ from decisions concerning the time at which that decision is to be realised, as from decisions concerning ways or means of realising it. Secondly, we should distinguish the decision to φ from decisions that facilitate deciding to φ – decisions that concern preconditions of the decision. Moreover, we should also be careful to distinguish deciding to φ from other φ-facilitating mental moves, for instance, attempting to muster motivational force in order to φ, where one has already decided to go φ. Such reflexive mental moves, often expressed in self-commands like "Come on!" or "Pull yourself together!" can obviously be made because one intends to make them. Again, they presuppose not uncertainty, but clarity on what one intends to do.

To return to Mele's example: even if we were to take it that the unusual intention produced by the unfortunate victim of manipulation were to be able to do the explanatory work that Mele thinks it can do, it would be of no help in the attempt to accommodate an action of deciding within a causal action theory. This is simply because, conceivable or not, such bizarre aetiologies are blatantly not the way we normally come to decide. And what the causal action theorist needs is an aetiological pattern that is at least generally given when we make decisions.

Mele's second argument (2000, 92f.; 2003a, 25) is designed to show that this could be the case. Here his strategy consists in drawing attention to cases in which occurrences that clearly qualify as actions do so without fulfilling the strict requirement that a representation of their content have cropped up in a prior optative attitude. This is, as he points out, the case where individual action parts together make up a complex sequential action, such as playing a piece on the piano. Mele then suggests that there may be a parallel in the case of deciding such that "an intention to decide what to do" can play a significant causal role in bringing about our decisions.

Unfortunately, Mele says nothing more about what the parallel might consist in or about the precise causal role we are to envisage. In as far as there is a genuine parallel here with complex action that might enable us to see specific decisions as attitudinally caused mental actions, that parallel can, as far as I can see, only concern the conferral of intentionality, and thus action status, on the pieces of behaviour not represented in the content of some intention. In the same way that the action parts derive their intentionality, and a fortiori their actional character, from the intention to perform the action whole, the decision to φ might seem to derive its intentionality from the intention to decide whether to φ or to ψ.

The analogy, however, would be strained beyond breaking point. It is at least a necessary condition of action parts deriving their intentionality from the intention to perform the entire complex action that the agent be the bearer of a – perhaps dispositional, but consciously accessible – belief that the latter consists of the former. This is supported by the fact that there are countless events in our bodies that inevitably occur when we act, but which are unintentional so long as we are unaware of their occurrence.

When we turn to the making of decisions with specific contents, it ought to be clear that no one has the belief that deciding to φ is a necessary constituent of deciding whether to φ or to ¬φ. The relation here is the converse: in order to decide to φ, we must decide whether or not to φ. But deciding whether or not to φ is not necessarily deciding to φ. Compare pulling a numbered ball out a bag that you know contains balls numbered 1 to 20. Taking out ball number 7 is not something you do intentionally because you pull out *some* ball as a result of intending to do so. This of course differs from taking all the balls out of the bag, which necessarily involves taking out number 7. Here, if you do the former intentionally, you also do the latter intentionally.

The better analogy might appear not to be with the parts of complex actions, but with acting in a way that brings about certain consequences that were not specifically aimed at, but whose possibility was taken into account in performing the action. Specific decisions would then be analogous to the production of foreseen "side effects". But this is also phenomenally inadequate. One could imagine a case of deciding that may seem analogous up to a point. Indira, who is chronically indecisive, has been told by her therapist to make five decisions every day. She therefore goes around deciding for the sake of deciding. What she ends up deciding to do is, in a sense only a by-product of her basic project of deciding regularly. Nevertheless, the attitudinal results of Indira's individual decisions cannot count as side-effects of those decisions. Decisions are not events that are knowingly brought about, but without being aimed at. Rather, deciding involves constituting what one decides on as an aim by deciding on it. One has the upshot of one's decision in focus in the moment of deciding. So there is nothing collateral about deciding. Whether accepted collateral consequences should be subsumed under the term "intentional", "unintentional" or "non-intentional",[41] decisions just don't fit the bill. And as there is no prior attitude anywhere along the line from which they could plausibly derive intentionality, specific decisions are, like judgements, just not candidates for the ascription of intentionality.

Finally, as Mele points out himself (2000, 102f.), if deciding were to be an action, it would have to be a *basic* action, that is, an action that is not performed by performing some other action. Decisions are neither causally, conventionally nor otherwise factually complex. As such, were they to be actions, it would have to be at least conceivable that they have *non-derivative intentionality*. The intentionality of raising my arm may derive from the intentionality of signalling my desire to say something. But arm raising could not be a basic action were we not able simply to raise our arm intentionally. Similarly, even if it were to be thought that the decision to φ could derive its intentionality from the intentionality of deciding either to φ or to ¬φ, it would still have to be possible intentionally to simply decide to φ. Again, this possibility is missing.

[41] For an account of these various terminological preferences, see Mele 1992b, 200ff.

8.7 Decisions as Deliberation-Terminative Optative Occurrences

Decisions are puzzling phenomena. On the one hand, they are forms of mental behaviour that often feel like basic actions. Indeed, they are frequently characterised by a peculiarly emphatic active phenomenology, a feeling of activity that actually goes well beyond that experienced during most of our actions, where actions are forms of behaviour which are "due to us" in the sense of being appropriately caused by our being motivated to perform them. As such, decisions may appear to form their own class of "hyper-actions". On the other hand, that active character does not entail their intentionality: the question as to whether a decision was intentional or not is misplaced and the attempt to explain specific decisions in terms of wants* to make those decisions is necessarily unsuccessful.

Short of abandoning a causal theory of action,[42] there seems to me to be only one solution. This is that decisions are *a species of consciously occurrent wants**, whose occurrence is sometimes, as it happens, characterised by a specific active phenomenology. This proposal, which may at first sound more like a reformulation of the puzzle than its solution, raises two questions, the discussion of which, I think, does lead to a plausible conception. One concerns the specification of the particular species; the other concerns the status of the active phenomenology.

8.7.1 Activity

To take the second point first: to talk of "active phenomenology" is to claim that the agent's acquisition of the attitude *seems* to be in some sense active. The question this naturally raises is whether things are indeed as they seem, i.e. whether the phenomenology is veridical or not. Either way, we have good reasons not to conceptualise decisional intention acquisition as an action.

On the one hand, if the phenomenology should turn out to be illusory, then a decision's aetiology presumably involves causation by some non-attitudinal occurrence in its bearer's body. In cases of deadlock-terminative and counter-motivational decision, there is, by hypothesis, no attitudinal factor to which causal responsibility could be assigned. In other words, if the phenomenology is illusory, then at least certain kinds of decisions are happenings involving us during which we become radically opaque to ourselves.[43] The rational conclusion here would be that these are events we would be well advised to avoid as far as possible.

[42] This is, of course, very much a live option again (cf. Dancy 2000; Bittner 2001; Schueler 2003; Alvarez 2007; Stoutland 2007).

[43] This is the sort of solution advocated by Ullmann-Margalit and Morgenbesser (1977, 773f.).

However, although people do sometimes try to avoid making decisions, this tends to be because they are averse to the responsibility and effort involved, not because they believe that, in deciding, they surrender the responsibility for their deeds to non-attitudinal features of their physiology. By and large, we tend to prefer deciding what to do to what generally appears to be the alternative – allowing causal chains that somehow involve us less to be "decisive" for what we end up doing. Apparently, we are convinced that there is *something* veridical about deciding's active phenomenology.[44]

Indeed, where decisions feel most strongly active, they appear simply to emanate from *the person herself* without further attitudinal mediation. For this appearance to correspond to reality, there would have to be such a thing as originary and direct causation of optative attitudes by their bearers. This would have the drastic metaphysical consequence of establishing a cleavage in causation of the kind envisaged by agent causationism. Note, though, that where agent causation was originally, in the work of Chisholm and Taylor, conceived as an explanatory model for all human action,[45] what we would have here would be a limited phenomenon restricted to the production of optative attitudes under certain, highly restricted conditions.[46] Therefore, even if the active phenomenology of deciding could be shown to be veridical, that would not provide particularly persuasive grounds for conceptualising decisions as actions. And even if the metaphysical difficulties of vindicating agent causality were to be solvable, the conception would still not plausibly be applicable to common-or-garden actions, but only to a particular sub-set of those mental occurrences we call decisions. That, however, would not align decisions with standard examples of action, but would rather be a reason for assigning them a class of their own.

[44] Gary Watson has suggested that the activity he takes to be at work in decision is a matter of "sensitivity to reasons" (Watson 2003, 125). However, whether such a conception is meant to identify the *explanans* of the active phenomenology or whether it is supposed to provide an analysis of *what is meant* by talk of activity itself, a sensitivity-to-reasons conception doesn't draw the lines in the right place. The latter interpretation would tend to classify as active all behaviour that is understandable as a response mediated by a reliable reasons-responding mechanism. This would include habitual and subintentional actions as well as many emotional reactions. Were, on the other hand, a sensitivity-to-reasons conception to be proposed as an explanation of the active phenomenology that sometimes characterizes decision, it would be simply false. Even if countermotivational decisions both result from sensitivity to reasons and feel particularly active, deadlock-terminative and counterevaluative decisions, which may also be accompanied by active phenomenology, cannot be explained by the agent's sensitivity to reasons.

[45] According to Chisholm, "at least one of the events that is involved in *any act* is caused, not by any other event, but by the agent" (Chisholm 1966, 29; my emphasis).

[46] Timothy O'Connor conceives agent causation as necessarily executed via "decisions". However, the concept of decision he employs is very different to that reconstructed in this text. For O'Connor, a decision is a "primitive action", consisting in the agent's causing an "immediately executive or action-triggering state". Decisions thus construed seem, on the one hand, not to require any step back from the flow of action, yet on the other hand to necessarily bring about their content immediately on being formed. See O'Connor 1995, 181, 198 (note 15), 200 (note 36) and 2002, 348. This is not our everyday concept.

What should be emphasised here is that discussion of the veridical status of the undoubted active phenomenology of at least some of our decisions is a *metaphysical* issue. But whether or not that phenomenology is veridical is no part of the *concept* of deciding. In fact, neither is the phenomenology itself. There are plenty of cases of deciding in which it is conspicuous by its absence. Faced with the momentary optative problem of whether to have a ham or a cheese sandwich, Sam hesitates momentarily before plumping for the ham. Need he experience himself as particularly active in the process? I very much doubt it. Indeed, such cases are obvious candidates for decisions explicable in terms of changes in the motivational strength of the relevant wants* effected through minimal deliberation. One look at the ham and the evoked memory of his last tasty ham sandwich immediately sways Sam – and is thus sufficient to explain his decision.

Still, if there are such phenomena as countermotivational and deadlock-terminative deciding, then, at least in some cases, our decisions are not the effects of the motivational strength of the prior wants* involved. It is no doubt these cases that are primarily marked by a particular sense of activity on the part of the agent. This sense of activity suggests a basic capacity to revoke any tendency to act that may result from the causal interaction of previously given wants*.[47] It seems likely, then, that it is largely the same cases that are characterised by both salient *active phenomenology* and serious explanatory difficulties. Our sense of activity consists, at least in part, in the assumption that we could have "vetoed" the workings of what otherwise appear to be perfectly good explanatory factors in us.

Nevertheless, action in the face of considerations that rationally lead to motivational deadlock is not always experienced as actively instigated. Think again of actions such as picking one orange from those lying in a fruit bowl or choosing one of a set of cans of beans on a supermarket shelf that are "to all intents and purposes" identical (Sect. 8.2.2). The act of choosing one particular rather than another is, in such situations, often experienced as completely effortless, if indeed it leaves any particular experiential traces at all.

Note, though, that many such cases of *apparently unmotivated picking* – in which the agent no doubt wants to take *some* particular, but appears bereft of an attitudinal source of the motivation to take *that* one – are not plausibly construed as the products of decisions. This is because the agent is not afflicted by any prior doubt that requires resolution by further optative movement. Rather, the complete lack of any active phenomenology in such cases plausibly correlates with the fact that what tips the optative scales is some sub-attitudinal factor such as the tendency

[47] B. Libet (1985, 536ff.) talks of the possibility of the agent himself "vetoing" the action tendency inherent in electrically measurable prior "readiness potentials". These are presumably correlates of motivationally qualified optative attitudes – a plausible assumption in spite of Libet's unhelpful talk of "*the brain* 'decid[ing]' . . . before there is any reportable subjective awareness that such a decision has taken place" (my emphasis).

to take items on the left, or items that are more shiny, etc, and that this mechanism clicks in before any optative conflict can even arise.[48]

But even once we exclude such cases, the field of decision encompasses a whole experiential range from the pole of strenuous active intervention to that of effortless attitude formation. Think back to the man at the cross-roads. Perhaps, after much "um"-ing and "ah"-ing, he just finds himself having taken the one fork rather than the other. He had been afflicted by genuine uncertainty, he opted for one of the alternatives and yet cannot even introspectively reconstruct how he came to take the decisive mental step. Should we therefore refuse that step the title of decision? Surely not.

Essential to the concept of decision is not any necessarily active quality, but rather that the concept leaves room for the possibility of such phenomenology. Obviously, it would be a great help if we could clarify precisely what is meant by "activity" here. I have suggested that it has agent-causationist phenomenal character. What it cannot be equivalent to is the exercise of agency, that is, the kind of activity that is constitutive of acting under normal circumstances. That notion of activity is, with certain qualifications, a matter of causation of what we do by wants* whose contents match the descriptions of what is done or of something for the sake of which the thing is done. But precisely this aetiology appears to be missing in the very cases where the particularly active phenomenology is present.

Another way of making the same point is this: a necessary condition of an event involving a person being her action is that that event *could have* been brought about by her deciding to bring it about. But this, as I have argued, is simply not true of specific decisions.

8.7.2 Specifying the Species

So far, I have been primarily concerned to say what deciding is not. The positive claim that it is a form of optative stand is obviously massively insufficient. Are there specifications that might add up to necessary and sufficient conditions? Well, two necessary conditions were already part of the brief characterisation I gave at the beginning of 8.7: that the relevant want* be both occurrent and conscious. Clearly, neither dispositional wanting nor the activation of some want* without the awareness of its bearer can be cases of deciding. But can we hope to adduce sufficient conditions?

An influential feature of Harry Frankfurt's suggestion that I have thus far left undiscussed may seem pertinent here: the claim that deciding is essentially a higher-

[48] Studies by the psychologists Nisbett and Wilson (1977, 243f.) have famously provided strong evidence that, even where people have a conscious sense of actively choosing for reasons, a subattitudinal factor, such as the tendency to choose items on the right, may be "decisive". Where subjects were asked to choose between four identical pairs of nylon stockings, the right-most pair was heavily "over-chosen", although the choosers almost unanimously denied that the object's position was in any way relevant.

order optative matter (1987, 170ff.). Decisions, Frankfurt argues, primarily concern ourselves, a fact given linguistic expression in the reflexive French verb "se décider" or in the English idiom "to make up one's mind". Certainly, it has considerable plausibility that in *some* cases of deciding, for example, in countermotivational cases, an agent might enlist higher-order wants* in the service of a decision. Nosmo's decision not to smoke at the party might be well served by his forming the want* that his desire to smoke not be realised. Where deciding is supported by such an attempt at reflexive self-control, it will involve the occurrence of what can be termed a – negative or positive – realisation-oriented higher-order want*.

However, such optative ascent is, even in consciously occurrent form, neither necessary nor sufficient for decision. Sandwich bar Sam hardly needs to climb to higher reflexive echelons in order just to plump for the ham.[49] And someone may climb the optative ladder without thereby deciding one way or the other. This is most obviously true if he climbs such ladders for two or more competing options. Del, for instance, both wants to go out for a drink and wants to do some work this evening. He imagines the taste of a good beer and the relaxing atmosphere in the bar and finds himself wanting to satisfy the former want. Then he turns to thinking about the how bad he will feel tomorrow if he doesn't make some progress with his work and he develops the desire to satisfy the latter want. And so it might go back and forth. If this is the case, obviously Del has not decided on anything at all.

Assuming that Del manages to come to a decision, must he not do so by adding to his optative "Let it be the case ..." that *executive* je-ne-sais-quoi that somehow puts an end to this otherwise interminable reflective process (cf. Mele 1992a, 160–162)? Fortunately, I don't think so. What I want to suggest that Del needs is an *optative occurrence that brings his – minimal or extended – deliberation to an end*. Decisions are at core (minimal) deliberation-terminative conscious occurrences. This is an idea that is going to need refining a little. The necessary refinements concern both the deliberative episode and the optative occurrence.

First, for someone to have taken a decision, it need not be true that he cease *all* deliberation. If Del has decided to go out for a drink, then he can of course go on to deliberate about whether to go on foot or by bike. The deliberation brought to a stop by the relevant optative occurrence has to be a minimal deliberative episode motivated by a specifiable want*, whose content, we can say, constitutes *the deliberative issue*. The deliberative issue is the disjunction of at least two representations of actions or forbearances, in its minimal form φ v $\neg\varphi$. In other words, it represents the options uncertainty about which inaugurated the deliberative episode in the first place. It is the minimal deliberation on *this* particular issue whose cessation plays a criterial role.

Subordinate deliberation motivated by the desire to adopt ways or means to achieve the end thus decided on is another matter. The same is true of *meta-deliberation* concerning whether to begin or, once one has begun, whether to continue deliberating or not. The point is that decisions are consciously occurrent wants* both *internal to* and *terminative of* a particular episode of (minimal)

[49] For Frankfurt, Sam would be simply choosing, not deciding.

deliberation, an episode individuated by the disjunctive content of its motivating want*, that is, by its deliberative issue.

The second refinement of the basic idea of deliberation-terminative wanting* is necessary because not just any optative occurrence of an agent that happens to cause a cessation of the relevant deliberative episode counts as a decision. For instance, if Del's deliberation on whether to go to the pub or stay at home and work is brought to an end by the spontaneous welling up in him of the desire for a cup of tea, as a result of which he puts the kettle on, thereby losing track of what he had been thinking about, then he has obviously made no decision on the issue at stake. In such a case, the reason why the optative occurrence is of the wrong sort is easily pinpointed: its content again has nothing to do with the deliberative issue. As such it fails to qualify as internal to the relevant deliberative episode.

Note that the deliberation-terminative want* "having to do with" the deliberative issue is not a requirement that the solution to the original incompatibility have already been explicitly represented by the agent as one of those disjuncts. Intervening forms of theoretical inquiry and subordinate deliberation, even unexpected flashes of inspiration, can bring further options into play, one of which may end up as the content of the deliberation-terminative want*. What is decisive is that the agent see the newly introduced option as a replacement for one of the original disjuncts. This is most obviously the case where the incompatibility was given between "doing φ" and "doing something else", i.e. $\neg\varphi$. Here, some ψ need simply be seen by the agent as a way of satisfying the predicate "$\neg\varphi$".

However, as things stand, the requirement that the deliberation-terminative want* be "internal" to the deliberative episode cannot prevent further inappropriate optative episodes counting as decisions. An optative occurrence with one of the contents at issue might still have an effect that brings the agent's deliberation to a close in the wrong way. For example, Del's optative thought to the effect that he spend the evening working might trigger in him an imaginative representation of such a strenuous evening that he faints and thus ceases deliberating.[50] Nevertheless, the thought was no decision, but merely one mental movement in the unfinished process of weighing alternatives.

If the idea of deliberation-termination is understood purely causally, then problems of deviant causality threaten the analysis. Clearly, then, there must be more to the notion of deliberation-termination than the mere causation of an end to deliberation. That this must be so is implicit in the last sentence of the last paragraph, which describes Del's faint-inducing thought as causing the cessation of a deliberative episode that nevertheless counts as *unfinished*. Moreover, it not only counts as unfinished if, on regaining consciousness, Del at some point resumes his practical reflection. Rather, the fact that Del's deliberation has not been "terminated" in the sense at issue is independent of whether he happens to resume deliberation on the matter. The point is that the de facto cessation of deliberation must be seen by the agent as coming about because the content of the causally effective optative occurrence is, as remarked at the beginning of 8.5, *his answer* to the question that

[50] This twist to the example and the objection it illustrates are due to Michael Bratman.

initiated the deliberation in the first place. (Minimal) deliberation is initiated by the optative question, “What am I to do?”, posed with respect to some situation, and is terminated when the agent sees the question as having been answered.

In virtue of what, however, can a particular positive optative stand of an agent on his φ-ing count as *his answer*? I think we should see this as a matter of the agent’s disposition to refuse to reopen the case. The attitude’s content counts as specifying the agent’s answer if, at the moment of taking the attitudinal stand, he is more strongly dispositionally motivated not to deliberate than to deliberate should he be prompted to re-enter deliberation. Del’s deliberation counts as unfinished if he has retained the disposition to deliberate further on the matter, should he be prompted. That disposition can obviously be given even if no such prompting takes place. Where no such disposition accompanies his optative stand, it counts as his answer and its content counts as the content of his decision.

One final point that should be addressed here is this: Why should some optative thought token – for instance, “Let it be the case that I go out for a drink” – suddenly appear to its bearer to answer his original optative question, although he had played host to tokens of exactly the same thought type earlier in the episode without seeing them in this light? It seems likely that there is no general answer to this question. In many cases, we can imagine that deliberation has led to a shift in motivational strengths of the wants* involved in the original incompatibility. But the phenomena of deadlock-terminative and countermotivational deciding indicate that this explanation won’t work in every case. Perhaps the simple fact of re-focussing on one of the want* contents, coupled with other, contingent non-attitudinal changes in the agent, can bring about this effect (cf. Ullmann-Margalit and Morgenbesser 1977, 174). But perhaps sometimes there are the puzzling phenomena at work that only appear explicable in agent-causationist terms. In such cases a person might have a sense that that token of the optative occurrence somehow emanates directly from their being. In neither of these latter two cases can the agent be seen as having a “decisive reason” to terminate deliberation. That is the nature of things when people “just decide” what to do.

Having assembled the requisite materials, I can now advance my proposal: a decision, I submit, is *a conscious optative occurrence that causes its bearer to cease (minimal) deliberation on some issue in virtue of him seeing it as his answer to the optative question the concern to answer which had initiated the relevant deliberative episode*. Thus understood, a decision turns out to be distantly related to Hobbes’ concept of “the Will” as “the last Appetite in Deliberating” (L 36). Crucially, however, the simple property of chronological ultimacy is not sufficient to confer on the optative occurrence its unique practical status.[51] Rather, this derives from the specific conjunction of taking-as-one’s-own-answer and causal efficacy that is necessary for a want* token to qualify as bringing deliberation to a proper close.

[51] And, of course, pace Hobbes, the want* in question need by no means be the last want* prior to action. There need be no immediate transition from decision to action. Indeed, a decisional intention might turn out never to be realised.

Chapter 9
Intentions, Decisional and Nondecisional

Deciding at t to φ entails coming to intend at t to φ. That is, I think, a fairly clear conceptual truth (cf. Raz 1978, 133).[1] However, it seems equally clear that intending to φ does not entail having decided to φ: there are various ways in which we can come to have such intentions. But if this is the case, why should there be a conceptual connection from deciding to intending? For functionalist and normative functionalist conceptions, according to which the conceptually decisive functional roles are causal or normative consequences of intending (cf. Sect. 7.3), the connection to decision must appear to be contingent. But this is surely wrong. Intuitively, the upstream, rather than the downstream connections between intention and decision are primary[2]: it is natural to think that it is the taking of a decision that is responsible for the consequences of the intention thus formed. What speaks most strongly against such an "upstreamist" view is the fact that intentions may develop in other ways and that such non-decisional intentions also appear to have the same consequences as those formed by decision. For the theorist who believes that the connection with decision is primary in the sense just explained, there are three options: firstly, he can argue that the commonalities between decisional and nondecisional cases are so weak that it would be best to give up labelling both with the same term. Secondly, he can attempt to show that there is, alongside decision, an additional kind of genetic condition that explains why non-decisional intending generates precisely the same consequences as its decisional relative. The third option is mixture of the other two. To adopt it is to argue that there is a significant, but incomplete overlap in the consequences of the attitudes generated in the two differing ways, that the consequences that are common to the two kinds

[1] Bratman has argued that what he calls a "choice" doesn't entail having an intention with the same object, as a package can be chosen only part of the content of which need be intended (Bratman 1987, 152ff.). I respond to this objection in Roughley 2007c.

[2] The contrast between the two families of theories was suggested by Michael Bratman in written comments on an earlier draft.

N. Roughley, *Wanting and Intending*, Philosophical Studies Series 123,
DOI 10.1007/978-94-017-7387-4_9

of case can be equally explained by either aetiology and that these commonalities suffice to explain our subsumption of both cases under the same concept. It is this third strategy that I shall be adopting.

In this chapter I shall begin by explaining how we get from the analysis of decision to an analysis of intention. Because things we make don't persist indefinitely, we need to say more about the conditions for decisional intending than merely that it is brought about by deciding (Sect. 9.1). After doing so, I turn to nondecisional intentions and distinguish five distinct kinds (Sect. 9.2). I make no claims that the typology is exhaustive (it isn't). My aim is to distinguish a sufficiently wide variety of nondecisional cases so that the analysis I go on to propose can plausibly be thought to cover those attitudes we would think of as intentions without decisional aetiologies. I then argue that these cases differ from decisional cases relative to the question of belief constraints (Sect. 9.3). This difference supports the claim that the word "intention" doesn't always refer to the same type of phenomenon, i.e. that the unity of intention thesis is false. In the next three sections I develop the idea that the contextually unique practical status of nondecisional intentions grounds in a combination of motivational strength (Sect. 9.4), conscious wanting* (Sect. 9.5) and unresponsiveness to the triggering of the dispositional want* to deliberate that characterizes full agency (Sect. 9.6). On the way, a discussion of the relation of non-decisional intentions to conclusive reasons judgements gives me occasion to propose a further intention-antecedent requirement of practical rationality that is broader than *CRD*. I conclude by bringing together the results of the discussion in a disjunctive definition of intention (Sect. 9.7).

9.1 Decisional Intentions

Although a decision to φ is, trivially, a necessary condition of decisionally intending to φ, someone who has decided to φ doesn't necessarily intend to φ. The obvious reason for this concerns the persistence of the attitude thus generated. Any attempt to give the concept of decisional intention clear contours must therefore propose conditions for intention's persistence. In particular, it needs to name non-ad-hoc conditions whose satisfaction defeats the inference from "*X* has decided to φ" to "*X* intends to φ". To develop my proposal, I will discuss, first, deliberative and, second, non-deliberative intention revision, before considering the question of whether doxastic changes might be sufficient to count as revoking an intention that was decisionally formed.

To make things relatively simple, I will focus on a short-term intention, such as one generated when Janice decides to go and buy a jam donut in the course of the morning. Under what circumstances should we cease to ascribe her the intention? As we saw in the discussion of *DIP* (Sect. 7.2.3), the answers correspond, up to a point, to the reasons why someone who has taken a positive conscious optative stand on some *p* might no longer want* *p* (Sect. 5.1.2). Most obviously, she might change her mind (i) and decide to buy a muesli bar instead. However, as we have seen, there are various ways of shedding intentions that are not helpfully counted as changing one's mind: Janice might become so absorbed in her work that she permanently forgets (ii) what she had decided to do. Obviously, going and buying the donut, i.e. realizing her intention (iii) will generally result in the intention's dissolution. Moreover, it seems that, under the advent of conditions that make the intention irrelevant (iv) – if, for instance, a visitor comes round laden with cakes – Janice might also shed her intention without needing to spare a thought to doing so.

(i) Change of mind is perspicuously thought of as active abandonment rather than unnoticed shedding of an intention, where the activity involved is the activity characteristic of decision. Change of mind is thus a matter of deciding anew and requires prior minimal deliberation. A change of mind takes place if some optative competitor to a prior intention terminates a new episode of minimal deliberation whose issue includes the prior intention's content. The want* may survive this process, but simply no longer count as an intention: Janice might still welcome a donut brought round by a visitor. As every more recent deliberative episode cancels out the criterial relevance of earlier episodes of deliberation, our criterion needs to specify that the relevant decision *have terminated the agent's most recent minimal deliberation on the issue*.

This point holds even for cases in which the agent hasn't changed her mind, but has merely re-entered deliberation. An optative attitude is disqualified as an intention as soon as the content of that attitude is fed into new deliberative machinery (cf. Sect. 7.2.3; Holton 2009, 121) – until the point at which it perhaps re-emerges in a new deliberation-terminative role. As wants* are the material out of which intentions are fashioned, old intentions can be put back into the deliberative process and end up being recycled as new intentions. Whether that happens will depend on whether they end up playing the deliberation-terminative role anew. Change of mind and reconsideration are ways in which deliberatively formed intentions can be revoked. This is done by reengaging the deliberative mechanisms that led to the intention being formed in the first place. It is therefore a simple and consistent move for our theory to exclude this step.

(ii–iv) Things are, however, not quite so simple when we turn to ways in which intentions may be *non-deliberatively shed*. If the intention permanently slips Janice's mind,[3] if someone comes round with Black Forest Gateau or if she goes and buys the donut, in the normal run of things that will be the end of her intention. In the first case, there is ex hypothesi no event of conscious intention abandonment. In the latter two cases, there could be. When the cakes arrive, Janice might just think "So much for the donut" and in the case of intention realization, she may have a whole list of things to buy that she mentally ticks off as she purchases them. However, the intention's demise obviously doesn't require any such events. What unites all these

[3] Note that, if a person has temporarily forgotten that he has decided to φ, that does not (pace Audi 1973b, 65f.) entail that he no longer intends to φ. Someone may react to her memory being jogged with respect to something she has been planning with the words "I've been meaning to do that for ages, but almost missed the opportunity, as it had slipped my mind" (cf. von Wright 1971, 105f.). In spite of having forgotten her intention, such a person clearly still intended to φ. Permanent forgetting, on the other hand, involves intention dissolution.

cases is that Janice not only ceases to intend to get the donut, she also ceases to want* to do so.

What a genetic theory of decisional intentions needs here, then, is simply to exclude cases in which the agent has ceased to want* the intention's content. Because of intention's "supra-optative" component, this is a meaningful condition. Moreover, it picks out the feature common to the three kinds of case where there is no relevant conscious thought token. Note that, absent want* dissolution, an intention's becoming irrelevant or fulfilled is not a defeating condition of its persistence, but at most a defeating condition of its rational persistence. It is conceivable that someone who forgets that she has achieved her goal might continue to intend to achieve it. And even if Janice's decision to buy the donut had been based on the belief that she needs to maintain a certain blood sugar level, something that the Black Forest Gateau would also achieve admirably, it is conceivable that Janice might find herself strangely and stubbornly holding onto the donut intention, perhaps because she resents people second-guessing her needs. As such cases illustrate, there is such a thing as irrational intention persistence. For this reason, mere changes in external circumstances are insufficient for an intention's dissolution.

To the claim that, absent change of mind, abandoning an intention requires the agent to cease wanting* its content, it might be objected that there could be a non-deliberative degeneration of an intention's commitment component without the want* dissolving entirely. Janice, it might be thought, could under certain circumstances cease to intend to get the donut, whilst still hoping for its arrival in the company of a visitor. She might simply have come to think it too much bother to go out and get it.

The decisive question here is whether this is conceivable without any form of even minimal deliberation. I think that we should answer in the negative. Firstly, there appears to be no phenomenological evidence that such non-deliberative movements of mind take place. Secondly, the only theoretical reason to assume such a possibility is an irreducibilist conception of the commitment component of intention, the nature of which is the issue at stake. Coming to think that the walk to the bakery involves too great a cost to make the benefit of the donut worthwhile is naturally taken to be a deliberative process. It obviously involves weighing up costs and benefits of means and ends. Now there are plausibly analogous processes carried out subpersonally in human and animal agents that are immediately involved in behaviour. But these are best thought of as structuring the agent's motivational processes. Janice may, without further deliberation, find herself strongly unwilling to go the bakery, a fact that may in turn affect her motivation to realize her intention. However, as cases such as that of Nosmo (Sect. 8.2.3) show, even if she were to find herself more strongly motivated not to buy the donut than to buy it, it would not follow that she had ceased to intend to do so. Clarity on the shift in her motivation could, of course, persuade her to rethink her intention. But this would involve reconsideration.

So the account sees decisional intention persistence as curtailed only by either want* dissolution or the reinauguration of minimal deliberation. However, it may appear that strong counterexamples to this account can be found if we turn once again to the role of belief. Perhaps Janice suddenly remembers that today is a bank holiday and that therefore the shops are closed. It follows that her envisaged donut purchasing activities are a non-starter. She continues to long for a donut and maintains the instrumental want* to go and buy one. But, so it seems, her attitude towards something she has now come to believe unfeasible cannot continue to count as an intention. Moreover, this appears to be documented by the incoherence of a thought such as "I intend to go and buy a donut, although I know that it is impossible for me to do so" or, even more starkly, "I intend to go and buy a donut, but I won't".

In spite of first appearances, however, I think that even the advent of such a belief can leave the relevant want's* status as an intention intact. This has a fairly simple reason: attitudinizers don't necessarily conjoin attitudes that are relevant for each other.[4] The condition under which Janice would not be able to avoid putting them together would be her playing host to both *consciously*. I argued in Section 6.3.3 that the incoherence of self-ascribing an intention to do something and expressing the belief that one won't, or cannot do so, depends on the fact that both first-person attitudes must be thought of as conscious at the time of their self-ascription or expression. Note that peculiar cases of alienation, in which an agent expresses the intention to φ in conjunction with the inferentially based self-ascription of the belief that she won't φ ("I'm going to go and buy a donut, but apparently believe that I won't do so") are perfectly coherent. This is so because self-ascribing the belief expresses the speaker's second-order belief that she is the bearer of a first-order doxastic attitude. There is no strict incompatibility between the attitudes of the two different orders because the higher-order self-ascription of the first-order belief leaves it open whether the first-order belief is true or not, for instance, whether the first-order belief might have been caused by mechanisms that should be doxastically irrelevant. Where there is a conceptual incompatibility between a belief that one won't perform some action at *t* and a decisional intention to perform it at *t*, this hinges on the agent taking the conscious stand of truth-taking relative to her non-performance of the action.

Because the only strict conceptual relations between decisional intention and belief relate either conscious tokens of the two types of attitudes or else conscious beliefs and processes of deliberative intention formation, no doxastic transformation necessarily defeats decisional intention persistence.

My claim in this section, then, has been that a decisionally generated intention stands unless the agent either ceases to want* the intention's content – as a result of permanent forgetting, of realizing the content, of the content's becoming irrelevant, perhaps as a result of physiological changes – or reenters deliberation on the issue of the intention's content. The claim is actually not quite correct, as we will see in Section 9.6. In the meantime, we should be clear that there is nothing ad hoc about the two conditions: they don't introduce new, otherwise unmotivated criteria

[4] Cf. Section 4.1.2, where I argued that it is perfectly possible to have contradictory beliefs, but not to agglomerate them and thus come to believe a content with contradictory conjoins.

merely to deal with recalcitrant cases. Rather, they are simply required in order to ensure that the event central to the genetic conception – the termination of minimal deliberation about something wanted* – retains its influence on the agent. We can thus provisionally define decisional intention as follows:

DI'
X decisionally intends to φ iff:
X wants* to φ and
has, on the basis of his most recent minimal deliberation on whether to φ, decided to φ.

9.2 Five Ways to Nondecisionally Intend

It is fairly obvious that we don't only see people as intending to perform actions they have taken decisions to perform.[5] There are at least five kinds of nondecisional intentions of which a genetic theory of intending has to take account.[6]

Intentions are, firstly, sometimes generated *spontaneously*. In such cases, we take the first available option without comparing other possibilities or hesitating in the slightest. In Section 8.5.2, I mentioned the case of someone simply plumping for a ham sandwich after a mere moment's hesitation. In such a case of "pure minimal deliberation", the agent vacillates momentarily between taking the sandwich and not taking it, without weighing reasons. Spontaneously generated intentions come into being without even such a moment's hesitation. Someone takes a drink from a tray (Raz 1978, 134) or a biscuit from a plate (Velleman 2007, 197f.); similarly, on seeing an elderly person in need of help, someone might form the spontaneous intention to cross the street and help them (cf. Watson 2003, 125).

Secondly, we can acquire intentions as a result of *specifying* of the content of an intention we already have. An agent, arriving at a train station with the intention of taking the next train to *Z*, consults the timetable and, on seeing that the next train goes at 10.32, acquires the intention to take the 10.32 train to *Z* (cf. Meiland 1970, 55). The step is non-decisional, as the agent does not need to overcome any kinds of doubts in order to take it. Indeed, it is automatic in the sense that the agent doesn't need to attend to any features of the machinery involved in making the mental move in order to make it. Nevertheless, there is at least a sense in which a new intention is acquired: the person is now set on a way of fulfilling his previous intention that both opens new epistemic opportunities (he knows he's got time to buy a paper) and subjects him to new requirements (he's only got 10 min to do so).

[5] That there is such a divide within the class of the mental states we call intentions has been repeatedly remarked upon (Hampshire and Hart 1958, 3, 12; Chisholm 1970, 645; Raz 1978, 133ff.; O'Shaughnessy 1980 II, 297; Velleman 1989, 112; Pink 1996, 20).

[6] At least six, in fact. See below Section 9.5.

A third mode of nondecisional intention formation is intending *out of habit*. Habitual attitude formation appears to ground in dispositions to take on the relevant attitude on perceiving certain kinds of things in certain contexts. Think about what it takes attitudinally to get a suburban commuter to her place of work in the morning. After waking, she perhaps drowsily forms an intention to turn on the light (Pink 1996, 20); somewhere in the course of the next hour or so she comes to intend to take some particular train as a part of her regular routine (Meiland 1970, 55). And so on, where the intentions thus formed are taken on unhesitatingly as a matter of course.

There are, fourthly, non-decisional intentions whose genesis is *gradual* in nature. Someone might, so it seems, gradually form the intention to leave her home town. Here various metaphors suggest themselves: intentions can "grow on you" or can, to appropriate a formulation of O'Shaughnessy's, "crystallize" over time (O'Shaughnessy 1980 II, 301).

Finally, some intentions are generated by nondeliberative rational processes on the basis of reasons judgements. Sometimes people acquire the belief that under certain circumstances they would have conclusive reason to φ, or not to φ. This belief might itself result from a judgement passed outside of practical deliberation, i.e. without the agent's reflection on reasons having been motivated by optative uncertainty. Such *conditional conclusive reasons judgements* can result in the agent being set to intend to φ under the relevant circumstances without having to reflect further. For instance, someone might as a young person have been convinced that there are conclusive reasons for her to refuse heroin if it should ever be offered to her. This may lead her to be organized in such a way that, if she believes that she is being offered heroin, she unthinkingly forms the intention to refuse it.

Being dispositionally organised in this way complements the disposition to conform to *CRD*, the rational principle relating conclusive reasons judgements to decisions (Sect. 8.5.2). Roughly, *CRD* demands that rational agents decide in line with their conclusive reasons judgements in those cases in which they take action to realise the judgement to require a decision. It thus presupposes that there are cases in which no decisions are necessary. A substantial ground for this presupposition is that rational agents are characteristically disposed to nondeliberatively intend in line with conclusive reasons judgements. Rational agents are the bearers of such subpersonal connections, connections that qualify intentions, like beliefs, as "judgement-sensitive" in the sense proposed by Scanlon (Scanlon 1998, 20ff.; cf. Broome 2007a, 367).

There are, then, a number of ways in which intentions can be nondecisionally acquired. There might, however, be doubts as to whether all the kinds of example I have mentioned are actually cases of φ-ing as a result of intending to φ. Such doubts may particularly concern spontaneous and habitual intentions and can be more or less strong. The stronger doubts question whether the agents are even behaving intentionally. Weaker doubts ground in the conviction that there are spontaneous or habitual actions we perform intentionally without having intended to perform them. In other words, it may be thought either that spontaneous or habitual actions should not be seen as forms of intentional action at all or that they are intentional actions performed without corresponding intentions.

Leaving aside for the moment the question of the relationship between the intended and the intentional, let me simply say why I take it to be obviously true that there *is* such a thing as spontaneous and habitual intention formation.

First, someone who sees an acquaintance on the street and greets him will under normal circumstance “mean to” greet him. Note that spontaneous intention formation need not lead to spontaneous *action* in the sense of a split-second reaction to some feature of one’s environment. I may spot my neighbour at some distance off, form the spontaneous intention to greet him – no considerations occur to me as why I shouldn’t – and finally greet him when our paths cross. This is intentional action, and intentional action on the basis of setting oneself to behave in the relevant way. So neither the strong nor the weak doubts have any purchase here.

Second, what about habitual action: do we have any good reasons to think that acting out of habit is necessarily acting without an intention?[7] I think not, for reasons that exactly parallel those just given for spontaneous intentions. However many routine actions may be performed without the formation of prior intentions,[8] all that is necessary here is that people sometimes form intentions as a matter of habit. Again, this seems demonstrable by plausible cases in which the performance of a routine action is postponed for some reason. When the alarm rings in the morning, that appears to trigger the habitual formation of the intention to get up, at least in some people. One sort of evidence for this would be the thoughts that the agent might have if another agent were to employ means to dissuade him or her from getting up at that moment. Or take someone who cycles to work. The fact that certain stretches of the journey are harder work than others may well automatically engage her consciousness in thoughts such as “OK, let’s do some serious pedalling to get up the hill today!” If such routine situations can generate such conscious intention-expressive thoughts without any doubts needing to be overcome, then there are such things as non-decisional habitual intentions. The phenomenological evidence seems strong enough for us to justifiably take it that there are.

9.3 Doxastic Conceptual Constraints

One significant difference between decisional and non-decisional intentions concerns their relationship to belief. If, as I argued, the relations between belief and decisional intention ground in the latter’s genesis in minimal deliberation, then

[7] There are philosophers and psychologists who suggest that this may be the case. An example of the first is Timothy Schroeder (Schroeder 2004a, 22), examples of the second are David Neal and Wendy Wood (Neal and Wood 2009).

[8] Timothy Schroeder (ibid.) describes an example meant to show that habitual action is not to be explained in terms of “desire”. It concerns an agent who has an ingrained habit of turning left at an intersection on the way to work, who changes jobs, leaving him with no desire to take a left turn there, but who, on returning one day in order to go to a grocery store on the right of the intersection, nevertheless turns left. The case involves neither action out of an intention nor even action describable as intentional.

any connection of nondecisional intention to belief will have to have some other explanation. I shall argue that there is in fact no such general conceptual connection in the case of nondecisional intentions.

We can begin with the fact that nondecisional – like decisional – intentions are able to exist alongside beliefs of the agent that she won't realize their content, where the agent is in some sense absent-minded. If Janice either habitually forms the intention to go and buy a donut every day or if her intention is a one-off formed spontaneously on seeing a donut pictured in some advert, she might do so absent-mindedly in spite of her belief that the shops are closed today. On the face of it, the examples seem to be pretty much the same as in the decisional case. A significant difference between nondecisional and decisional cases, however, lies in the fact that nondecisional intentions may also be *formed* non-absentmindedly in the face of the agent's conscious belief in the impossibility of his realizing its content.

Take the case of young Vic, who is regularly victimised by Billy, to whom he has never had the courage to stand up. Vic has always believed he could never hit back at Billy. However, one morning on which his powerlessness is bothering him particularly acutely, he surprises himself by suddenly thinking "Right! ... " and swiping out at his tormentor. This narrative seems to me coherent, but I can see no necessity that the sudden welling up of anger and courage that led to Billy's action be accompanied by a retraction of his belief in his own inability to act as he goes on to do. Vic's consciously playing host to a belief with negative content means that he can only come to satisfy the negative condition by making *some* doxastic move. But the spontaneity of his intention formation is at odds with any such prior doxastic step.

Now, the relevance of the example could be disputed by simply denying that Vic can have hit out as a result of intending to do so. More strongly, someone might deny that Vic's hitting out is even intentional. However, although examples can certainly be constructed in which an action like Vic's might be less than intentional, there appear to be no a priori arguments why the case must be of this kind. Compare the examples of Inge and Pablo: Inge, suddenly indignant at her mother, spontaneously goes to strike her, but then retracts at the last microsecond; Pablo, a passer-by overcome by a sudden stirring of "fellow-feeling", spontaneously, goes to hug a roadside beggar, but pulls back at the last moment. In either case, it seems plausible that, had the agent performed the action they began spontaneously, they would have done so unintentionally. Inge and Pablo might then express this by saying that they "didn't mean to" do what they did.

But – and this is the question to which we need an answer here – in virtue of what would they be right? As far as I can see, the answer has to appeal to a *conscious thought*, a conscious thought that is naturally seen as intention-formative. If either Inge, Pablo or Vic perform some spontaneous action without consciously "going for it", I suggest that we should see them as acting subintentionally, that is, as φ-ing, where they want* to φ, but intend neither to φ nor to bring about any state of affairs the bringing about of which involves their φ-ing (cf. Sect. 5.1.3). This is one version of the cases under scrutiny. There is however a second, which involves the protagonist thinking a conscious thought expressible as "I'm going to φ". Where such a thought is indeed consciously tokened, I submit that their action is both intentional and intended; where no such thought crosses the agent's mind, a spontaneous action of the sort we are considering should be classified as subintentional.

The decisive point in our context is that a spontaneous thought of this kind appears compatible with not taking the step of negating a standing, contemporaneously conscious belief that one won't perform the action in question. Here we have a significant difference from decisional intention formation. Agents form decisional intentions *in order to* enable the results of their deliberation to control their action in the face of optative uncertainty. To decide to φ – rather than to do something other than φ-ing – is to take a step that is necessarily conscious as a practical mental move made in order to overcome doubt about the deliberative issue. As the concluding step of a process set in motion with a particular aim, then, it is incompatible with a conscious belief in the realization of defeating conditions for that aim. In contrast, spontaneous intention formation involves no such perspective on the purpose of the acquisition of the intention.[9] Indeed, it seems that, in such cases, the intention-formative thought can simply be acquired out of the blue: an agent can, as Vic does, take himself by surprise in forming the intention.

Because other kinds of non-decisional intentions are also acquired outside a framework set up by the agent in order to facilitate their action, their acquisition is also compatible with conscious beliefs that they won't, or can't be realised. There is one exception, to which I will come at the end of this section.

Gradual intention acquisition may be simply a slow version of the formation of spontaneous intentions. A certain habit of mind – the disposition to regularly token conscious doxastic thoughts to the effect that one is incapable of doing something – may still be in place when a gradually developing intention reaches crystallization point. The Vic and Billy scenario might be conceived in just this way. If Vic's sudden resolution is the effect of a gradual motivational transformation, the case may indeed be better thus described.

The relationship to conscious belief that may arise in the case of habitual intending can be thought of as the converse of the gradual case. Take an agent brought up to pray to their god every morning before breakfast, who is convinced by an atheist that his god doesn't exist and that he therefore cannot pray to him. This new convert to atheism may nevertheless find himself with the intention to begin his usual prayer as he sits down to eat his breakfast. Whereas spontaneously or gradually acquired intentions can co-exist with rationally incompatible habitual or accustomed beliefs, habitual intentions, generated by stable internal mechanisms, can co-exist with rationally incompatible, reflectively acquired convictions.

A gap can also open up between judgement-derivative intentions and the agent's beliefs as to whether she will perform the intended action. Part of what it is for

[9] It is thus no coincidence that, where Jay Wallace makes the important point concerning the agent-relative modality of "the instrumental principle" (cf. Sect. 7.2.3), he explicitly restricts his discussion to the "notions of necessity and possibility at work in practical deliberation" (Wallace 2001, 24).

an agent to be rational is his being structured in such a way that, should he be the bearer of a judgement that he'd have conclusive reason to φ in situation *s* and should he come to acquire the belief that he is in *s*, he tends to automatically acquire the intention to φ. As these rational processes run at a subpersonal level, it is conceivable that they may result in an agent intending to φ, in spite of his consciously believing that he is unable to φ. Someone may, for instance, dispositionally believe it is best always to take an umbrella if there are dark clouds in the sky, develop the perceptual belief that there are dark clouds in the sky and as a result form the intention to take her umbrella, even though she at the time believes that the umbrella was chewed to bits by the dog the previous day.

The exception to the compatibility of an agent's nondecisionally acquiring an intention to φ and consciously believing herself unable to φ is provided by certain cases of specificatory intention generation. If an agent intends to φ at the next possible opportunity, acquires the belief that doing so requires that she ψ and has no thoughts concerning reasons for not ψ-ing, she will under normal circumstances automatically acquire the intention to ψ. The reason why she cannot acquire this specificatory intention to ψ whilst consciously believing herself unable to ψ is that acquiring the belief that triggers the specificatory intention together with a belief in her inability to ψ would require that she believe a contradiction. This is because the belief that the next possible way of φ-ing is to ψ not only entails, but contains the belief that it is possible to ψ. The doxastic constraints here are thus imposed by a specific feature of the thought processes required for such case of specificatory intention generation. They certainly don't carry over to all non-decisional intentions.

9.4 Being Set: Nondecisional Intention and Motivation

9.4.1 Being Set

There are, then, no general conceptual constraints relating nondecisional intention to belief. This indicates that the attitudes we call intentions outside of minimally deliberative contexts differ significantly from their deliberatively generated relatives. Intentions are optative attitudes that have an eminent status relative to action control. One way in which they can attain this status is through decision. If the thesis of the unity of intention is false, then there must be a second way in which such an eminent status is conferred. I shall be claiming that it is conferred by a combination of various factors.

We can approach the first of these via one of the metaphors often used in the literature to pick out that extra something involved in intending that goes beyond optative attitudinizing. As remarked in Section 6.1.2, it is frequently said that an agent intending to φ is "set on" φ-ing, something that clearly need not be the case if she merely wants* to φ. In nondecisional cases, agents are, it seems natural to say, "set on" performing an action without having "set themselves" to perform it. As Robert Audi has argued (1988, 243f.), the most natural interpretation of the notion at work here is motivational.

Leaving aside general skepticism about the concept of motivational strength (cf. Mele 1998, 27f.; 2003a, 164f.), it seems that an agent performing action φ, where she does so neither unknowingly, by mistake, luckily or as a mere side-effect, must be the bearer of a *motivationally unrivalled want** to φ.[10] A motivationally unrivalled want* is a want* that is motivationally stronger than any wants* that can be labelled "optative competitors". An optative competitor is in turn a want* to perform (or to omit) any other action or set of actions the performance (or omission) of which would prospectively either (i) make the agent's φ-ing impossible or (ii) lead to a reduction of the level of her motivation to φ below that required for her φ-ing.[11] The motivational rivals of a want* to φ are those optative competitors backed by sufficient motivational strength to lead the agent not to satisfy her want* to φ.

Now, if an agent's being motivationally set on φ-ing can explain her φ-ing and if the same action is also explainable by an agent intending to perform it, then there is a prima facie case for identifying unrivalled wanting* and intending. For the reasons discussed in Section 8.2, such an identification would be obviously wrong if intention were a unitary phenomenon: decisions and therefore decisional intentions are neither exclusively explicable by, nor identifiable with motivationally unrivalled wanting*. However, once we consider intentions with no deliberative aetiology, the case for seeing their "supra-optative" feature in motivational terms begins to look considerably more attractive.

Consider for a moment Mele's claim that the nondeliberative generation of intentions is to be understood as a "default process", whereby preponderant motivation leads directly to the acquisition of corresponding intentions, so long as nothing prevents this happening (Mele 1992a, 168, 190). I want to suggest that Mele is postulating *one attitude too many* here. Under certain further conditions to which I will turn shortly, our motivationally unrivalled wants* *are* our intentions. If we take it that both the irreducibility thesis and the unity of intention thesis are up for grabs, then we can in this way provide a more economical explanation of actions that are not preceded by deliberation. Nondeliberative intending, and a fortiori nondeliberative intended action can be given more parsimonious explanations.

Importantly, this parsimony is not only locally quantitative, keeping down the number of instances of the entities the theory works with in one place. It is also a consequence of the qualitative parsimony of a reductive conception that rejects

[10] This is an alternative formulation of the principle adduced in Section 8.2.3, note 14. The action φ may be an attempt to perform some further action ψ. This is particularly relevant where strengths of wants* to realize incompatible aims are modified by the subjective probability of attaining those aims by the available actions. Under such circumstances an agent can be more strongly motivated to $\varphi 1$, i.e. to attempt to $\psi 1$, than she is to $\varphi 2$, i.e. to attempt to $\psi 2$, although she would be more strongly motivated to $\psi 2$ than to $\psi 1$, if she believed she could directly perform either of the latter actions. Al Mele has been at pains to make this point (1992a, 52ff.; 1998, 29f.; 2003a, 166f.). Note, however, the appropriateness of the *conditional* ascription of stronger motivation to $\psi 2$ than to $\psi 1$. Strictly, the agent is more strongly motivated to $\varphi 1$ than she is to φ, but would, under other doxastic conditions, be more strongly motivated to $\psi 2$ than to $\psi 1$.

[11] This is basically the concept of motivational competition proposed by Mele (1992a, 66).

the need for a new sui generis attitudinal type.[12] Clearly, though, these points have to be weighed against the quantitative increase in explanatory structures invoked by a disjunctive theory that rejects the unity of intention thesis. In a prominent passage, Elisabeth Anscombe claims that, although we are tempted to see the word "intention" as having different senses, such semantic equivocation is implausible (Anscombe 1957, 1). However, as Velleman points out, Anscombe provides no argument for the thesis of intention's unity (Velleman 1989, 113).

Velleman himself conceives his own distinctive theory of intention (cf. Sects. 6.3.1 and 8.5.1) explicitly as a theory of what I have been calling decisional intention. Alongside "the state of being decided upon one's next action", the term "intention" can, Velleman claims, also refer to "an agent's ultimate motivating desire", a state which Velleman thinks is also picked out by the phrases "the intention with which one acts" and "the agent's goal" (Velleman 1989, 112). Velleman is, he says, only interested in decisional intentions, which he associates with "plans", rather than with "goals".[13] I think we need a theory that distinguishes the two kinds of states, but also clarifies their relationship. In what follows, I first argue for the conceptual significance of motivation for non-decisional intentions, before going on to explain why optative attitudes distinguished by this feature should be conceived as covered by the same concept picked out by a decisional genesis.

9.4.2 Motivation and the Typology of Nondecisional Intentions

A first thing to notice about nondecisional intentions is that all the five types we have looked at can plausibly be understood in motivational terms.

To start with, this looks fairly obvious in the specificatory case, which appears easily understandable in terms of the transfer of motivation from some aim to an unchallenged, but simply more concrete version.

[12] The distinction between qualitative and quantitative parsimony, which derives from David Lewis, is quoted by Bratman, along with Lewis's normative prioritisation of the former, in Bratman's defence of a reductive theory of shared intention (Bratman 2014, 106). Bratman, of course, doesn't think that qualitative parsimony is a decisive consideration for an understanding of individual intention.

[13] Velleman seems here to be simply adopting Bratman's terminology for the first type of intention (cf. Sect. 6.2, note 14). Whether one should talk in this way depends on the concept of plan at issue. Obviously, if a plan involves having thought through the ways or means of realizing a goal, then its genesis will have to have involved deliberation. Note, however, that deliberation on this level need not involve a deliberative genesis of the superordinate intention. Thus, even on such a deliberative conception of plans, there is no necessary equivalence between decisional intending and planning. There is no necessary association between planning and deliberation if the concept of plan in question requires only the existence of causal connections between goals and representations of ways or means of their realization. Habitually generated intentions often come complete with subordinate intentions that enable a whole series of actions to be performed in order to achieve the automatically generated aim.

Turning to spontaneous intention generation and the case of Vic and Billy: it seems perfectly plausible that the explanation of Vic's fist-swinging needs simply to refer to the sudden welling up in him of new levels of anger. If anger either involves or generates a want* to launch some kind of attack on the emotion's object (cf. Aristotle Rhet 1378a 30ff.), then the intention might be simply the want* involved as from the point at which its motivational force becomes unrivalled. Vic undergoes an affective and motivational transformation, where the latter change, once it involves the crossing of the motivational threshold, is precisely the genesis of his intention. Both the onslaught of the new level of motivation and the intention equally explain his behaviour.

This conception, whilst removing the conceptual necessity of any belief condition, nevertheless explains why spontaneously formed intentions are characteristically accompanied by a significant level of confidence (cf. Sect. 6.4). Someone who is aware of being strongly motivated to perform some action is ceteris paribus more likely to believe that she will perform it than she would be were she to take herself to be only weakly motivated to do so. Significantly, the confidence that characteristically accompanies intending could rationally result either from such awareness of one's motivation or from having decided to perform the relevant action. So the same feature of the intentional syndrome could be explained by two distinct phenomena, either of which may justify talk of intending.

A motivational analysis also helps to clarify why gradual intention formation is basically a slower version of the spontaneous case. Motivation can, so it seems, surge suddenly or just increase gradually, presumably without the awareness of its bearer. Someone who has moved to a new city may gradually, as she begins to feel more and more at home, become increasingly motivated to spend the rest of her life there, a process that may lead to a corresponding intention crystallizing, without her needing to make any decision. The crystallization of the intention is, I am suggesting, at core the crossing of a motivational threshold.

It has been suggested that cases of this ilk are well described as examples of someone taking an "unconscious decision". Velleman tells the story of an agent who has "decided" to sever a relationship without realizing that he has thus decided, only becoming aware of his "decision" later when he tries to make sense of his abrasive behaviour (Velleman 1992a, 126). This way of talking corresponds to certain everyday uses. However, I think such uses are best understood as expansions of the core use, which picks out the conscious resolution of optative doubt (cf. Raz 1978, 133f.). "Decision" does appear to be the everyday term best suited to mark the distinction between intentions thus formed and others – although in the end the important point is the distinction, not the term.

There is an intermediate usage to be found in Dennett and O'Shaughnessy, who describe agents who at t_1 are optatively uncertain with respect to the alternative of φ-ing or not φ-ing, but who find themselves intending to φ at t_2, without having consciously decided to do so. In these cases, "decisions" – if the resolution of optative uncertainty is sufficient for decision – have been taken non-consciously: in the agent's sleep (O'Shaughnessy 1980 II, 301) or in the process of going about one's everyday business (Dennett 1984, 80). My response is the same as that to examples of Velleman's sort: it is important to mark the cases in which the resolution of optative uncertainty terminates minimal deliberation. The examples in which that uncertainty is no longer there in spite of the agent not having taken any conscious step to terminate it are plausibly cases in which the balance of the agent's motivation tips in one particular direction as a result of nondeliberative processes. Dennett certainly sees his usage of "decision" as marking, indeed necessarily marking motivational changes, changes that are sufficient to have effects on the agent's action (cf. Sect. 8.2.1, note 4). Although my terminological proposal – to use "decision" independently of motivation and to think of the acquisition of intentions through motivational transformation as "nondecisional" – runs counter to Dennett's, I agree that gradual motivational changes can be constitutive of coming to intend.

For routinely generated intentions to cause particular problems with my suggestion, there would have to be some reason why such automatic processes cannot lead to motivational changes. However, if we have accepted that automatic processes can not only lead directly to automatic actions, but can in at least some cases generate intentions, which in turn lead to action, it is difficult to see why that generation of attitudes should not lead to them being backed by specific levels of motivational force.

Turning, finally, to forms of automatic judgement-derivative intending, it certainly looks plausible that the practical dispositions of bearers of conclusive reasons judgements should rationally be a matter of their motivation. In order to see why, consider the relationship of the form of rationality at work here to the parallel requirement codified by *CRD* (Sect. 8.5.2). Very roughly, *CRD* requires agents to decide in time to do what they (continue to) judge themselves to have conclusive reasons to do. Nondecisional cases, it seems natural to assume,[14] are going to be covered by a broader principle, from which *CRD* derives.

That principle may appear obviously to be more or less what Broome has called the enkratic requirement,[15] the primary difference being merely the fact that the consequent of the broader requirement concerns intention, whereas that of the narrower requirement concerns deciding. However, I shall claim that there is a requirement that is, at least in one dimension, broader than any principle that merely relates conclusive reasons and intentions and that the recognition of this requirement will help us clarify why nondecisional intentions should be understood as forms of motivationally unrivalled wanting*.

Broome is explicit that the automatic generation of judgement-derivative intentions isn't covered by the requirement he calls "enkrasia". Like his version of the requirement on subordinate intentions (cf. Sect. 7.2.2, note 39), Broome's enkratic requirement only covers the generation of intentions without which the

[14] I remain on the fence here as to whether this assumption is in fact true. *CRD* cannot be derived from the requirement I sketch below. Whether there is some way of shedding the restrictions it incorporates, so that some such derivation might be possible is a question that goes beyond the issues I need to address in order to defend my analysis of nondecisional intention.

[15] Cf. the references in Chapter 8, note 27.

agent believes she won't perform the action in question (Broome 2013a, 434). In arguing for my formulation of *SI*, I opted against including such a doxastic premise state whose content refers to the agent's own intentions, as doing so massively restricts the requirement's scope. There are, it seems, clear cases in which not intending a subordinate action is irrational if the agent believes the action is a necessary means or way of realising his superordinate intention, independently of any opinions she may have about the necessity of intending.

By the same token, failure automatically to generate judgement-derivative intentions may also be irrational in the absence of the belief that such an intention is necessary. A driver trying to find his way to some unfamiliar destination may judge it best to turn right at the next opportunity. If, when the next street on the right appears, he doesn't form the intention to turn right, something has gone wrong. But this is not only the case if the driver believes he needs to form the relevant intention.[16] Still, it seems phenomenologically clear that in such cases, people often do what they do – take the turning, for instance – because they form the relevant intention. The driver might think something like "OK, there's the right turn, here we go".

Alongside such cases, there importantly also seem to be cases in which an agent φs, believes she has conclusive reasons to φ and yet doesn't φ as a result of intending to φ, because she has no specific intention to φ. We may form a conclusive reasons judgement that picks out some φ that is a component of a composite action that is itself intended, but whose performance is so well-oiled that the component actions don't require any individual intentions. Imagine someone on an advanced drivers' course concurring with the instructor's judgement that you should always look over your shoulder before overtaking, and thinking smugly that that is what they always do automatically.

The fact that the agent doesn't need to intend to perform the action for which she takes herself to have conclusive reasons is not a blemish on her rationality. Being disposed to do such things automatically is actually an excellent way to be organised, as both social psychologists (cf. Sects. 9.5.2 and 9.5.3 below) and Aristotelian virtue ethicists have insisted. Moreover, it seems a fairly clear case of what we would normally see as a rational, not merely reasonable way of being organised. Being thus disposed is efficient, enabling the agent to get the things done she takes herself to have conclusive reasons to do and freeing up further capacities to do other things she wants to do.

If we take the cases of the two drivers as paradigms for a whole set of further examples of rational organisation of agents in the face of their conclusive reasons judgements, we can conclude that the broadest requirement in play when agents make such judgements doesn't only demand coherence between those judgements and the agent's intentions, however formed. Rather, its consequent should specify

[16] A dispositionalist about belief will presumably deny the necessity of any explicit doxastic thoughts concerning the necessity of intention formation. But what might the relevant manifestation conditions of such a belief consist in other than in the formation of the intention itself?

that the agent either intends to bring about the judgement's content or is disposed to do so automatically. I propose the following formulation:

(CRA)
It is rationally required of *X* that:
if *X* judges that she has conclusive reasons to φ in *s*,
X takes herself to be in *s*
and *X* doesn't believe that φ-ing in *s* is impossible for her,
X intends to φ or is disposed to φ automatically.

Three brief comments: First, if I am right that *CRA* is a requirement of rationality, it is a further reason to doubt the claim that rationality supervenes on the mind (cf. Sect. 7.2.1). Whether someone takes it to be a strong reason will depend on whether she thinks that dispositions to automatic behaviour should be classified as mental states. As I take it that they should not, I see *CRA* as speaking strongly against the supervenience thesis.

Second, the inclusion in the consequent of mere dispositions to act may appear implausible as it might seem incompatible with a natural assumption, viz. that being subject to a rational requirement entails having the ability to satisfy it. Dispositions to perform some action on taking myself to be in some situation may simply be absent. Obviously, such dispositions cannot be created spontaneously on (rational) demand. But, even if "rationally required" were to imply "can" in some sense,[17] this would not generate an objection to *CRA*, as what the principle demands, where the proto-doxastic ("takes") and doxastic ("believes") premises are held stable, is either a relevant disposition or a relevant intention. An ability condition might thus be covered by the disjunct picking out the intention.

Third, the requirement concerns φ-ing in a particular situation.[18] This is necessary in order to cover the automatic cases, in which the disposition may be triggered by mere perception of the situation. Moreover, as mere perception, without the formation of a corresponding belief may be sufficient, the second antecedent specifies "taking oneself to be in *s*", which I use to cover both doxastic and merely perceptual "takings". The content of the conclusive reasons judgement here is thus more specific than that named in the first premise of *CRD*, which concerns the performance of an action without reference to a particular situation. It follows that *CRD* is not merely a specification of *CRA*. *CRA* is broader than *CRD* in covering non-decisional cases; it is, however, also narrower, in only covering cases in which the reasons judgement concerns actions in specific situations. Is the lack of derivation here a

[17] I questioned whether any such implication holds in Section 7.2.3.

[18] This feature of the content of the conclusive reasons judgement replaces the temporal feature in *CRD* introduced by the temporally conditional content of the second premise state. Like *CRD*, *CRA* should also be supplemented by a parallel requirement generated by the application of the "perceptual obviousness operator" (Sect. 7.2.2). Here, its effect is to stipulate that the requirement's applicability is conditional on *X* obviously being in the situation specified in the conclusive reasons judgment, whilst cancelling the second, proto-doxastic condition internal to the requirement. Cf. Section 8.5.2, note 31.

problem? It presumably is if we expect the a priori standards of rationality to form not only a coherent, but also a non-overlapping set of architectonically structured principles. However, this seems to me to be an all-too Platonic picture. There may, instead, be various consistent and coherent ways in which the terrain of rationality can be carved up. I tend to think Davidson was right when he claimed that the standards of rationality "can be constituted in various ways" (Davidson 1985, 352).

Return now to the typology of non-decisional intentions, specifically to automatic judgement-derivative cases. If we hold the first three premise states of *CRA* stable, the requirement covers three different ways of being rationally structured, two of which involve intending to φ in s. In the first, the agent has the intention in question because *CRD* guides her behaviour and she decides accordingly. In the second, she intends to φ in s because she is the bearer of subpersonal connections between conclusive reasons judgements and intentions. In the third, she judges she has conclusive reasons to φ and is also, independently, disposed to φ because φ-ing in s is a constitutive component of ψ-ing, where she intends to ψ.

Our interest here concerns the second way of being rationally structured covered by *CRA*. In particular, it concerns the question of whether the intention thus generated without a corresponding decision is plausibly analyzable in motivational terms. *CRA*, which requires agential organisation particularly apt to realise the content of a conclusive reasons judgement, helps us to see that the answer to this question is affirmative. The surest way of coming to φ, if you aren't disposed automatically to φ, is to be predominantly motivated to φ. *CRD* specifies the relations between those agential parameters epistemically accessible in practical deliberation, where forming a decisional intention is the best we can do. Usually it will be enough – either because the agent's motivational dispositions explain the decision or because the decision succeeds in mustering the requisite motivational force (Sect. 8.2.3). Decisionally intending under the relevant conditions is, it seems, a good way to fulfil the function outlined by *CRA*; being preponderantly motivated is generally even better.

9.4.3 Being Set and Having Settled

If, as I am proposing, nondecisionally intending is a matter of being preponderantly motivated to act, it follows that if X nondecisionally intends at t_1 to φ at t_2 and doesn't forget her intention, isn't unaware of the time, isn't prevented from φ-ing and doesn't change her mind, X φs, or at least tries to φ at t_2. The relation between nondecisionally intending to φ and φ-ing requires no rational mediation. In contrast to the decisional case, there can be no nondecisional intentions that are not backed by sufficient motivational force for corresponding action or attempts – absent the defeating conditions just named.

The point can be made clearer by returning to Nosmo (Sect. 8.2.3) and comparing him with his hyper-rational relative, Nosmissimo. Both acquire the intention not to smoke as a result of the judgement that they have conclusive reasons not to smoke.

Nosmo follows up his judgement with a corresponding decision, a mental step that doesn't guarantee that he will realize its content, even if he doesn't change his mind. Nosmissimo, in contrast, had judged earlier in life that he would give up smoking if he ever were to become convinced that it causes cancer. On reading a scientific study which convinces him that the antecedent is satisfied, he finds himself nondecisionally intending to give up smoking. His new attitude can, I submit, only be a motivationally unrivalled want*.

But, it might be asked, is it not conceivable that an attitude described as an intention could be nondeliberatively generated by a corresponding conclusive reasons judgement and yet be motivationally too weak to lead to the realization of its content? That depends on the conceivability test one chooses. It seems to me not unlikely that experimental philosophy polls would show that people on the street don't find such a scenario incoherent. However, people may simply be unaware of the mechanisms that underlie their use of the term "intend" and its cognates.

Compare the objections brought forward by Michael Smith against Moore's open question argument: the fact that "*x* is good, but *x* is not *y*" appears coherent to everyday speakers of the language, whatever one inserts for *y*, doesn't exclude the truth of a definition of *x* in terms of *y*, if the correctness of the definition isn't clear to the speakers (Smith 1994, 36f.). The non-intuitive character of the equivalence might have causes of the sort pinpointed by Kripke: its truth has perhaps not been discovered yet. Alternatively, it may have reasons concerning the economy of everyday thought. If intention is, as I am suggesting, a disjunctive concept, that may be because it lumps together two sorts of processes that are sufficiently similar in most cases so as to be indistinguishable without considerable theoretical effort – an effort that already goes well beyond the effort to focus on one's own intentions. In view of this possibility, the fact that people may see no logical problem in picking out a motivationally weak nondecisionally formed attitude with the term "intention" is not much of an objection.

If, then, such nondecisional intendings are indeed basically a matter of the welling up of motivation, one could say that the relation between the *persistence* and *motivational strength* (Sect. 7.1.2) of nondecisional intentions is different from that at work in their decisional cousins. There is no sense in which, for instance, a spontaneous intention can live on irrationally in its bearer *as* an intention if it is motivationally too weak to be realized in a situation he takes to be the opportunity for its realization. In such cases, insufficient motivational strength entails the end of the attitude's persistence as an intention.

One can certainly make good sense of this: if a person has not deliberated on considerations that in one way or another transcend the immediacy of his present motivational situation, then it is difficult to imagine how anything but the motivational strength of presently activated wants* can guide his action. If the will that can turn out to be weak is indeed a sort of intention,[19] then strict weakness of will – where the action or omission and the willing with which it fails to align are

[19] Cf. Section 8.2.3, notes 8 and 9.

contemporaneous – is impossible for nondecisionally generated episodes of willing, whereas, if the considerations adduced in Section 8.2.3 are valid, this is not the case for willing that results from decision.

My claim, then, is that the possibility of strict weakness of intention (or perhaps "will"), which involves contravening the principle of executive consistency, *EC*,[20] grounds in the fact that the relevant forms of intending are generated by their bearers taking a mental step back – however minimal – from immediate agential involvement with the world. It is such a step, taken in at least minimal deliberation, that enables agents to "set" aims that transcend the purview of their present motivational constitution. The flip side of this capacity for "practical transcendence" is, quite naturally, that there may be times when the attitudes thus generated simply don't possess the motivational clout necessary for their realization in preference to optatitve competitors. In non-decisional cases, the possibilities both of that "transcendence" and, a fortiori, of a corresponding attitudinal inefficacy are quite simply missing.

The point can be put in terms of another metaphor that is often used to focus intuitions in discussions of intending, the metaphor of being "settled on" doing something (Sect. 6.1.2). "Settling" is naturally understood in terms of "putting an end to a dispute" (cf. Nowell-Smith 1957b, 64) and thus presupposes that there was a "dispute" in the first place. But precisely this need not have been the case when you come to intend to greet someone on the street, to reach for the nearest biscuit on the tray or to get up when the alarm goes off. Nondecisional intenders can thus be said to be "set on" doing something they are not "settled on" doing.[21]

Mele (1992a, 162, 167) has argued that intending is a matter of being in an executive state which, if it doesn't involve "settledness" in the sense of "having settled", at least involves what he dubs "thin settledness". "Thin settledness" is supposed to be essentially the kind of state generally caused by an event of "settling", but with that aetiology subtracted.[22] My contention is that there are two different kinds of "executive state", only one of which is informatively characterised in terms of the metaphor of settledness. Where no "settling", in the sense of dispute-resolving, has taken place, it is at least misleading to talk of "settledness", whatever its girth. Where there has been no "settling", there is only room for being motivationally "set".

[20] Holton's intention-based conception of weakness of will contains no provision for this, but rather sees all weakness of will as involving the contravention of standards of intention stability.

[21] Audi (1988, 243f.) makes this point, albeit in the context of the defense of a global definition of intention in motivational terms.

[22] Mele (1992a, 161f.) completes his definition of "thin settledness" by the subtraction of further components that everyday understanding might see as necessary for "settledness": "firmness" (ease of revocation) and (factual) "duration". Velleman's usage, according to which a paradigmatic intention is an "agent's attitude towards outcomes that are settled, from his perspective, at the close of deliberation" (Velleman 1989, 112), accords with my suggestion.

9.5 Nondecisional Intention and Conscious Wanting*

Even if it is true that nondecisional intentions are necessarily motivationally unrivalled, there is little plausibility to the claim that this motivational condition can be sufficient for intending. Indeed, the difference between the cases of Inge and Pablo, on the one hand, and Vic, on the other, was that, although all three are moved to act by the spontaneous welling up of motivation, the actions of the former two, in contrast to that of the latter, are not intentional and a fortiori not intended under any description (Sect. 9.3). The reason, I claimed, is that Inge and Pablo spontaneously go to perform some action on which they have taken no relevant conscious optative stand. In the version of the case in which they haven't had any conscious thought directing their performance of the action, their behaviour can be described as "meaningful" (cf. Vollmer 1993, 324), as it corresponds to and is caused by a want* of theirs. Nevertheless, they will be justified in saying that they "didn't mean to" act as they did.

The proposal I am developing here depends for its plausibility on us being able to distinguish conscious wants* from conscious beliefs about one's wants*. Playing host to a consciously occurrent want* involves taking a conscious stand expressible by sentences of the form "Let it be the case that *p*" (Sect. 4.2.1). That is quite a different matter to believing or otherwise thinking that one is playing host to whatever motivational, hedonic or neurophysiological dispositions might be thought constitutive of desiring. An agent can correctly believe that she is the bearer of a motivationaly unrivalled want* to φ, and yet still not intend to φ.

What is criterially required – for nondecisional as for decisional intentions – is the occurrence of a relevant conscious optative stand on the matter. In order to see why, it will be helpful to return to David Velleman's example of an agent who has, in Velleman's terminology, "unconsciously decided" to sever a relationship to a friend. Even if I am right that talk of deciding should be reserved for the conscious resolution of doubt, Velleman could, so it seems, be justified in stating that in such a case the agent may be the bearer of non-decisional "subconscious intentions", for instance, to alienate his friend (Velleman 1992a, 126). Now, it is plausible that such an agent may, without being aware of it, be pursuing a goal with the content Velleman specifies, a goal constituted as such by a desire to which the agent at the time of acting has no conscious access (cf. Sects. 5.1.4–5.1.6). Nevertheless, in such a case the standard mechanisms of action control are circumvented in a way that deprives the agent of any chance of accepting or rejecting his pursuance of the goal.

Forming an intention is the movement of mind that constitutes such an action-controlling stand of acceptance or rejection. For this reason, intention is, as I shall put it, the *anchor* of responsibility for action. It is true that intending is neither sufficient nor necessary for responsibility: as compelled intention formation is a legitimate excuse and as we are responsible both for the "foreseen" consequences of actions and for our reckless or negligent actions.[23] These distinctions between ways of being responsible are, however, important to us. Responsibility for negligent or reckless behaviour is in a significant sense secondary to responsibility for intended action. Should Velleman's agent come to retrospectively distance himself from the nonconscious goal that guided his action towards his (now ex-)friend, the latter would be more likely to accept an apology if the agent explains that the goal was one of which he hadn't been aware than if he describes his action as having been deliberate. In the light of the agent's regret on realizing that he had been pursuing a goal he hadn't thought about, we might describe his behaviour in terminating the friendship as negligent. There is a clear sense in which the undermining of the friendship would then be less strongly attributable than if he had consciously opted for it.

It will, no doubt, be objected that agents have countless intentions that involve no conscious thoughts. I think this impression – if it is not born of a prior commitment to functionalism – derives from a failure to distinguish various ways in which intentions can explain actions. In what follows, I shall discuss three different kinds of relation of intentions to actions to whose explanation they contribute. These discussions draw on empirical research on nonconscious mechanisms of action control at work in such cases. I shall argue that none of these cases speak against the claim that intentions are necessarily consciously inaugurated: even given the importance of such forms of nonconscious control, there must be a relevant conscious optative stand on the matter if the action is to count as intended.

9.5.1 Proximal Intentions

Now, it might at first glance appear that for an action to be nondecisionally intended, it must be immediately preceded by a consciously occurrent want*.[24] This certainly seems highly plausible for proximal intentions. Spontaneous proximal intentions, such as the intention to take a biscuit offered on a tray ("Mmm, I'll have one of those!") or the intended versions of Vic's, Pablo's and Inge's cases, fit the bill nicely. The same applies to gradual and judgement-derivative proximal intentions. The point is that, if an agent φs in some situation in which she has just acquired the attitude that is responsible for her φ-ing, a lack of any corresponding conscious thought on her part seems to leave her without any control over what she is doing.

[23] For "foreseen" read "accepted". See Roughley 2007c. Note that "acceptance", like intention, but in contrast to negligence, requires a conscious optative stand on the matter for which one takes responsibility.

[24] Compare Goldman: "we are inclined to call acts *un*intentional if there is no corresponding *conscious* desire present" (Goldman 1970, 123). The inclination is important, even if it is not quite right, even for intended actions.

9.5.2 Distal Intentions

Once we turn to distal intentions, however, it becomes clear that there can be no requirement that the relevant conscious optative thought be immediately antecedent. Peter Gollwitzer's research on implementation intentions demonstrates that the control over our actions that we gain by intention formation can be facilitated by, as he puts it, "delegating" that control "to the environment" (Bargh and Gollwitzer 1994; Gollwitzer 1999, 493ff.). Forming conditional intentions with highly specific antecedent clauses ("When the clock strikes 5 pm on Monday, ... "; "As soon as I receive the list of courses, ... ") apparently organizes an agent's action system in such a way that the mere accessibility of the features mentioned in the clause ("situation cues") frequently leads to the performance of the action independently of any further conscious thought concerning the action itself. Indeed, one of the main reasons Gollwitzer's research has been taken up by health care professionals is that this circumvention of conscious control brings a marked increase in the efficacy of positively health-related intentions – concerning, for instance, dieting, doing sport, self-examination for cancer (Gollwitzer and Sheeran 2008).

If this is correct, then distal intentions formed consciously with some particular situation in mind require neither conscious temporal updating nor any other form of conscious re-tokening in order to guide their bearer's actions appropriately. However, it might appear that the results of these experiments are irrelevant for our present topic, as the participants are instructed to take the explicit step of drawing up their conditional intentions, a step which generally involves choosing both the conditions under which they are to realize their more general goal and the way of realizing it. It may thus appear that implementation intentions are necessarily of the decisional variety.

This is certainly true where the conscious intention-formative thought is preceded by the selection between options. The question is whether implementation intentions are necessarily formed as a result of such selection.[25] As far as I can see, this question is not asked explicitly in the literature. Where definition-like characterizations are provided, they make no reference to such selection processes, but focus instead on the "if-then structure", also generally mentioning the purpose of their formation and their efficacy in leading to action (Gollwitzer 1999, 494; Gollwitzer and Sheeran 2008; Achtziger et al. 2008, 381f.). However, certain experimental designs effectively provide the participants with the precise wording with which they are to fill in the if-then structure. For instance, in one experiment, participants, having identified in a questionnaire the high-fat food they have eaten most of in the last week (*F*), were then instructed to form the implementation intention with the pre-given wording "If I think about *F*, I will ignore the thought" (Achtziger et al. 2008, 384).

[25] I raised a similar question with respect to the transition to the "implemental mind set" and its effects. Cf. Section 6.4.1.

Independently of how implementation intentions are defined, the explanatory question that is decisive both for health care professionals and for our concerns here asks after the mechanism that makes these intentions so efficacious. Gollwitzer's explanation is that their formation leads to a "heightened activation of the situational cues", making their representation significantly more accessible to perceptual, attention and memory processes (Gollwitzer 1999, 497f.). What we would need to know would be whether the strengthening of the causal connections between semantic or perceptual features of the situation and the intended behaviour can get by without choice. The example cited indicates that this is the case. Indeed, the way the example continues suggests that other mechanisms, such as repeatedly rehearsing the intention's content, may equally work to strengthen the connections even in the face of temptation. Moreover, if the decisive feature of implementation intentions is an explicit wording that includes a conditional clause (and is not merely logically equivalent to some such wording[26]) and if an agent is aware of this, then she may simply specify some intention she already has, employing the requisite formal structure. Forming certain implementation intentions could then be a matter of forming nondecisional specificatory intentions. Where implementation intentions are generated by merely specifying the content of a more general intention ("goal intention") and where that intention leads an agent automatically to realise its content on finding herself in the situation the content specifies, we have a case of non-decisional distal intending in which the intended action is initiated non-consciously.

There is another interesting consequence of these experiments for the topic of nondecisional intentions. They indicate that there is a sixth kind to be added to the typology: if someone is in the right kind of motivational state, i.e. is prepared to accept and act on instructions from another agent without further reflection, then instructions or orders of the first person may be nondeliberatively translated one-to-one into intentions of the second – at least in those cases in which the participants don't have to make choices as to how to fill in any gaps. We could talk here of *prescriptive* intention generation. In organizations that are structured by rigid and unquestionable hierarchy (armed forces) and in contexts in which the recipients of instructions assume those instructions will be unproblematic (psychological experiments), unthinking obedience in the formation and carrying out of intentions is presumably the norm. Participants in the relevant contexts enter them with the unrivalled motivation to do what the authority figure prescribes – up to a certain point –, a source of motivation which ceteris paribus transfers to the optative attitudes with contents corresponding to those prescribed. Where the normal flow from prescription to intention content is disturbed – for instance, by doubts as to whether what is being prescribed is OK, as in the Milgram experiment –, any intentions then formed are deliberative and thus decisional.

[26] There is, apparently, a dramatic difference between the effect of intending to φ on Wednesday at *n* o'clock and intending to φ "if it is *n* o'clock on Wednesday" (Gollwitzer and Sheeran 2008, 12f.).

It seems, then, that at least some of the intentions formed by the participants in the experiments are nondecisional. Others, where the most suitable way of realizing a "goal intention" has to be chosen, are clearly decisional. The experiments show that, when they are effective, conditional intentions formed in either way can do their work without *further* conscious thought. Nevertheless, in either case, intention formation requires an initial conscious attitudinal stand.[27]

9.5.3 Intentions to Perform Habitual Actions

There is a further challenge to the claim that intending to φ in s requires having consciously taken a corresponding optative stand. The challenge is posed by habitual action. Habits involve more comprehensive forms of automaticity than the "strategic" variant (Gollwitzer and Schaal 1998) inaugurated by implementation intentions. Like implementation intentions, habits also involve acting as a result of having "delegated" control of our actions to our environment. Here, however, that delegation is not deliberate, but results from repeated intentional performances (Bargh and Gollwitzer 1994). The explanatory claim is that repeated realization of a goal type in contexts distinguished by particular kinds of cues leads to the establishment of subpersonal connections in the agent's system of action control, connections that are activated by the relevant cues, nonconsciously initiating and controlling the agent's behaviour. No-one needs a conscious optative stand on their depressing the clutch and accelerator in order to set their car in motion again once the traffic lights turn green. Nevertheless, the car driver is surely pursuing a goal in doing so.[28]

It seems considerably less plausible than in cases of conditional distal intentions that earlier conscious wants* might confer a goal on a later automatic performance of a routine action.[29] Here it would have to be the repeated tokening of wants* of a certain type which conferred the corresponding goal on later nonconsciously performed tokens of an action of the same type as those in the original series. But spare a thought for the bus driver who, when taking his family out for an excursion in the car, finds himself repeatedly pulling in and stopping at bus stops along the way (Norman 1981; Bargh and Barndollar 1996, 458ff.). That he does so is well explained by the genesis of subpersonal connections through repeated

[27] In Gollwitzer's words, "an act of will" is necessary (Gollwitzer and Sheeran 2008).

[28] This is compatible with the claim in Section 9.2, where I argued against the suggestion that there may not be any automatically generated intentions, as all habitual actions may seem to be explained by some other mechanisms. I claimed that there are at least *some* cases of habitual action that we plausibly take to be controlled by conscious intentions, citing as evidence examples in which agents expend conscious effort to achieve their habitually generated aim.

[29] Something along these lines was suggested by Larry Wright (Wright 1976, 127ff.). I am grateful to Ezio Di Nucci for this reference.

performances. But what is his goal supposed to be in doing so? It can't be to let out bus passengers, as he knows perfectly well that he has none in his car. He certainly need not have forgotten that, indeed he might be in the middle of a conversation with his family as he pulls in.[30] The characterization of such cases as "action slips" (Norman 1981) or "capture errors" (Reason 1979) is appropriate because something has gone wrong. That something is surely well described as the person's behaviour ceasing to be guided by his goals.[31] The aetiological conception cannot explain this.

What is more plausible is that the relevant conscious stand has as its content the composite action of which the automatically triggered action is a routine component part. Habitual or skilled actions are generally seen by their agent as parts of larger intended actions, for instance going to work. People tend to have routines that help them get on the way in the morning without them needing to spend too much time thinking about what to do and in what order to do it. They do all these things, so it seems, because they intend to go to work. There are thus a whole set of component actions that together make up the composite action of going to work. Routine actions are, then, normally parts of composite actions, where composite actions consist to a large extent of sequentially organized subordinate actions.[32] The relation of "consisting of" is no necessary relation, as different subordinate actions can together make up the same composite action: I can go to work by car, bike or underground; and by one of various possible routes. Actions subordinate to composite actions nevertheless make up, or are ways of realizing the superordinate action.

In such cases, I think we should be saying, it is the motivationally backed conscious optative stand on the composite action that secures the agent's control over its component actions and thus bestows goal-directedness on them.[33] However, as

[30] The resistance of habits to change through consciously formed intentions is one of the main reasons the authors mentioned in note 7, above, give for seeing habitual action as not goal-directed at all. However, the occurrence of unintentional versions of a certain form of behaviour doesn't invalidate the ascription of goals in cases that we see as intentional. Indeed, we need some criterion to distinguish the two kinds of case.

[31] Bargh and Barndollar reject this – commonsensical – idea, suggesting instead that the disposition to behave in such ways under repeated environmental conditions should be seen as constituting the having of a goal, indeed of intending: "To our minds, the unconscious intention is just as 'intentional' as ... the momentary conscious goal" (Bargh and Barndollar 1996, 465). Note the difference from Velleman's agent who breaks off a friendship without having been aware of the reason for his action (Sects. 9.4.2 and 9.5). This latter person is legitimately described as pursuing a goal, even though the lack of conscious access to that goal prevents us seeing him as intending the goal's content.

[32] Composite action descriptions differ from complex action descriptions, where the latter are descriptions of one and the same event or process in terms of different properties it instantiates. Composite action descriptions tie together descriptions of temporally distinct actions. Although I tend to think that the coarse-grained approach to action individuation exemplified in the first two sentences of this footnote is correct, I will nevertheless adopt an idiom more appropriate to fine-grained individuation in the main body of the text – talking of complex and composite actions – merely in order to avoid the somewhat tortuous syntax required by insisting that the distinctions in question are at the level of action descriptions (cf. Sect. 3.2.1).

[33] A solution along these lines is suggested by Mele (Mele 1992a, 113f.).

the conscious stand does not have the component actions as its content individually, those actions are plausibly not intended. A full account of the goal-directedness of component actions would have to make clear what conditions secure that feature of the relevant segments of behaviour, as we certainly shouldn't see every segment of a composite action, indeed of any action, as goal-directed.

Here, in Section 9.5, I have outlined a coherent conception of the criterial centrality of conscious optative attitudinizing for nondecisional intending, which seems to accord sufficiently with our everyday intuitions on these matters. These seem, particularly in relation to habitual composite actions, to be fairly indeterminate. The account explains why we are right to see certain automatically performed actions as distally intended and certain routine components of composite actions as goal-directed, but not intended.[34]

9.6 Leaving the Question Open

Unrivalled motivation to φ is insufficient for intending to φ. A relevant conscious optative occurrence is also required. However, the conjunction of these two conditions is still insufficient, as the intervention of deliberation on the agent's part will, as with decisional intentions, deprive the want* of its contextually unique practical status. The agent may thus suspend commitments relative to φ-ing or settle on not φ-ing. But motivationally backed conscious optative stands are not only disqualified from counting as intentions where an agent has begun to deliberate. There are, so it seems, also cases of noncommittal unrivalled motivation in which no deliberation has been initiated. Remember the person who, relative to the coming Saturday evening, is most strongly motivated to go and see a certain play, but who, because she believes other options might crop up that she may want to consider, is not (yet) settled on going to the play (Sect. 8.2.1). For this person, the question of whether she's going to go and see the play is still open, even though she has also taken a positive conscious optative stand on going to see it.

If the disjunctive theory cannot explain why our theatre-goer doesn't (yet) intend in consonance with her conscious and motivationally backed want*, then it clearly fails. This is core territory of the irreducibility theorist: his answer is that the unanalysable step of committing oneself has not taken place. Does a motivationally

[34] Things can get complicated. Take the story told of Hilbert, who, on being asked by his wife to change his tie before the dinner guests arrive, not only takes off his tie, but ends up undressing completely and going to bed, where his wife later finds him asleep (Heckhausen and Beckmann 1990, 41). Hilbert plausibly intends to change his tie, but because his absent-mindedness allows his tie changing to trigger his going-to-bed routine, goes to bed unintentionally. It is quite possible that, in the middle of his absent-mindedly undressing, he encounters a problem in untying his shoes, a problem that triggers conscious thoughts concerning the goal of disentangling his shoelaces. In this case, he may intend to untie his shoelaces as a means to performing the composite action of going to bed that itself is completely unintended. Cf. Roughley 2007a.

based theory have the resources to deal with this problem? Positively put: is there something more informative that can be said about what it means for such a question to be "left open"?

I think there is. It grounds in the idea that human agents are generally disposed to see the contents of their conscious action wants*, particularly where those wants* possess significant motivational clout, as matters up for deliberative grabs. This disposition results from the acquisition of certain kinds of knowledge about the structure of agency, a development we can think of as part of becoming a full agent. We come to realize that our present optative constellation may alter with time: that the strength of some of our wants* can change and, above all, that we often acquire new wants* as our situation changes. We also develop the awareness that some of the things we want* now, even desire passionately, may be things that are axiologically or probabilistically problematic.

The development of these forms of understanding and foresight is naturally accompanied by the development of the dispositional want* to think through those want* contents we consciously token and see as potentially significant for our future action. Indeed, such a development is not only a natural accompaniment; it is also, perhaps even primarily, the consequence of the socialisation into a culture that demands such circumspection. An agent's consciously tokening at t_1 the want* to φ at t_2 will tend to trigger her dispositional want* to deliberate on whether to φ – at least where her φ-ing is not an obviously trivial matter.

Sometimes, so it seems, no such want* is triggered. Where it is, there are three possible reactions. The agent can, *firstly*, override or simply ignore her deliberative want*. She may do this because the triggering want's* content appears so obviously unproblematic. In such cases, the dismissal of the deliberative option may take place so quickly as to be hardly noticeable and may thus be phenomenologically indistinguishable from the case in which the deliberative want* is simply not triggered. It may, however, also be the result of the agent's thought that she has no time to think the matter through. Another reason might be her sensing that entering into deliberation could endanger her doing something she very strongly wants to do. In such cases there may well be mechanisms at work that require no more than the occurrence of unpleasant sensations at the thought of deliberation for it to be rejected (cf. Sect. 5.1.6). A *second* reaction to the triggering of the deliberative want* consists in the agent entering into deliberation. A *third* possibility is that she postpone, without rejecting, deliberation on the matter.

What I want to suggest is that this typology enables us to map the cases in which we would intuitively see the question of whether to φ as being "open" or "closed" for the agent. The claim is that, if either no deliberative want* is triggered or that want* is ignored or overridden, a positive conscious optative stand of the agent's on her φ-ing, where this is backed by unrivalled motivation, is sufficient for her intending. In contrast, either the second or the third type of reaction to the triggering of deliberative motivation involves cancelling the want's* status as intention-constitutive. Either entering into deliberation on the content of an action want* or coming to see that content as a matter for deliberation involves taking up a kind of distance to the action want's* content that prevents it from counting as an intention.

This way of construing things involves rejecting an assumption of the irreducibility theorist that seems natural. The assumption is that what has to be explained is the “closing of the question” for the agent of whether to φ or not. However, what this natural assumption ignores is the need for an explanation of there being a question there for the agent in the first place. I am proposing that the question arises when the agent herself “opens” it. This happens when she comes to see her φ-ing as a matter for deliberation. But should an agent who fulfils the optative and motivational conditions happen to have no thoughts on the matter, she has not “opened” the question. Where this is the case, she is, by default, the bearer of a non-decisional intention. Thus, to stick with the metaphor of open and closed questions, we don’t only have two, but three kinds of case: the question can be closed for the agent, if she has come to a decision; it can be open for the agent if she either sees it as a matter for deliberation or has begun thinking about it; or there may be no question for the agent at all. This will be so if she has not opened it, either because she ignored the desire to deliberate or because no such want* was triggered.

A footballer who deliberately commits a foul, someone spontaneously greeting a friend on the street or someone reaching for his glass at the bar may all be understood as doing what they do because they are *set on* doing it. But this may only mean that they are at that moment *motivationally organised* in such a way that this is what they are led to do. It would then be misleading to say that for them the question of what they are to do is “closed”, as it was never in the relevant sense “open” in the first place. The theatre-goer mentioned at the beginning of this section is presumably in a different condition. If the question is open for her, that is because she sees it as a matter that is still up for deliberative grabs.[35]

An agent leaves the question open as to whether to φ in spite of his playing host to a consciously occurrent want* to φ, backed by unrivalled motivational force, if he sees his φ-ing as a matter for deliberation and has not reached a decision on the matter. This solution ties in with the claim I advanced as to what makes the content of a deliberation-terminative want* the deliberator’s own answer to the question to which the deliberative episode is directed (Sect. 8.7.2). Deliberation on an issue is completed not simply when the agent actually stops reflecting, but where a deliberation-terminative optative stand taken by the agent is accompanied by unrivalled motivation not to reenter deliberation. Closing a deliberative question involves both a conscious optative stand and unrivalled

[35] The component of the analysis that provides the solution here is related to a suggestion of Michael Ridge (1998, 163ff.). Ridge analyses “commitment” in terms of unrivalled motivation to act, in conjunction with the “desire” not to deliberate. What I am suggesting need be conjoined with unrivalled motivation in nondecisional cases is either the non-triggering of the dispositional want* to deliberate or the triggering of a disposition not to realize that want* when it is triggered. I don’t think we need specify whether the disposition not to realize the deliberative want* itself need be motivational or merely behavioural. Although I agree with many of Ridge’s arguments for a reductive analysis of intention, further central differences between our views concern my claim that intention is a disjunctive concept and my related views on the roles of decision and conscious optative occurrences. Ridge’s purely motivational interpretation of intention commits him to seek strategies to explain away phenomena such as weakness of will and countermotivational decision (cf. Ridge 1998, 168ff.).

deliberation-aversive motivation relative to the deliberative issue. The bearer of a conscious optative stand backed by unrivalled motivation doesn't open the question as to whether to realize the content of his want* if he doesn't come to see it as a component of a deliberative issue.

What this shows is that intending, of whatever ilk, is essentially related to practical deliberation.[36] This goes a long way towards vindicating an "upstreamist" conception of intention, which gives conceptual priority to intention's genesis relative to the downstream features focused on by functionalist and normative-functionalist conceptions. The idea is that the commitment constitutive of intending, whether in the decisional or non-decisional variant, is established by the agent in a way that requires her conscious participation. The step taken by the agent establishes "where she stands" on a particular practical issue. Because human persons are essentially deliberative creatures, establishing where they stand, even in some minor, contextually specific matter, results from the way they use their deliberative capacities. This is true both when they actively employ them and when they sidestep them by not entering into deliberation about whether to act in a way the probability of which is raised by goings on in their motivational system. Whether such sidestepping is a matter of instantaneously opting at the prompting of a deliberative want* or whether the agent is simply disposed not to deliberate when certain sorts of conscious action wants* are triggered, either way, refraining from deliberation expresses towards the matter at hand a practical perspective that is, at that moment, the agent's own.

This conception of intending has consequences for agents without the capacity for deliberation. Many non-human animals and small children don't have such capacities. Agents of whom this is true can neither be bearers of "open questions" as to what to do in the future, nor can they close them by settling on some particular action. They can, however, plausibly be set to do certain things, in as far as they are the bearers of wants* and beliefs. And those things they are set to do can be sensibly described as their "goals". To what extent their goals correspond to consciously tokened wants* will, of course, hardly be determinable. Theoretically, we could either say that they are constitutively unable to intend or describe them as always intending the objects of those conscious wants* they are most motivated to realise. Taking the latter course involves describing them in the light of our own capacities. We do sometimes talk this way, and in the case of small children, it makes a lot of sense to just that. They are on the way to becoming like us and we don't know at what point precisely their deliberative capacities set in. They are non-paradigmatic intenders or, if you prefer, proto-intenders.

[36] When, for instance, Gilbert Harman describes intentions as attitudes concerning ones future action "arrived at and maintained by practical reasoning" (Harman 1975/76, 451; 1986b, 375), he is focusing too strongly on the paradigmatic case. Such a claim excludes the kind of important cases covered by Audi's and Ridge's equally one-sided motivational conceptions.

9.7 Intentions, Decisional and Nondecisional

In order to conclude the chapter by pressing the results of the foregoing discussion into a definition, we need clarity on the relationship between the two types of intentions. If I am right that it is possible to decisionally intend not to φ, in spite of being motivationally set on φ-ing, and if nondecisionally intending is at core a matter of being motivationally set, there are potentially situations in which agents are the bearers of both types of intention where the content of the one negates the content of the other. Such a conflict is avoided by a lexical ordering of the conditions, an ordering that grounds in the significance of practical deliberation. Because where a person stands on a practical issue is essentially a matter of her taking on an optative attitude in the light of her deliberative capacities, the overriding mechanism of taking such a stand involves the active use of those capacities. Where those capacities are set in motion or where their use is merely envisaged, earlier postures on the matter are cancelled out, whether these are paradigmatic intentions, themselves formed as a result of practical deliberation or whether they are of the secondary kind, formed by sidestepping deliberation on the content of a motivationally unrivalled conscious want.*

I submit, then, that the concept of intention can be defined as follows:

(ID)
X intends to perform action φ iff:
1. *X* wants* to φ,
2. *X* doesn't see his φ-ing as a matter for at least minimal deliberation
and
3.1. if *X* has minimally deliberated on whether to φ, he has, in the course of his most recent minimal deliberation on the issue, decided to φ,
or
3.2. if *X* hasn't engaged in any such minimal deliberation, his want* to φ has been consciously tokened and is motivationally unrivalled.

Now, some philosophers may worry about the disjunctive feature of this definition. A disjunctive relation, it might be thought, doesn't establish the unity necessary to justify talk of one thing. An analysis that sees something as being an instance of a type of entity because it instantiates either property *x* or property *y* may appear to be working with a conception of a thing's identity that would allow jadeite and nephrite to count as the same thing (cf. Schroeder 2007, 69).

I want to say two things about this worry. The first concerns the type of conditions under which an attitude counts as an intention. None of them are intrinsic conditions. Rather, they are, with the exception of the second conjoin of the second disjunct, all relations to the content of the want* that, under the relevant conditions, *is* the intention. This means that intentions are not good candidates for the status of natural kinds of mental states, however one might understand this phrase in a psychological context (cf. Griffiths 1997, 6). In this, they may be thought to contrast with certain basic emotions and perhaps with the basic attitudes of believing and wanting*, which may turn out to be inherent and innate features of human psychology. What makes an optative attitude an intention is primarily a want's* coming to be the object of further psychological features: a status that can only come about through the development of further psychological capacities, in particular the capacity for practical deliberation. In contrast to what modularity theorists tend to believe, a great deal of the structure of our mental life plausibly develops in interaction with features of our overall life form, in particular with the linguistic and social practices with which we are confronted in development (cf. Astington 1996).

Second, there are certain parallels between my proposal and the analyses that go under the heading of "disjunctivism" in other areas. In Hornsby's conception of reasons for action and McDowell's conception of epistemic experience, there is a secondary variant of the relevant concept that only qualifies as such because there is a primary variant that gives the concept its cogency. According to Hornsby, acting for reasons involves acting out of knowledge that *p* or out of the mere belief that *p* (Hornsby 2008, 252). The latter kind of case, she claims, would be insufficient to generate the concept of acting for reasons. It is thus inconceivable that all cases of acting for reasons could have the structure picked out by the second disjunct (ebd., 258). My claim is similarly that the idea of intending could not be generated simply from cases of consciously tokened, motivationally unrivalled wanting*, i.e. instances of 3.2. Such cases are co-opted as part of the concept as surrogate versions that can stand in where the full deliberative apparatus is not available or appropriate. These parallels should, I think, help to dispel worries about the claim that a basic mental concept can legitimately be given a disjunctive analysis.

There are, of course, also important disanalogies between the disjunctivist conceptions just cited and the conception of intending formulated in *ID*. The most important is that in both Hornsby's analysis of reasons for action and McDowell's analysis of epistemic experience, the secondary cases involve a cognitive deficiency on the part of the agent (Haddock and Macpherson 2008, 4ff.). As intending is, according to *ID*, not a cognitive matter, no such deficiency can be in play here. The secondary cases of intending picked out by *ID* are less demanding, requiring less explicit thought on the part of their bearer. However, they are in no sense deficient relative to the primary, paradigmatic cases.

Finally, it is, I think, helpful to see that there is a developmental story, supported by much of the empirical psychological literature, that dovetails nicely with the analysis proposed in *ID*. Janet Wilde Astington has argued that there is something "paradoxical" about the concept of intention, as it has seemed on the one hand to be the attitude most easily ascribed to young children and yet on the other hand to have components that make its ascription more demanding than that of belief (Astington 2001, 85). There seems to be a sense in which we might think of 3-year-old children as expressing intentions when they say that they "wanna" or they're "gonna" do something (Astington 1999, 305, 1991, 167f.). Preschoolers also begin to use terms such as "try" and "on purpose" (Astington 1999, 307). Moreover, already 18-month-old infants have been shown to impute goals that adults appear to be trying to achieve: in experimental conditions they are able to imitate a complete action, when adults only demonstrate a failed attempt (Meltzoff 1995) and they frequently help adults to attain goals that obstacles or apparent lack of skill have prevented the adults from attaining (Warneken and Tomasello 2006).

What is plausibly expressed or imputed in these early cases is, as the psychologist Louis Moses puts it, "an intention/desire notion that is closely wedded to action" (Moses 2001, 78; cf. Bartsch and Wellman 1995, 68). A reliable and systematic understanding of the terms "intends to", "means to" "plans to" and "on purpose" was found by Astington to not be forthcoming until the age of 7 (Astington 1999, 309; cf. Schult 2002, 1744ff.). She has therefore wondered whether it isn't "unwise to use the same term to refer to the early and the later understanding" (Astington 2001, 85), a question that seems particularly pertinent if you assume, as Moses does, that a later conception of intention "requires an appreciation of humans as actively constructing their mental agendas, a conception of psychological functioning that is unlikely to emerge before the close of the pre-school years" (Moses 2001, 82).

Astington's terminological worry is, I think, highly pertinent. Intentions, we can say, fix goals or purposes, but not all goals or purposes correspond to intentions. *ID* provides criteria for classifying some goals formed by adult humans as intentions. Those goals of young children that satisfy condition 3.2 can also be thus classified if we view them from within the framework of adult agency. Up to a certain point, that is something we do with children, here as in other areas ascribing them mental states before the capacity for those states has fully crystallized. What is strictly speaking a category mistake looks to be a highly desirable feature of the way we socialize children. Ascribing them intentions prior to the development of the capacity for deliberation helps bootstrap their mental functioning into the structures of adult social interaction (cf. Gibbs 2001, 120).[37]

[37] Against Searle's claim that intentions have self-referential contents, Mele argues that the claim would prevent us assigning intentions to 8-month-old children (Mele 1992a, 204). This is true, but, for the reasons just given, no objection.

Chapter 10
The Intention-Consequential Requirements and Anchoring Attributability

I have argued that intending is essentially related to practical deliberation. Particular wants*, I have claimed, are given a contextually unique practical status by functioning as deliberation-terminative or by being allowed direct access to the agent's system of action control in spite of the possibility of deliberative mediation. Taking the relevant kind of stand or effectively opting to leave things unreflectively in the hands of prior motivation are the ways in which agents themselves exert direct control over their own action. It is in these ways, one could say, that agents constitute who they are practically: this is how they determine "where they stand" on practical questions. This "upstreamist" approach, which ties intending conceptually to one of two sorts of aetiology, has consequences for a number of important issues that surround the notion of intention.[1] In this last chapter, I shall argue that it provides a distinctive, and distinctively plausible explanation of the requirements of practical rationality.

The explanation contrasts with conceptions that make use of what are claimed to be necessary doxastic features of intending, that is, with cognitivist explanations of the *IC* requirements. Having argued in Section 6.3 against strong cognitivist understandings of intention, according to which intending is or entails believing that one will realise its content, I briefly discuss a weaker, prima facie more plausible form of cognitivism, that of Jay Wallace, noting in particular the use Wallace feels constrained to make of the distinction between deliberative and non-deliberative contexts, in order to make his explanation stick. One important reason why it doesn't stick is that, unlike what Wallace claims, the *IC* requirements are applicable to both deliberatively and non-deliberatively acquired intentions (Sect. 10.1).

My non-cognitivist proposal has important similarities to that advanced by Bratman. I discuss Bratman's model in some detail, as it is, as far as I am aware, at present the most elaborated attempt to provide a non-cognitive explanation of the

[1] For its consequences for two issues I don't take up here, the doctrine of double effect and intentional action, see Roughley 2007c and Roughley unpublished c.

N. Roughley, *Wanting and Intending*, Philosophical Studies Series 123,
DOI 10.1007/978-94-017-7387-4_10

requirements and because a presentation of both its strengths and weaknesses helps me to clarify and justify the precise contours of my own model.

Bratman's proposal first needs a little hermeneutic unpacking, which I do in Section 10.2. He thinks that the source of the requirements' force is our reason for what he calls self-governance, because conformity to the requirements is a constitutive condition of such self-governance. I go on to claim, first, that self-governance cannot be the source Bratman takes it to be, adducing three reasons why not (Sect. 10.3.1). Second, I argue that the Bratmanian normative functionalist understanding of intention erects a fundamental barrier to the provision of a non-circular answer to the explanatory question (Sect. 10.3.2).

Clarity on the problems with Bratman's conception helps pave the way for my own suggestion, according to which the source of the *IC* requirements is not a particularly strong form of agency that we perhaps all have reason to pursue, but rather intentional agency itself.

Intentional agency, I claim, grounds in our capacity for deliberation and decision, where decision is, as I have argued in the last two chapters, the paradigmatic genesis of intention. Drawing on support from findings of social and developmental psychology, I suggest that deliberative intention formation plausibly develops at least in part in response to social pressure to conform to norms, in the face of which children learn to *take personal responsibility* for what they do. In terms adapted from Gary Watson, I argue that to postdeliberatively opt for certain actions is to provide an anchor for their attributability and that it is a concern to provide such an anchor that has conferred on intention the shape it has according to the analysis offered in *ID* (Sects. 10.4.1 and 10.4.2).

The reason for intending's subjection to the *IC* requirements, I then argue, is that their contravention renders responsibility taking unintelligible. As the taking of responsibility in the sense developed here is the core of intentional agency, an agent who intends in violation of the *IC* requirements becomes opaque to himself, thus necessarily losing his hold on his own agency. This fact, I argue in Section 10.5.1, provides agents with a distinct kind of reason, one that is both stringent and under certain circumstances strikingly weak. This conjunction of features, I claim, suffices as an answer to the scepticism expressed by Broome as to whether rationality can be shown to generate reasons.

The disjunctive structure of my theory of intending requires that the explanation be shown to be applicable beyond deliberative cases. Accordingly, I argue in Section 10.5.2 that the notion of taking personal responsibility can be extended to cover cases of nondecisional intending. I claim that we conceive nondecisional intention as involving the exercise of the same capacity explicitly and concurrently exercised in the decisional case. Importantly, the reason we conceive it as fulfilling this role is at least in part normative. It grounds in the demand that persons deliberate on whether they are willing to see themselves as the realisers of their motivationally unrivalled wants*, deliberation that need not be concurrent with the conscious tokening of those wants*, but may have to concern dispositional wants* in advance. It is because of this normative background, in which our intentional agency is deeply entrenched, that nondeliberative implicit acceptance of the action controlling tendency of a conscious want* counts as responsibility taking and is thus equally subject to the *IC* requirements.

It turns out, then, that, although intention is a descriptive concept, the subjection of the full range of its decisional and non-decisional variants to the *IC* requirements presupposes a normative component in the understanding of their aptness for the role that explains that subjection. Indeed, intention is so strongly interwoven with our culture of normative address that it seems extremely unlikely that a creature without such a normative life form might develop a concept with anything like the contours that intention has for us.

10.1 Intention Noncognitivism and the *IC* Requirements

The question as to how we might explain the intention-consequential norms of practical rationality has been broached by a large number of authors.[2] In what follows, I develop an explanation that grounds in the upstreamist, disjunctive, optative conception of intending for which I have argued.

Various authors have claimed that the only way to understand the norms of practical rationality is to see them as variants or derivates of the norms of theoretical rationality.[3] This strategy can get going if intending to φ entails or is equivalent to the belief that one will φ or perhaps if, less strongly, intending to φ entails the belief that one can φ (i.e. if one of the doxastic conditions labelled *B1* and *B3* or *BM* in Section 6.3.2 is valid). Such views may seem to promise unitary explanations of why attitudinal conjunctions such as the following count as central cases of practical irrationality:

(i) intending to φ and intending not to φ;
(ii) intending to φ and not intending to ψ, whilst believing that your ψ-ing is a necessary means of your φ-ing;
(iii) intending to φ and intending to ψ, whilst believing that your φ-ing requires your not ψ -ing.

If we take it that intending entails believing you will realise the content intended, we get (i) beliefs in contradictory contents, (ii) three incoherent beliefs, which together specify that the believer won't do something she needs to do in order to do something she will do, and (iii) three further incoherent beliefs, which together specify that the agent will do something which will prevent her doing something she is going to do.

[2] Harman 1975/76, 1986a; Korsgaard 1997; Raz 1999; Wallace 2001; Scanlon 2004; 2007; Kolodny 2005; Setiya 2007; Velleman 2007; Cullity 2008; Broome 2007a; 2008; 2009; Schroeder 2009; Bratman 2009a; 2009b; 2009c.

[3] Wallace 2001, 104ff.; Setiya 2007, 663ff.; Velleman 2007, 204ff.; Broome 2009, 77ff.; Schroeder 2009, 236ff.

If intending is taken to have a weaker doxastic condition concerning, for instance, only the agent's ability to perform the action intended, demonstrating the implied doxastic incoherence is considerably less straightforward. There is, then, certainly nothing incoherent about any of the conjunctions of doxastic states entailed by (i), (ii) and (iii). Believing that you can φ and believing that you can omit φ (i) is a not unusual condition prior to any practical decision. Believing that your ψ-ing is a necessary means of your φ-ing, whilst believing that you can φ and having no particular beliefs as to whether you can ψ (ii) is a perfectly satisfactory doxastic state of affairs, as is the belief that your φ-ing requires your not ψ–ing in conjunction with the further beliefs that you can φ and that you can ψ (iii).

One modification that has been taken to help solve the problem for cases along the lines of (ii) involves replacing the agent's belief in the necessity of her ψ-ing as a means to her φ-ing with a belief concerning the functioning of her own intention, that is, with a belief in the necessity of her intending to ψ in order for her to φ. Wallace has argued that the constellation of beliefs that this generates is incoherent (Wallace 2001, 105f.). The agent, he claims, now believes that her intending to ψ is a necessary means of her φ-ing, whilst also believing that she can φ and believing that she doesn't intend to ψ.

This set of beliefs is indeed incoherent. However, it is not generated merely by the move from a belief concerning the agent's action to a belief concerning her intention. It also requires the additional move from the fact that she doesn't intend to ψ to the belief that she doesn't intend to ψ, i.e. a second move to a higher-order attitude. Wallace claims that this move is unproblematic, as it will be taken by any agent who is "minimally self-aware". That level and type of self-awareness, he goes on to say, can be presupposed in someone who is practically deliberating (Wallace 2006a, 118). This last fact is in turn decisive, he thinks, because what he calls "the instrumental principle" is only in play (Wallace actually says "relevant") in deliberative contexts.

It is, first, noteworthy that such an attempt to explain the force of the *IC* requirements in terms of their doxastic implications also sees itself constrained to avert to further consequences of a non-logical nature.

Second, Wallace admits that those consequences are only likely to occur in the restricted context of the practical deliberation of a minimally competent deliberator: in that context, we can expect an agent with that minimal competence to keep track of whether she has formed relevant intentions, something we cannot reasonably expect in non-deliberative contexts. That is plausible. What is implausible is the claim that this restriction fits a restriction on the applicability of the requirement of subordinate intention formation. An agent who has non-decisionally – or decisionally – formed an intention to do something on that day and then in the course of the day fails to form a subordinate intention to bring about some state of affairs she sees as a necessary means to her intended end is behaving irrationally, independently of whether she sits down and thinks the matter through. The important distinction between decisional and non-decisional intention, which Wallace picks up on here, is importantly not a difference in the subjection of intention's different species to the *IC* requirement I have labelled *SI*. This point alone provides a sufficient reason to reject Wallace's strategy.

A third point of importance is that, even if this explanation of the force of *SI* were to be plausible, there appears to be little hope of extending the strategy to the other intention-consequential principles.

Finally, the strategy of substituting a second-order belief about the necessity of one's intending for a first-order belief about the necessity of one's action already involves, as I remarked in Section 7.2.2 with respect to Broome's similar move, a further significant restriction on the scope of a principle of intention subordination, a restriction I argued that we don't need to impose.

There are arguments that are independent of the *IC* requirements which speak against either identifying intention with some cognitive attitude or even taking it to involve any positive belief condition. I developed such arguments in Chapter 6 (Sects. 6.3.1 and 6.3.2), looking in particular at problems with the stronger proposals of Anscombe, Hampshire and Velleman. We have just seen that the weaker proposal advanced by Wallace, which seems phenomenologically more plausible, is unable to provide an explanation of the force of the requirements.

The claim that intimate conceptual connections between intending and believing generate norms of practical rationality as derivations of the norms of theoretical rationality has been helpfully characterised by Bratman as "cognitivism" about intention rationality (Bratman 1999a, 250f.; 2009a, 30; 2009b, 229).[4] The contrasting claim, that practical rationality is in some sense *essentially* practical and thus on a par with, rather than derivative of theoretical rationality, is naturally labelled "non-cognitivism".[5] The analysis of intention that I have developed in this book, like Bratman's, does without any positive cognitive or doxastic component. Clearly, for a conception according to which intentions are at core optative attitudes, moreover, optative attitudes flanked by only a weak, negative doxastic condition, the force of the *IC* requirements cannot be derived from strictures on belief. In view of the conclusion in Chapter 7 that the subjection to the *IC* requirements is intending's most striking hallmark (Sect. 7.3), any non-cognitivist conception of intention is faced with the challenge of explaining what it is that is wrong with those relational states proscribed by the requirements, and doing so in a way that appeals to features that are in some way unique to the practical realm.

The challenge appears particularly steep in view of the conception of optative attitudinising developed in Chapter 4. I argued there that wanting* ($p\&\neg p$) is as impossible as believing it: a content empty of meaning cannot be the object of any sort of attitude. However, an agent may well have good reasons not to agglomerate

[4] Bratman argues at length against intention cognitivism, in particular against the positions of Wallace, Harman and Velleman (Bratman 2009a). Among his objections to Wallace are versions of my first and third worries.

[5] Cf. Section 7.3, note 58. As Bratman emphasizes, this is a different matter to meta-ethical noncognitivism. Perhaps some of the motivation for the latter might derive from the failure to distinguish it from noncognitivism about practical rationality and to work out a plausible version of the latter. Once we distinguish clearly between conclusive reasons judgements on the one hand and intentions on the other (cf. Sects. 8.5.2 and 8.5.3), we have two different questions whose relationship is up for grabs.

even contradictory want* contents (Sect. 4.2.2). This is because it may be rationally perfectly in order to want* two things that are for some reason incompatible.[6] This, in turn, results from the essence of the optative mode, which sets up the content of relevant attitudes as a subjective standard. A standard thus set up as a measure for reality is, in as far as it is no more than that, subject to no constraints other than that its content not be incoherent. In this it contrasts with belief, whose constitutive aim of truth imposes rational constraints on its relationship to the contents of other attitudes of the same ilk. But, if intentions are optative attitudes and optative attitudinising is at core a matter of setting a subjective standard, what is so problematic about the attitudinal combinations (i), (ii) and (iii)?

The lines along which a non-cognitivist answer runs are naturally taken to be orthogonal to the concerns expressed in cognitivist answers. According to the intention non-cognitivist, whereas theoretical rationality grounds in our commitment to the way the world is, including the way we are, practical rationality grounds in our active relationship to our selves, in what we can call our activity of *self-forging*. The second term in this compound expression picks out the activity component, whereas the first indicates that that activity takes place under constraints that ensure that the result counts as one's own. The central constraints with this function are, the non-cognitivist thinks, the *IC* requirements. In this chapter, I will develop my proposal as to how we should understand this process of constrained self-forging and why it is the *IC* requirements that do the constraining.

10.2 Bratman's Proposal: From Intention Holism to Self-Governance

Whereas Broome declares himself unable to identify grounds for the intuitive belief that rationality provides reasons (Broome 2013a, 192ff.), Bratman believes that Broome's question can be answered for the *IC* requirements. He attempts to do this in two steps. In a first step, he considers a strategy that begins with a conception of the constitutive aims of the attitudes involved, viz. belief and intention. This strategy, however, proves to be unworkable because of what he takes to be the only plausible characterisation of intention's aim. As this characterisation doesn't pick out what is specific about individual intentions, but only about the intention system as a whole, the strategy cannot show that the rationality of intending in particular situations delivers reasons for corresponding intentions.

[6] It might be objected that this is only true for extrinsic, not for intrinsic wants*. As it is likely that there are very few things that we want* intrinsically, this could possibly turn out to be true. It would, however, exclude so many cases of wanting* that a discussion which parallels that of believing would be impossible. Conversely, in order to preserve the parallel, we ought also to exclude everything that is believed on the basis of other things believed. That doesn't look like a very promising move either.

In response, Bratman claims that the individual attitudes generated within the intention system turn out to enable their bearers' to live a kind of life persons generally have a strong reason to lead. The relevant kind of life is, according to Bratman, a life of self-governance. It turns out, he claims further, that guidance by the *IC* requirements is constitutive for self-governance in agents kitted out with the capacity for intention. The requirements have reason-giving force, as they are derived from this reason. However, as the reason can under certain defeating conditions be inoperative, the *IC* requirements are not necessarily reason-providing.

10.2.1 An Ersatz-Aim

It is natural to assume that the a priori restrictions on the patterns of attitudinal combinations an attitude can enter into derive from the essence of the attitude in question. If "aiming" at truth is at least a central feature of belief's essence (Sect. 4.1.1), then such a strategy certainly seems appealing. If truth is in this sense belief's formal object, contradictory belief is the primary kind of doxastic irrationality, as it is the most direct way in which belief can undermine its own aim. We can of course fail to believe what is true as a result of not attending sufficiently to perceptible things in the world or to some of the things that others could communicate to us. Such failures to be aware of reasons are generally not failures of theoretical rationality.[7] We are defective in doxastic rationality when we are attitudinally organised in such a way that we undermine the way our truth-tracking mechanisms work with the epistemic inputs they have. If irrationality is at core a matter of self-frustration – and this is plausibly the reason why rational requirements are at least generally of wide scope – epistemic irrationality is plausibly at core self-frustration of our doxastic mechanisms.

Such a view of the relationship between belief and requirements of theoretical rationality raises the question of whether a parallel strategy can be run for intention and the intention-consequential requirements. Velleman has suggested that requirements such as *SI* can only be made sense of if intending also aims – via the specific mechanisms discussed in Section 6.3.1 – at truth (Velleman 2007, 206). If this were correct, the explanation of intention's subjection to the *IC* requirements would not run parallel to the explanation of belief's subjection to requirements of theoretical rationality. The explanation would be essentially the same. But if the intention noncognitivist is to produce an explanation that does run along parallel

[7] I argued in Section 7.2.2 that there are certain constellations in which such a lack of awareness may be a failure of *practical* rationality.

lines, he needs to name a noncognitive aim of intention that would allow such a parallel construction.[8]

In response to Velleman's challenge, Bratman, in spite of professed doubts about the strategy (Bratman 2009a, 51), suggests a characterisation that might be thought to pick out intention's constitutive aim. In doing so, he returns to characterisations of intending that he had, from the inception of his theory, argued support the normative functionalist perspective. Early in *Intentions, Plans and Practical Reason*, he characterises intention both as an action-controlling attitude (Bratman 1987, 16) and as an attitude which functions to coordinate an agent's behaviour over time (Bratman 1987, 17). These characterisations recur in the terms in which he specifies intention's "aim": "coordinated control of action that is effective in the pursuit of what is intended" (Bratman 2009a, 54), or more succinctly, "coordinated, effective control of temporally extended action" (Bratman 2009b, 231).

The suggestion, then, is that intention can be seen as having a constitutive aim with two decisive features: effectiveness in realisation of the attitude's content, but an effectiveness that is only aimed at in so far as it involves "coordination" with the realisation of attitudinal conspecifics. These two features, Bratman argues in a first step, can be seen as explaining the two main norms of rationality to which he sees intention as essentially subject: effective realisation of an intention's content requires "means-ends coherence" (non-contravention of *SI*) (Bratman 2009a, 54); coordination requires "intention consistency" (adherence to *IBC*) (Bratman 2009a, 52).

There are two important things to say about this move. The first concerns the relationship between the two components of the aim – realisation and coordination. The second concerns the grounds on which Bratman assigns them.

On the first point: the dual characterisation seems to fulfil the desideratum of explaining why norms of rationality attach to intending, whilst not attaching to other optative attitudes. Bratman's constitutive aim plausibly picks out a way in which intending is essentially more than mere optative attitudinising. That is, the characterisation is an attempt to pick out what makes intending eminently practical (cf. Sect. 6.1.1). What is particularly interesting about Bratman's move here is that it specifies the eminently practical feature of intending in terms of coordination. As wanting* already aims at realisation – in the sense opposed to content-adjustment so as to fit the world – the question of the way in which intending can be "more practical", i.e. more realisation-oriented than wanting* already is, may appear mysterious. The irreducibility thesis is basically the underwriting of this mystery of an executive mode about which nothing more can be said.[9] The

[8] The noncognitivist about intention rationality is thus faced with a task parallel to that faced by the noncognitivist about moral judgements. Compare Bratman's remarks on Gibbard's expressivist meta-ethics (Bratman 2008, 96).

[9] Searle's self-referential conception of intention, according to which it is a feature of intention's content that it be realised as a causal result of the agent having the attitude with that content, is an attempt to get to grips with the same feature (Searle 1983, 85ff.).

originality and elegance of Bratman's solution consists in not seeking any further feature of the "vertical", causal relationship between the attitude's content and some particular action of the agent, but rather in turning the focus to take in "horizontal" relationships between the agent's potential actions. The extra oomph attaching to intending to φ is, thus understood, not a matter of some additional, executive push towards φ-ing, but rather a matter of the – combined causal and normative – pressure exerted by the overall attitudinal and behavioural constellation into which the performance of the action has to fit.

If this is correct, of the two components specified in Bratman's aim for intending, it is the holistic feature that is decisive. This fits in with the emphasis of his conception of intending as "planning". In *Intentions, Plans and Practical Reason*, Bratman identifies the failure of earlier theories to focus on intention's longer-term, thought-structuring function. He expresses the criticism there in his rejection of what he called the "strategy of extension" (Bratman 1987, 6ff.) from proximal to distal intentions. Focussing on the former had led, Bratman diagnoses, to a neglect of the planning features that only become obvious in distal cases. The planning theory focuses on the latter cases, apparently assuming that the eminently action-related feature of proximal cases can be explained in terms of planning. This could be seen as inverse strategy of extension to that of the theories Bratman criticised (Roughley 2007b, 219).

As elegant as it is, this inversion does not seem to do justice to the phenomena. Let me briefly note why not: as important as planning phenomena are for intention, it looks plausible that they remain subordinate to intention's primary function of controlling their bearer's action. The point appears to parallel what we should say about belief. Just as believing a content primarily directs the believer's mind to a state of the world that that content is supposed to fit, intending some content primarily directs the intender's mind to her fitting that content into the world. The holistic features of attitudinal rationality in both cases are naturally understood as secondary relative to this primary aim. Because incoherence among beliefs is an epistemic guarantee that belief's essential aim has been missed somewhere in the net, believers operate under the rational demand to avoid doxastic incoherence. And something similar looks to hold for the practical case: it is plausibly because a lack of co-realisability of an agent's intentions guarantees that some intention's primary aim will not be achieved that intenders operate under the rational demand to avoid non-co-realisability of intentions.

If we insist on this, we are left without the advantage of Bratman's "horizontal" solution to the question of the supra-optative surplus. Instead, we are forced to return to the "vertical" question of how to make sense of the commitment to an individual action involved in coming to intend that specific action. However, once we turn to the second point mentioned above – the grounds on which Bratman assigns the dual components of intention's aim – it becomes clear that, even on Bratman's view, the "vertical" question remains unanswered.

Why, then, does Bratman take it that coordinated realisation is intention's "aim"? Well, attitudes which aim at coordinated realisation of their contents fulfil what Bratman calls a "pragmatic rationale", that is, they instantiate characteristics that are particularly conducive to their bearers getting what they want*, all in all (cf. Roughley 2007b, 218). Bratman is making the same point when he calls intentions "universal means" (Bratman 1987, 28, 53; 1999a, 5f.). The Bratmanian aims of intentions are those functions the attitude needs to fulfil in order generally to enable intenders to get what they want* all things considered, whatever that may be. The "aims" of intention in this sense are functional features which should persuade us choose to be intenders were we to have the choice.

Once these grounds for Bratman's characterisation of intention's "aim" are clear, we can see that his characterisation cannot do the job of explaining the rational requirements we're trying to understand. The fact that intending is generally useful for attitudinisers in as far as it characteristically fulfils functions x and y does not impose on those attitudinisers who are intenders the requirement to make sure their intentions fulfil functions x and y in individual cases. This point is mirrored in Bratman's attribution of intention's "aim" to the whole "system of intentions" (Bratman 2009b, 231), rather than to individual tokens of the attitude. The comparison with contractarian moral theory is germane here[10]: if the contractarian manages to achieve his first aim of showing that the introduction of moral norms is instrumentally rational for the individuals involved, he has still the problem of showing that it is rational for individuals to adhere to them in individual cases. One standard worry about contractarian moral conceptions is that a convincing moral theory would have to name intrinsic grounds for moral behaviour, whilst the structure of contractarianism seems to exclude any such possibility.

10.2.2 An Intrinsic Reason

Now, Bratman is perfectly aware of the problem of the rational purchase of the *IC* requirements in individual cases (Bratman 2009b, 231f.). Therefore, he explicitly sets out in search of a non-instrumental reason for intenders to conform to them. We should note here that there is a problem with this attempt within the framework of Bratman's dispositional-normative functionalism.[11] Because functionalism, whether normative or standard, rejects the demand for an essential

[10] Bratman repeatedly compares his two-tier planning theory to rule utilitarianism (Bratman 1987, 64; 2009c, 418; 2012, 77). The analogy with contractarianism is, I think, more precise here, as the contractarian claims to demonstrate the instrumental rationality of "morality", a claim in the face of which he has to show that it is rational for individuals not only to introduce norms, but also for them to see those norms as rationally binding them in individual cases. Utilitarianism is typically uninterested in the first step, that is, in generating moral norms out of the self-interest of individuals, instead confronting them with an overarching perspective of the interest of the collective.

[11] I introduced the characterisation of Bratman's position as "dispositional-normative functionalism" in Section 7.3 to distinguish it from Zangwil's "pure" normative functionalist position. As it is the normative dimension of Bratman's theory that is decisive in what follows and as the compound epithet is rather clumsy, I shall in what follows usually refer to his position simply as "normative functionalism".

characterisation in individual cases (cf. Sect. 3.3.1), any proposal that is consistent with functionalism must seek the source of a reason for intending in line with the *IC* requirements in something other than the states that are intentions.

Applied to Bratman's specific theory: if the essence of intending is the aptness of whatever is picked out by the term and its cognates to produce certain effects together with the appropriateness of invoking certain norms to regulate behaviour when the relevant state is realised, the question of why these norms should be valid will, have to be answered with reference to facts that are independent of intending's essence. That essence is then the explanandum, not the explanans. The explanation of intention's subjection to rational norms thus cannot ground in what intention essentially is.

Consistently with this, Bratman claims that the source of the *IC* requirements is furnished by an overarching and central feature of the life-form of agents such as us. The properties of intending for which there is a pragmatic rationale also turn out to be constitutive components of a more comprehensive feature of the life-form of persons: a feature he calls "self-governance".

"Self-governance" is Bratman's label for what is frequently called "autonomy". The alternative terminology may be in part motivated by the desire to distinguish the topic from that of free will, to which it may be thought to entertain various sorts of relationships.[12] The feature of self-governance that Bratman sees as the key to explaining the *IC* requirements is the feature that confers on an attitude the status of "speaking for the agent" (Bratman 2005, 134; 2009b, 236). It is such a feature that is sought by Frankfurt where he attempts to understand what it is in virtue of which the desire of an unwilling addict should be seen as not really *his*, even as it motivates his action. Bratman rejects each of the criteria successively proposed by Frankfurt – desiderative hierarchy, "decision" and satisfaction – at least as individually sufficient.[13]

In their place, Bratman suggests beginning with a functional criterion: that of establishing "Lockean" ties of psychological connection and continuity. Attitudes that play this role knit together to provide a "psychological, cross-temporal quilt" (Bratman 2007b, 5), a fabric that is constitutive of personal identity. Attitudes thus woven together are – let's not worry about the mixture of metaphors – removed from the "psychic stew" in which they all swim around unconnected and devoid of any particular status (Bratman 2000a, 23; 2004, 225). They get to "speak for the agent" or to count as attitudes with "the agent's weight behind them"

[12] The debate that Bratman enters by broaching this topic, particularly in the articles collected in *Structures of Agency* (2007a), began as a debate about compatibilist options on free will (the initial protagonists being Frankfurt and Watson), but shifted its focus. "Self-government" is a term used by the early Dewey to pick out the basic capacity for moral responsibility (quoted in Watson 1996, 261).

[13] Frankfurt proposes these criteria in his "Freedom of the Will and the Concept of a Person" (1971), "Identification and Externality" (1976) and "Identification and Wholeheartedness (1987). In "Identification, Decision and Treating as a Reason" (1996), Bratman argues against the sufficiency of any of these criteria. I claim in Section 8.5.1 above that Frankfurt's use of "decision" is designed to pick out something quite different to our everyday concept of deciding.

(Bratman 2004, 243) or as articulating the agent's "practical standpoint" (Bratman 2005, 130; 2009b, 236). At least they do so if they are also higher-order attitudes (Bratman 2001, 99).

Bratman claims that the aim that most of us have of "governing ourselves" generates reasons for agents to see themselves as bound by the *IC* requirements in individual situations. In as far as you have a reason to form attitudes that speak for yourself in an emphatic sense, you have a non-instrumental reason, Bratman claims, to avoid attitudinal conjunctions such as those described in (i) to (iii).[14] It follows that there is no such reason where the antecedent is not satisfied. This is the case, Bratman argues, where the attitude in question is the result of compulsion, as compulsive attitudes are not expressions of self-governance. An addict does not, for instance, have such a reason to come to intend the means to get the drug. However, in as far as the end set is not unchangeable for the agent, she has a non-instrumental reason to adopt the subjectively necessary means to her end and avoid conflicting ends, as long as she doesn't subject the relevant end or belief in necessity to revision.

This theoretical construction attempts to explain the binding character of the norms of intention rationality by tracing them back to the reason for self-governance, which Bratman assumes we all have. That reason in turn generates reasons for adjusting one's attitudinal and practical life in accordance with norms such as *SI* and *IBC* in every individual case as long as self-governance is possible. The requirements to intend the means believed necessary to one's ends and not to undermine one's intentions by generating ends taken to be incompatible are seen by Bratman as circumscribing essential components of what it is to have a practical standpoint: intending incompatible ends means that there is with respect to their relationship "no clear place" where the agent stands (Bratman 2009b, 236; 2009c, 431). Self-governance-derived reasons bite, Bratman believes, for as long as the relevant end is not unalterable for its bearer. However, where no intention revision is possible for the agent, she is, Bratman assumes, in no position to govern herself and thus has no self-governance-derived reason for the attitudinal and practical adjustments demanded by the *IC* requirements.

Bratman sees this result as compatible with Broome's diagnosis that, as the requirements take wide scope, there appears to be no demonstrably necessary correspondence between requirements of rationality and reasons (Broome 2013a, 204). Pace Broome, however, Bratman believes that the *IC* requirements can be shown to be in general reason-providing and that, where this is so, their capacity to deliver reasons itself grounds in a specific kind of reason, a reason only overridden where a particular kind of defeating condition is satisfied. Moreover, whereas Broome sees requirements along the lines of *SI* and *IBC* as applicable wherever an agent intends some end, Bratman believes that, where no self-governance is possible, the *IC* requirements have no purchase.

[14] As remarked in Section 7.2.2, Bratman's primary focus is on contraventions of the requirements on intention-belief consistency and subordinate intending, rather than on intention-intention consistency.

Bratman's noncognitivist explanation of the norms of intention rationality, then, grounds in an important aim that persons generally have reason to pursue. It takes that aim to constitutively require transtemporally unified psychological functioning, a form of functioning secured by the causal and normative ties essential to intention according to Bratman's planning theory. In the spirit of Bratman's Gricean conception of creature construction (Bratman 2000b, 49ff.; 2007b, 11), we might say that the standards of intention rationality are first introduced to fulfil the pragmatic rationale of enabling the creature-under-construction to get what it wants* most, all things considered, before at a later stage being enlisted in the project of self-governance that is made possible and perhaps necessary by further features that have been added to the creature. It is only with the latter step that the standards become rationally binding for the creature in specific action contexts, that is, take on the character of requirements. This step dissolves the analogy with contractarian conceptions of morality, which are only able to offer instrumental answers to the question as to why individuals should see rationally established moral norms as binding for them in specific situations.[15]

10.3 The *IC* Requirements, Self-Governance and Normative Functionalism

I think Bratman is right that the requirements of practical rationality are essentially demands we make on ourselves as agents, demands that are in a sense orthogonal to the functioning of reasons that ground in the way the world is, including the way we are. Moreover, I think that identifying a fundamental reason at the root of those demands is the right procedure. Nevertheless, I am unconvinced by the specific form Bratman gives to the theory. There are two kinds of reasons for this. The first sort of reason involves several problems with the Bratmanian conception of self-governance as the relevant normative source, the second has, once again, to do with problems bequeathed by normative functionalism.

10.3.1 Self-Governance as Source

Bratman's basic claim is that the *IC* requirements gain their authority for agents on individual occasions because and in as far as those agents have reason to be self-governing. Bratman backs this up with the claim that adherence to the *IC* requirements fulfils a function that is part-constitutive of self-governance: that of

[15] Kantian conceptions of the relationship between the self and morality are perhaps the model here, as the Kantian – certainly Christine Korsgaard – takes it that self-governance actually involves adherence to the moral law, a law that one gives oneself (cf. Korsgaard 1996, 98).

tying the various time-slices of the agent together, a cohesion without which no agent could be autonomous. The kinds of ties thus established are those that Parfit calls "psychological connectedness" and "continuity" (Parfit 1984, 204ff.). Bratman's use of this Lockean idea differs from Parfit's in two ways that are important here: first, where Parfit focuses on assertoric or factive states of mind, primarily on memory, Bratman's model of self-governance works with optative attitudes. Second, where Parfit sees certain purely quantitative relations as decisive, Bratman picks out particular attitude patterns, namely those that fit the *IC* requirements.

In this section, I shall point out a number of serious problems with the claim that our reasons for adhering to the *IC* requirements ground in our reason for self-governance. I shall distinguish three such problems, all of which concern the relationship between the requirements and their putative ground. These concern (a) the former's categoricity in contrast to the latter's gradability, (b) the modality of the reference to the latter and (c) a difference in their extension.

(a) A first difficulty can be labelled the *gradability problem*. It comes into play at two, possibly three points in the attempt to ground the *IC* requirements in self-governance. First, psychological continuity is a gradable property, as is the extent of an attitude's contribution to its realisation.[16] However, the requirements of intention rationality codified in principles such as *SI* and *IBC* are stringent: if you intend to bring about some end and you believe your φ-ing is necessary for you to bring that end about, then, as long as you maintain your first intention and your belief, you are required to intend to φ, not just to thus intend to a significant extent or in a significant proportion of cases. As Parfit points out (see below, Sect. 10.3.2), the mere persistence of an attitude contributes to psychological continuity. Clearly, this fact generates no stringent requirements to ensure that your attitudes persist – although it no doubt supports the diagnosis that something is seriously wrong if, for instance, an agent shows no consistency in optative or emotional reactions to recurring situations. Serious inconsistency in such reactions may justify recommendations that someone sit down and think through what is really important to them or seek psychiatric help. But it is unclear how categoricity enters the picture: the explanandum and the explanans don't fit.[17]

But gradability is not just a problem for the correlation between the binding character of the *IC* requirements and psychological continuity. It is also a problem for the relation between the requirements' authority and self-governance. This is because self-governance itself isn't plausibly a matter of yes or no: we can surely govern ourselves to a greater or lesser extent. If a reason for self-governance is to do the job Bratman assigns it, it looks like it will have to be a reason for perfect self-

[16] Compare Bratman's description of the way certain attitudes "speak for the agent because they help constitute and support the temporal extension of her agency" (Bratman 2005, 208). Both supporting and helping to constitute are gradable concepts.

[17] Parfit's solution to this problem in the sphere of personal identity involves setting a threshold beyond which we stipulate that there is sufficient psychological connectedness and therefore see ourselves as justified in talking about the same person (Parfit 1984, 206).

governance. So, even if psychological continuity were to be a constitutive condition of self-governance, it would still be unclear why self-governance imposes stringent requirements.

Note that this criticism does not entail that an explanation of the categoricity of the *IC* requirements need be an explanation of their overridingness relative to other reasons. We need to explain why they are stringent within the sphere of rationality; whether rationality might be overridden by reasons with another source – for instance, by moral reasons – is a quite different matter. I will come in a moment to some cases in which I think rationality is simply overridden.

There is even a third point at which gradual considerations raise doubts about Bratman's construction. As the grounds of the doubts are controversial, I won't insist on them, but they do seem to be worth mentioning. They ground in a metaphysical worry, namely whether there really can be intentions that we are helpless to prevent forming (cf. Greenspan 1978; Watson 1999, 72f.). Greenspan and Watson have argued that talk of irresistibility reduces to the normative idea that it would be unreasonable to expect the agent to suffer the forms of unpleasantness that would result for them from not taking on the relevant attitude. If this were correct, we would here also find ourselves on a continuum: certain circumstances would make self-governance more difficult. Again, we could stipulatively set a threshold of unpleasantness, beyond which demanding the relevant form of attitudinal coherence is taken to be unreasonable. This would, however, involve thinking of such unpleasantness as a reason which overrides the reason for self-governance, rather than as taking the reason to be inoperative because of the impossibility of governing oneself.

(b) A second problem with Bratman's explanation of the *IC* requirements' hold on us concerns the modality of the reference to self-governance in the explanation. Why, one wonders, should the *mere possibility* of self-governance be decisive for the *IC* requirements' validity? An explanation in these terms appears to entail that the importance of self-governance is reason-providing even where an agent is not governing herself in adhering to the *IC* requirements. Plausibly, putting the tea into the pot in order to make ones preferred beverage is not a particularly plausible example of self-governance. Bratman doesn't seem to think it is, as he states in other contexts that the attitudes constitutive of self-governance are actually highly specific "higher-order plan-like attitudes" (Bratman 2005, 208), in particular policies to treat certain considerations as reasons. But these higher-order attitudes aren't even intentions in the normal sense of the term, but rather much more ambitious plan-like attitudes with a generalising component. But if this is the case, why should the possibility of acting according to such attitudes subject the tea maker to a norm of means-ends coherence? If self-governance were indeed to be the source of the *IC* requirements, why should they be in play merely because it is possible that we govern ourselves at some point although we are not doing so in performing the action we are actually performing?

(c) A final problem for a conception according to which self-governance is the source of reason-giving power of the *IC* requirements lies in what is intuitively a significant *difference in extension* of the two phenomena. It is quite simply far from clear that the norms of rational intending are in play when and only when self-governance is possible. Compare John Broome's example of etiquette (Broome 2008, 96f.). Etiquette is a code that prescribes certain forms of behaviour under certain circumstances. The fact that there is a prescriptivity internal to the code doesn't entail that the relevant prescriptions necessarily apply to agents. In this respect, the norms of intention rationality are obviously different from the rules of etiquette: they have purchase on us, even when we have no good reason to form the intention they focus on. Bratman's claim is that, even if an agent has no reason to form the intention in the first place, once he has done so, he has a reason of self-governance to organise various departments of his attitudinising in accordance with those norms. That reason, however, is supposed to lose its applicability where no self-governance is possible. In *Intentions, Plans and Practical Reason*, Bratman had expressed worries about the possibility of "inacceptably" "bootstrapping" a reason for an action into existence by simply intending the action (1987, 24ff.). It turns out that bootstrapping is not itself the problem. Instead, Bratman (now?) thinks that bootstrapping is only inacceptable if the original intention is not subject to the control of its bearer.

The attraction of Bratman's construal is that, if true, it would dispel the mystery that appears to exist if rationality is on the one hand the source of unavoidable demands on us, whilst on the other hand being completely unconnected to reasons. In order to be successful in this aim, the construal clearly has to name defeating conditions that do indeed pick out cases in which the *IC* requirements don't apply. However, I don't see any particularly good, non-question-begging reasons to think that this is the case.

Assuming that there are cases of compulsion that involve an agent's incapacity not to form certain intentions, there is little evidence that cases that appear to qualify as such are generally understood as exceptions to *SI*. Take a case of kleptomania: Ken is assailed by the uncontrollable urge to steal dental floss. An essential part of the urge's uncontrollability is its automatic transformation into an intention. As he is aware that the drugstores in his home town are policed by security, he knows it is not so easy to attain his goal. Nevertheless, there is his intention, clear and strong. He works out that the only way to realise it is to enter the store when the security guard goes for his coffee break. Ken, I submit, has to form the subordinate intention to do just that if he is not to be locally irrational. It is, of course, clear that his local irrationality pales into insignificance in the light of the weirdness of his superordinate intention and the global irrationality in which it very probably involves him. Moreover, Ken's formation of the subordinate intention is unlikely to be motivated by the thought that he would be irrational not to. Nevertheless, the usual strictures are surely in place. What Bratman's construal picks up on is the resistance one may have to saying that Ken also has a reason to enter the store as the guard goes for his break. Bratman's claim that the rational requirements become inapplicable seems to be a case of bullet-biting in the face of the idea that self-government is the requirements' source. But it seems at odds with everyday understanding.

Compare Ken's cousin, Ben, who has bet the members of his clique that he can steal a toothbrush every day for the next week. Ben, who is not a compulsive gambler, develops the same belief as his cousin about the necessity of entering the drug-store during the security guard's coffee break. Bratman must now think that *SI* applies to Ben, as he has a reason of self-governance to move from end to means. Now give Ben two additional features: the disposition to feel strongly ashamed if he doesn't face risks he has publicly committed himself to facing and the – literal – inability to live with the shame he would feel in such a case. It looks like under these conditions Ben would no longer be governing himself in forming the subordinate intention. Nevertheless, it doesn't seem as if *SI* has now been suspended. As Ben has the superordinate intention as well as the belief he shares with his cousin, it again seems clear that he would be locally irrational not to form the subordinate intention to enter the store when the guard disappears for coffee.

Or take Buck Finn, who in a slightly different version of the story involving his relative, Huck, intends to betray Jim, but doesn't adopt the necessary means, whilst still maintaining the intention to do so.[18] In such a case, it seems that rationality and reasons, or at least the reasons that are important, come apart. Buck would be irrational, but fortunately so. The important point for my purposes is that this seems to be true, *independently of whether Buck is governing, or losing control of himself*. In what might be supposed to qualify as a case of the first kind, he perhaps forms and maintains the superordinate intention to betray Jim in spite of himself, his failure to form the subordinate intention being an assertion of his 'true self', expressed in a pattern of attitudinising held together by 'Lockean ties'. In the second kind of case, the Lockean ties might bind together a pattern of attitudinising that fits the slaveholder ideology, whilst the failure to form a subordinate intention results from a compassionate impulse that is "out of character". There seems to be no non-ad-hoc reason why *SI* should have purchase in the one case and not in the other. And even if Buck's adherence to the slaveholder ideology in the second version has resulted from indoctrination, the effects of which are unalterable at that moment, that doesn't seem to change the fact that rationality and moral reasons point in different directions.

At best, one could say that Bratman is proposing a revision of our conception of rationality, a revision he takes to be justified because in certain cases we simply have no reason to see the requirements as having purchase. I will be arguing that revision is unnecessary, as there is a reason why the requirements are in place even where the agent is seriously unhappy with her own intention.

[18] The Huckleberry Finn case is usually, and with good reason, discussed in terms of the moral worth of Huck's behaviour. In discussions such as Bennett's and Arpaly's, the topic is the relationship between Huck's value judgement and the reasons of which he is unaware; whether he intended to turn Jim in and, if so, whether he maintains that intention after not doing so, are not in view. Nevertheless, it should be noted that Arpaly's use of the example is part of an extended argument for a very different conception of rationality to that at issue here, a conception that essentially picks out responsiveness to reasons, where the reasons in question are in part moral reasons. Cf. Arpaly 2003, 125.

Note, finally, that Bratman has conceded that “unalterability” of an optative attitude need not actually be a defeating condition for self-governance. There are, he argues, “Frankfurtian reasons of self-governance” that ground in volitional necessities, as in the case of love. Aims that ground in such necessities are subject to the norms of rationality, as they don’t entail lack of self-governance (Bratman 2009c, 438ff.). This move raises the question of whether there is a unified conception of self-governance in the offing at all. If it isn’t attitude unalterability that defeats self-governance, what does? Only when some such replacement condition is on the table will be able to judge whether it is does better on the coextensivity test. My suspicion is that the notion of self-governance is going to be too strong, however it is construed.

10.3.2 Self-Governance and Normative Functionalism

In conclusion of my discussion of Bratman’s position, there is a basic problem in the kind of model that he is forced to offer that results from the normative functionalist analysis of intending. Indeed, the problem looks to be unavoidable for the normative functionalist, whatever explanatory construction he proposes. Put succinctly, the problem is that the normativity that needs explaining is itself required to do the explanatory work. What we want to identify is the source of the purchase of the norms that come attached to intention. Bratman’s answer names their constitutive necessity for the self-governance of beings such as us. Now, the reason why such norms are constitutive of self-governance, that is, the reason they enable the attitudes that conform to them to stand in for their bearer, is that they tie her various psychological states together by establishing connections of Lockean – or Parfitian – psychological continuity. What, however, is the *mechanism* by means of which they establish such continuity?

According to Parfit, there are a variety of different ways in which the strong psychological connectedness can be established, overlapping chains of which are required by the psychological criterion for transtemporal identity. The paradigmatic connection that Parfit takes over from Locke is that of an experiential state of a person at t_1 being the object of a state of remembering at t_2. Parfit remarks that other kinds of psychological connection would also contribute: that between an intention and a later action or that which holds “when a belief or a desire, or any other psychological feature, continues to be had” (Parfit 1984, 205). If intentions possess a characteristic stability relative to many other optative attitudes, then they are good candidates to establish Parfitian continuity through persistence. Moreover, there is certainly a great deal of plausibility to the idea that intention establishes transtemporal identity through its forward-directedness in an analogous way to memory’s doing so through its backward-directedness. Finally, the fact that intentions are characteristically bound up in networks of practical deliberation and further intention formation means they play a much stronger role than that fulfilled by the simple two-place relation between intention and action.

Two features of the variants Parfit names en passant should be remarked upon. First, Parfit takes it that the attitudes conceivably involved in establishing psychological continuity can be of pretty much any kind. Second, the relation suitable for doing the work need not, he thinks, necessarily link an attitude and its object or, we can add, an attitude and further attitudes derived from the content of the first attitude. Instead, mere persistence of psychological states over time may equally do the trick. Certainly, the life of someone with a persistently strong desire, say, to drive big cars or never to be talked down to by women, will have a certain kind of continuity. The same thing can be said for the life of an agent with certain strong emotional dispositions. Now, Bratman notes this distinction and remarks that intentions are a particularly salient way in which "referential connections" are established over time (Bratman 2000a, 30). However, he doesn't argue that only such semantic ties are important, also seeing intentions' characteristic stability as contributing to their bearer's transtemporal identity (Bratman 2005, 207).

Bratman is explicit that there is no metaphysical reason why self-governance has to be delivered via intentions. The metaphysics of self-governance simply requires, he says, some mechanism to pick out those attitudes that stand in for the agent. That is done, he goes on to claim, in planning agents such as ourselves, by planning structures. The idea, as mentioned at the end of Section 10.2.2, seems to be that we could think of planning structures as first coming into being because they fulfil the "pragmatic rationale" identified as also providing intention's "aim", but could then, in a further step, be enlisted in the service of the self-government of an agent that has in the meantime acquired a reason to be self-governing. In agents that have become intenders, agents such as human persons, adherence to the norms of intention rationality is "a necessary constituent of self-governance". The necessity in question is not metaphysical, but, one could say, anthropological.

The question now is: if intentions, but also beliefs, desires and emotions establish the relevant connections and continuities, what is it about the particular kind of links established by intentions in virtue of which their maintenance is privileged in the constitution of self-governance? Without an answer to this question, the claim that reasons of self-governance are the source of the *IC* requirements' binding character is unsupported. Unfortunately, it seems that the answer a normative functionalist has to give here involves appealing to the normativity essential to the concept of intending. The key claim of the normative functionalist is that intentions' specificity derives from the binding power of the norms with which they are necessarily "associated" (Sect. 7.3). But this special connection is precisely the one we are trying to explain. In as far as the key transtemporal connections established by planning result from the planner's being bound by the *IC* norms, their binding character cannot be explained by their establishing those very transtemporal connections.

10.4 Anchoring Attributability

I have discussed Bratman's proposal in some detail because I think he is correct to tie the normativity of intention rationality to the agent's activity of self-determination. For a noncognitivist about intention, the associated norms of rationality cannot ground in commitment to the way the world is or to the way one is as part of the world, but must ground in a commitment to the ongoing process of forging oneself. The open-ended character of self-forging excludes the attitudes decisive for that process being exclusively assertoric; the fact that the process cannot be unconstrained for it to count as the forging of a "self", rather than as a series of erratic, unconnected attitudinal events, is what makes the process reason-giving.[19] From this perspective, there are two key questions: first, why do the standards in play here pick out the specific attitudes that they do pick out, in particular, why do they focus on intentions? Second, why do they have the content that they have, that is, why do they relate the agentially relevant features, particularly attitudes, in the way they do.[20]

In spite of my basic agreement with Bratman's broad perspective, I have disagreed with various features of his argument for a non-cognitivist justification. First, there is a lack of extensional equivalence, indeed, there may be a fairly large gap, between cases in which the *IC* requirements are in place and cases in which self-governance is at stake. Second, both self-governance and contributions to its maintenance are gradable properties which seem at the most able to ground levels of rational pressure, but not stringent standards such as the *IC* requirements. Third, the commitment to normative functionalism prevents Bratman from saying anything genuinely informative about why it is intentions, rather than mental states of other types, that figure in the relevant requirements of rationality. The only answer the normative functionalist can offer is circular. I hope to show that an appreciation of these problems can pave the way for the development of a more convincing noncognitivist justification of the *IC* requirements.

The concept of self-governance seems to be at once too strong and too weak. On the one hand, it is too strong because there is little plausibility to the claim that the *IC* requirements only apply where an agent's self-governance is at stake. Moreover, the plausibility decreases even further if self-governance is tied to the unalterability of the agent's relevant ends. On the other hand, the concept of self-governance is in a sense too weak because, as it picks out a gradable phenomenon, it cannot give rise to stringent requirements. There appears to be no way to overcome the gradability problem by stipulating a threshold beyond which agents would be

[19] The structure of the adequacy conditions on a noncognitivist theory of intention rationality mirrors that of certain noncognitivist meta-ethical positions, most prominently Hare's view that moral judgements are unique combinations of the exercise of our capacities for "freedom" and "reason" (Hare 1963, 1ff., 111).

[20] The claim that not all the relevant features are attitudes, the denial of the claim that rationality supervenes on the mind, was first argued for in Section 7.2.

genuinely governing themselves. The second problem is aggravated by a move Bratman makes that might appear suited to solve the first. He specifies the relation between the validity of the *IC* requirements and self-governance through a modal qualification of the latter term in the relation: it is the possibility of self-governance that needs to be upheld for the *IC* requirements to be operative in individual cases. But this move attenuates the connection between self-governance and the *IC* requirements to such an extent that there seems to be no good reason for an agent to see them as binding for her. If self-governance is possible, but is obviously not at stake, why should an agent in particular instances feel bound to satisfy one of its constitutive conditions?

If a noncognitivist explanation of the *IC* requirements is to work, it has therefore to ground them in a feature of agency that is both less demanding than self-governance and is a yes-no, rather than a gradual matter. This, I think, can only mean that it has to be a categorically necessary feature of agency, understood as the capacity for intentional action.[21] Moreover, in order to deliver the requisite connection to the *IC* requirements, it has to show why the feature's instantiation makes the requirements inescapable, whilst avoiding the normative functionalist claim that triggering the requirements is essential to what makes an optative attitude an intention. Finally, if the conception grounds the binding character of the *IC* requirements in a reason that is part-constitutive of agency – with no get-out clause of the Bratmanian type – it looks as though the requirements themselves are going to generate reasons. If this is correct, then advancing such a conception is going to involve some bullet biting. I hope to show how this can be done harmlessly.

In what follows, I shall be arguing that the upstreamist disjunctive conception of intention formulated at the end of the last chapter provides the key to the agential feature we are after. The agential capacity for deliberation and decision, I shall claim, is what at bottom explains the subjection of agents to standards of practical rationality. Locke characterized the deliberative capacity as the power to suspend one's mechanisms of desire satisfaction and saw it as essential to responsibility. I think, indeed, that the connection to responsibility is key. However, everything hangs on understanding both the nature of that connection and the relevant concept of responsibility.

In a slogan that covers both points, I shall be arguing that intending *anchors attributability*. On the first point: the claim that intending is responsibility's anchor itself covers two theses that maintain intention's necessity for responsibility both at the level of general capacity and at that of individual cases. The truth of these claims grounds in turn in the fact that to come to intend some action is to take responsibility for that action, where taking responsibility is a matter of making an action one's

[21] "Agency" can also be understood more broadly to also include what I called subintentional actions, such as scratching one's head absent-mindedly (Sect. 5.1.3). Such cases, in which agents pursue purposes, understood as mere wants*, need not have the feature we are after (cf. also the reflections on non-human animal agency in Sect. 2.6). Correspondingly, the *IC* requirements have no purchase.

own in a way that only deliberative agents can do. On the second point: the form of responsibility thus taken is the core component of what Watson called attributability, and opposed to accountability, to which the idea of holding responsible is central. In developing the capacity for deliberation-terminative wanting*, we come to take responsibility for what we do in response to the demands of our social environment that hold us responsible.

The connection to the *IC* norms is that they embody the conditions whose non-satisfaction would undermine the intelligibility of thus taking responsibility in individual cases. In line with the disjunctive strategy developed in Chapter 9, I then go on to argue that non-decisional intentions are able to fulfill the same function.

10.4.1 Deliberation and Taking Responsibility

Before we start, it should be noted that the noncognitivist emphasis on self-forging involves no claim that we should think of "the self" as entirely the product of such processes. Alongside the importance of active processes of self-constitution, emphasized in conceptions such as Bratman's and the Frankfurtian conception from which he takes his lead, there is also a clear sense in which a person is revealed in the emotional and other experiences to which she finds she is disposed. Persons thus plausibly entertain to themselves relationships of both self-determination and self-discovery.[22] The noncognitivist theorist of intention rationality claims that the former process is essentially a matter of ways of intending – i.e. of taking on a specific type of optative stand – and that the *IC* requirements codify basic conditions in the absence of whose satisfaction an intender isn't determining herself in the relevant sense. As already indicated, it will be important that the concept of self-determination at work here is relatively undemanding.

My central claim is that intending is, in a sense closely related to that employed by Locke, a "forensic" concept, picking out as it does the psychological phenomenon that is the key to an agent's "ownership" of what she does (cf. Locke E II, xxvii, §16, §26). As is well-known, Locke's discussion of the problem of action ownership focuses on transtemporal identity in an attempt to clarify which actions at some time t_1 can legitimately be imputed to an agent at a later time t_2. (We have seen that Bratman attributes to intentions the "Lockean" function of tying time-slices of agents together, thus providing a necessary condition of self-governance.) According to the position I shall be developing on intention rationality, the forensic importance of transtemporal continuity, however that continuity is brought about, presupposes that there is an agential feature which at t_1 – or at some time prior to the action – establishes the ownership that forensic investigations are concerned to trace. Intending is that feature. We could thus characterize intentions as *intrinsically forensic*.

[22] Marya Schechtman (2004) talks of "self-control" and "self-expression".

The point can be put in terms I used at the beginning of Section 9.5: intending, I claimed, is necessarily the *anchor of responsibility* for our behaviour. This metaphor summarizes two related, but distinguishable claims: first, at the level of capacities, if we weren't intenders, we would not bear responsibility, that is, we could not be legitimately held accountable for our behaviour; and second, at the level of individual cases, our actions or omissions – and their results – can only be our responsibility if there is at least a potential route to them from our intending.

I will first say a word or two about the second claim, which some authors find controversial, but which I think, understood correctly, fits well with our moral and legal practices. Justifying the less familiar first claim needs more argument. My case rests on a combination of empirical data from recent developmental psychology and the reinterpretation of an important distinction in the theory of responsibility.

We can call the second claim *the anchoring claim for individual cases*. The basic idea here ought to be fairly clear: where we bring about *p* because we intend to do so, we are pro tanto responsible for doing so; similarly where we bring about *p* because we intend to bring about *q* but believe that doing so will also bring about *p*. Indeed, in any case in which we are responsible for bringing about some *p* that we don't intend to bring about individually, but the bringing about of which we take to be a risk of bringing about some intended *q*, our responsibility for *p* is to be traced to our intention to bring about *q*.[23] There is no recklessness without the intention to do something, coupled with a relevant belief concerning the action's probable consequences.

Only in cases of negligence is there no requirement that the agent have intended a related action for her to be responsible. Even here, however, there must be something the agent could have intended somewhere along the line in order that she not violate the relevant norm. Certainly, inadvertent risk taking is culpable if it exceeds some threshold that many people will stay below out of habit. So avoiding negligence need not be the result of intending anything explicitly related to that threshold. Nevertheless, if the charge of negligence is to be a fair one, an agent not disposed to attend to relevant risks must have been able to see to it that she does thus attend. 'Seeing to it' here means successfully adopting intended means to that end. In the words of George P. Fletcher, certain kinds of circumstance place an agent with a certain kind of insensitivity under a "duty of inquiry" (Fletcher 1971, 423). At the bottom line, that duty must have been able to be fulfilled as result of deliberately focusing one's attention. The only apparent exceptions to the second intention-as-anchor claim are cases of strict liability. However, I think strict liability is best understood as a means of social engineering – of preventing certain harms

[23] Bratman has argued that, whereas the ascription of intentions aims primarily at action explanation, it is the classification of actions as intentional that aims at assigning responsibility (Bratman 1987, 124f.). In as far as "intentionally φ-ing" covers the kinds of case described in the second and third sentences of the paragraph to which this footnote refers, the concept of φ being done intentionally covers more cases of φ-ing for which an agent is responsible than does the concept of φ-ing because one intends to φ. This is, however, compatible with the anchoring claim for individual cases – indeed, it presupposes it.

or of distributing social costs. There is therefore a clear reason why it shouldn't be classified as a form of genuine responsibility.

The anchoring claim for individual cases, according to which at least a potential route from intention to an action or omission is necessary for the action's or omission's legitimate attribution, says nothing about sufficient conditions for attributability. The satisfaction of defeating conditions such as duress, insanity or inability to understand the relevant concepts can sever the link. It is a central goal of theories of moral or legal responsibility to catalogue and explain such defeating conditions. Such matters go well beyond a theory of intention. A theory of intention, however, should explain why it is intention that performs this anchoring function.

This brings us to *the anchoring claim for capacities*: that we could not be bearers of responsibility if we were not intenders. According to the disjunctive upstreamism I have been arguing for, being an intender is essentially a matter of being a practical deliberator. (Not all intentions result from practical deliberation, but we would have no occasion to assign a special forensic status to consciously tokened unrivalled motivational attitudes were we not to be practical deliberators, and thus bearers of decisional intentions.) If both the upstreamist theory and the anchoring claim for capacities are true, it follows that we could not be responsible if we were not deliberators. That is certainly what Locke believed. What Locke called "the power to suspend the prosecution of this or that desire" (E II, xxi, §47) is certainly not, as he himself in one of his moods thought, sufficient for what is generally called free will or moral responsibility. Nevertheless, he was surely right that the ability actively to step back from the immediacy of desire-driven action is where we need to start in order to understand persons' capacity for forms of behaviour we see ourselves as justified in criticising. In taking deliberation-terminative optative stands on a practical question, an agent is, as I shall put it, *taking responsibility* for the behaviour of hers represented in the content of the attitudes thus formed.

In order to get clear on what I mean by this, it is helpful to revisit a distinction made by Gary Watson in a deservedly influential article on what he called "Two Faces of Responsibility". Use of the full concept of responsibility, Watson claimed, involves reference to two dimensions of the concept which can, up to a certain point, be taken apart. Part of Watson's aim was to emphasize the importance of the dimension he called "accountability", the dimension picked out by talk of "holding" someone responsible. We hold agent's responsible within a framework of demands addressed by individuals seen as vested with relevant authority. This dimension is separable from the dimension Watson calls "attributability" and which concerns the conditions under which a form of behaviour is not only caused by, but also *expressive* of the agent in question. This is the case, Watson suggests, where agents have "adopted an end", thus "taking responsibility" for what they do (Watson 1996, 270f.).

Watson's distinction helps me to clarify the sense in which coming to intend is precisely to take responsibility. Such a clarification needs, first, to keep the concept of accountability, that is of holding responsible, separate from the notion of taking responsibility. Second, it also needs to show that Watson's concept of attributability should itself be thinned out somewhat.

The first point can be made by noting a legitimate use of the phrase "taking responsibility" that is not at issue here. The use is to be found in a recent paper by Stephen Darwall, in which the author says that when we feel guilt, we "(begin, at least to) take responsibility for what we have done" (Darwall forthcoming). This retrospective and reactive use doesn't only contrast with the prospective or present-directed use on which I am focusing. It also picks out an attitudinal reaction whose complexity goes well beyond the referent of the phrase in my use here. Taking responsibility in the accountability sense is a matter of holding oneself responsible from the perspective of the representative members of the normative community at issue. It is thus a three-place relation. Where the community is the moral community, to do so is to feel guilt. Intending clearly doesn't necessarily involve even prospectively taking on a perspective that relates one's intended action to the demands of authoritative players in some practice.[24]

To take responsibility in the attributability perspective is to "take a stand" on the question of the options for action at issue. The two-place expressive relation between the agent and the action is marked by Dewey in a passage quoted by Watson (1996, 260), where Dewey states that conduct for which we take responsibility "is ourselves objectified in actions". For this reason, responsibility thus understood "belongs", Watson says, "to the very notion of practical identity" (Watson 1996, 271).

The formulations just quoted seem to me to describe nicely what we are doing when we come to intend and act to realise that intention. Here, once again however, we need to sound the note of caution I repeatedly sounded in my discussion of Bratman's discussion of self-governance. The relevant kind of stand-taking is to be understood as importantly undemanding, more precisely, to demand only the satisfaction of those conditions named in *ID* (Sect. 9.7). Watson, however, takes the attitudinal preconditions of attributability to be significantly stronger.

For Watson, conditions of attributability are conditions of aretaic evaluation, that is, of evaluation of the agent's character. Such evaluations differ from those that presuppose accountability in that they raise no questions about the avoidability of the action that gives rise to them, as concerns about avoidability derive from worries about the fairness of sanctions imposed where demands are not met. Aretaic evaluations, rather, concern what has often been thought of as the "real self" of the agent, a concept Watson proposes that we explicate in terms of the agent's own values or higher-order attitudes. This explication, he thinks, should exclude attitudes generated by brain-washing or hypnosis. In other words, actions are attributable that are, in some sense of the kind discussed in Section 8.5.1, strongly owned by their agents.

[24] For this reason, Bratman is right to reject Stoutland's claim that intending requires that the intender be prepared to take "full responsibility" for her action (Bratman 2014, 62). As Bratman puts it, intending's "role in practices of accountability" is "not essential to what intending is". However, because accountability is only one of two dimensions of responsibility, it doesn't follow that responsibility and intending are not essentially connected.

In spite of his avowed aim, I think that Watson has allowed features of accountability to creep into his characterisation of attributability. On the one hand, he is certainly right that there is a dimension of our ethical evaluations that concerns others' character structure independently of the avoidability for them of the relevant attitudes and dispositions (although the relevant structure is surely not restricted to the psychological features of which the agent reflexively approves). On the other hand, this evaluative dimension is not necessarily in play when we ascribe full responsibility. Prior to asking the avoidability question, we simply want to know if the – directly or indirectly – normatively relevant action is one the agent opted to carry out. Whether he did so with reservations relative to his system of values is at this point irrelevant.[25] Of course, there are conditions under which we might then want to ask whether his thus opting was the result of brainwashing, compulsion or duress. But, in the context of responsibility, these questions all seem motivated by the concern with avoidability in the light of potential sanctions.[26]

Attributability, then, is at core a matter of an agent with an activated capacity for deliberation having opted for the relevant action. This is, as Watson rightly points out against Slote, more than the mere effect of whatever set of behavioural dispositions a human, like any other animal, happens to possess (Watson 1996, 288). However, in order to go beyond such a pure causal view, we don't need to advert to a conception of a real self or to other forms of strong ownership. All we need are mechanisms to confer weak ownership on actions. Actions weakly owned are actions that bear the mark of deliberation, or at least of deliberability.

10.4.2 On the Ontogenesis of Taking Responsibility and Holding Responsible

Watson both distinguishes attributability and accountability and claims that they are closely related. I have been suggesting that we whittle the first of these down to a concept of weak attributability and see this as explained by the disjunctive concept of intention codified in *ID*. The claim that coming to intend is essentially to take responsibility is supported by empirical evidence that intending's ontogenesis may be a response to being held responsible by the members of the social world into which one is socialised. The evidence I wish to present briefly is of two sorts: first, data that point to a correlative genesis of intention and the capacity for

[25] Watson's view of the relationship between aretaic evaluation and accountability seems to entail that normative ethical judgements are judgements of virtue plus accountability. But one reason why normative ethical, as opposed to virtue-ethical, conceptions have appealed to many authors is precisely the minimal demands of character assessment involved in holding someone responsible for contravening a norm.

[26] In legal contexts, the question of the agent's character can come into play on a third level, the level of sentencing. At this level, it is sometimes appropriate to ask whether the legally problematic action was "out of character".

deliberation, and second, data that indicate that intention and deliberation develop in an environment that is already strongly structured by normative demands, that is, by – partially inchoate – practices of holding responsible. If the story is correct, it provides developmental support for what I have been calling the anchoring claim for capacities.

As we saw at the end of Chapter 9, young children already have some idea of purposes. They are plausibly taken to be pursuing such goal states in either responding spontaneously to endogenous impulses such as thirst, or in their immediate responses to social norms or to what they at least take to be socially expected: endogenously or exogenously generated wants* appear to take control of young children's behaviour. This is plausibly reflected in the lack of differentiation between explicit conceptions of wanting, desiring or intending (Astington 1993, 89ff.; Bartsch and Wellman 1995, 68; cf. Sect. 9.7), something that begins to change at around five (Tomasello et al. 2005, 678) and appears to be linguistically anchored by seven (Astington 1999, 309). Between 3 and 6 years, children's self-understanding apparently changes significantly, in particular, they begin to develop what Wellman calls "a constructive or active mind" (Wellman 1992, 118). This is the time during which fully fledged metacognition, that is, the capacity explicitly to represent one's and others' mental states, develops (cf. Esken 2012). There is strong evidence that this may require the development of certain linguistic, particularly syntactic capacities (Hale and Tager-Flussberg 2003; Astington 2006). Whatever the precise ontogenetic relationship between an active conception of mental processes, the capacity to think about what one wants* – *as* the things one wants* – and the ability to produce syntactically complex representations, it seems plausible that these are all essential components in the capacity for practical deliberation. The developmental evidence thus dovetails with the claim that the full-blown concept of intention only becomes applicable with the development of the ability to practically deliberate.

At some point around the end of the preschool or early school years, then, children become able to step back from at least some of their impulses and some social expectations and take their own stand on the matter at hand. 'Own' here doesn't pick out a relation to a mysterious new entity, 'the true self', but is rather simply *the mark of deliberation*. That mark can first be imprinted on one's actions by having reflected on what to do and terminating one's deliberation by means of the relevant want* token. Later, the mark of deliberation is traced by the possibility of deliberative influence raised by the conscious tokening of a relevant want*.

This connection looks to be expressed by our use of the adjective "deliberate" – in the cryptic manner typical of linguistic connections that develop under holistic semantic conditions. By and large, adult English language users seem to employ "deliberate" and "on purpose" more or less as synonyms. As early as 2 years of age, children start to claim that they have done things accidently or on purpose depending on whether the deed done was represented as the content of something they wanted* (Astington 1993, 89ff.; 1999, 306f.). The introduction of the adjective "deliberate" indicates that this distinction has become more refined, introducing as it does the idea that the agent's purpose can now be seen as one formed in the light of possible or factual reflection on what will or might result from attempting to achieve the goal. States of affairs we bring about deliberately, like states of affairs we bring about intentionally, need not be states of affairs individually intended, but they are necessarily states of affairs brought about as a result of intending an action whose causal consequences or non-causal results we have epistemically represented. That representation takes place prototypically, although not necessarily, in the course of practical deliberation.[27]

There is, then, evidence that the full-blown concept of intention only develops with the capacity for deliberation, a transition marked by the acquisition of the concept of deliberate action, alongside that of purposeful action. I now want to adduce some additional data which suggest that these developments may be, at least in part, a response to being confronted with normative demands and associated practices of holding responsible.

The relevant work concerns the sensitivity of young children to social norms or to what they take to be the socially standardized way of doing things. Three-year-olds are frequently quick to imitate, and overimitate actions performed by adults and to protest or criticize when peers deviate from what they apparently take to be "the way of doing things" (Rakoczy et al 2008; Schmidt et al 2011; Kenward 2012). It seems, then, that children very early in their development see themselves as subject to social norms, to which the default response appears to be conformity and even prescription towards peers. There is also evidence that the acquisition of the relevant normative beliefs takes place without the children having any conception of why the presumed norms are in place. This is particularly clear in overimitation studies, in which senseless components of complex action sequences are rigorously imitated and also called for when others are acting (Kenward 2012; Keupp et al 2013).

This material shows that young children are strongly aware of growing into an environment that is deontically structured long before they are capable of engaging in practical deliberation. There is even evidence that they tend to project requirements where there are none, complying with norms that don't exist. In as far as compliance is automatic, no questions of responsibility arise. Here, however, as elsewhere, adult and older peers very quickly begin imputing responsibility in a way that with time bootstraps children into the adult world of accountability. When one three-year-old criticizes another – or a puppet – for not adhering to the rules (Rakoczy and Tomasello 2007; Rakoczy et al 2008), neither child has anything like the full set of psychological capacities necessary to judge whether another agent is responsible for such non-compliance. Such phenomena appear, however, to be manifestations of children's first, inchoate sense of taking part in practices in which the imputation of accountability is a constitutive element. It seems that a natural response to such imputations is to *take* responsibility oneself,

[27] Both Wellman and Moses see the full-blown concept of intention as requiring the capacity for "planning" (Wellman 1992, 290ff.; Moses 2001, 82). Planning, i.e. generating subordinate intentions on the basis of initial intentions and beliefs, often acquired as one goes along, is one central feature of deliberation, but not the whole story. Coming to initial intentions is also an important component process.

once the capacity to do so has crystallized. Becoming able to do so requires that one acquire an understanding of the concepts at work in the relevant norms. However, such an understanding will be of no use unless it feeds into at least minimal practical deliberation. Deliberation, including the deliberation-terminative opting for an answer to the question of what to do, appears to be the natural response of agents-in-the-making to being held responsible. Moreover, it is also importantly the response that in time comes to be demanded of young humans by their peers.[28]

It seems plausible that the idea of taking responsibility for what one does can only arise in the social context of being held responsible by others. It may well be as a response to such impositions that the practice of stepping back from one's motivation and reflecting on options develops. This would explain the intimate relationship between (weak) attributability and accountability. Perhaps there is a complex two-way dependency between the phylogenesis of the capacity for practical deliberation and that of deontically structured social environments. At any rate, it does seem ontogenetically fairly clear that the sense of being bound up in a world of norms precedes the development of practical reasoning. If this is correct, it is plausible that the active dimension of self-formation comes into being as a response to the imposition of all sorts of requirements, relative to which the developing agent begins to situate herself. That response is presumably to a certain degree automatic. It is however also a response that comes to be normatively expected of children as they move into the second half of the first decade of their lives.

This, I propose, is the root of the full-blown capacity to intend: children begin a process of self-formation in a normatively structured environment by learning actively to commit themselves to options, and by coming to understand that leaving epistemically available possibilities unopted for amounts to opting against them. It seems likely that these processes begin in restricted deontic contexts – in games and in response to early pressure from adults in specific areas of activity. With time, they presumably generalize and come to be understood as at work in all areas of agential activity. Again, such a development is plausibly in part the result of its being normatively expected by the child's social world. Once this understanding is in place, I think we can say that the child is well on the way to being an intender in the full sense of the term.

10.5 Taking Responsibility and Practical Rationality

In the story so far,[29] developing agents, in learning to form minimally deliberative intentions, come to take responsibility for actions in the light of demands imposed on the basis of which other agents begin to hold them responsible. Of course, just as the emotional reactions and protests of preschoolers don't amount to holding

[28] How important this is will become clear in Section 10.5.2.

[29] This story, in contrast to Bratman's narrative of creature construction, consists partly in empirical claims about the ontogeny of human persons and is as such falsifiable by research in developmental psychology.

responsible in a full sense, a young child's opting for an action after minimal deliberation is not equivalent to them assuming responsibility in a full sense either. At this age, neither party will adequately understand the concepts at issue or have any substantial conception of defeating conditions. Moreover, the mechanisms of emotion regulation and delay of gratification at the disposal of such young intenders will be too primitive for us to see their postdeliberative stands as sufficiently under their control. Nevertheless, the development of the capacity to step back, think and opt installs the machinery within which the relevant concepts will be able to play their role, which will be strengthened by control mechanisms and the operation of which will by default allow the legitimate application of the accountability apparatus. This is the ontogenetic dimension of the anchoring claim for capacities. With this developmental narrative in the background, we can now turn to the proposed non-cognitivist explanation of the *IC* requirements.

10.5.1 Decisional Intending and the IC Requirements

The explanation I wish to propose turns on the key idea of taking responsibility. Once we understand coming to intend as fulfilling this role, we can begin to see that there is some truth to the normative functionalist idea that there are intimate connections between intention and normativity. There are several such connections, although none of these are connections to which we need to advert in explaining what intending is.

The connection that I emphasised in Section 10.4 relates to valid judgements or normative reactions concerning an intender's accountability relative to social or moral norms. According to the anchoring claim for individual cases, intentions must play some role, if only in the background, if such judgements or reactions are to be fitting. In the last section, I also developed the ontogenetic conjecture that coming to take responsibility may be in part a causal result of the deontic pressure exerted by a social environment whose members are strongly disposed to hold agents, including developing agents, responsible. According to the anchoring claim for capacities, attributablity depends on the individuals to whom actions or omissions are attributed being able to deliberate and thus opt in the relevant sense.

However, none of these connections ties intending to such normative contexts. When we come to intend some action, we take responsibility for that action in as far as we take a stand that is legitimately seen as the default anchor for attributability. But we don't only put our weight behind an agential option where we think there is some probability of us being held responsible for the resulting action. We learn to do so whenever we are faced with a practical situation that, for whatever reason, raises doubts, however minimal, about how to proceed. In answering the practical question of what to do in the micro-situations of everyday life, we express our willingness to see ourselves as realisers of certain options. To come to see postdeliberative opting in these terms is to come to see ourselves as engaged in a project of self-forging. Once we begin to develop this perspective, we take what we can call *personal* *responsibility* for the actions we intend. Taking personal responsibility does not require an external bearer of standards who can measure one's options against those standards. What it does require is that the agent come to understand herself as an enduring entity able to see answers to practical questions as her own, indeed unable not to see such answers as her own.[30]

It seems likely that such a self-understanding – taking ourselves to be engaged in a project of self-forging – is developmentally secondary to the understanding that we need to respond to the normative pressure of our elders and peers. It almost certainly comes later. There is also a fair amount of plausibility to the idea that it may for empirical reasons require the prior confrontation with normative social pressure.[31]

Whether or not these genetic connections hold, we clearly can, and frequently do take personal responsibility outside any such explicitly normative contexts. In postdeliberatively opting for some particular course of action, an agent clarifies *where she stands* on the particular deliberative issue. In discussions of practical reasoning, it is frequently pointed out that, in deciding what to do, agents are not primarily thinking about themselves – for instance about their beliefs and desires – but rather about reasons. Nevertheless, deliberation is accompanied by an agent's consciousness that she is the one doing the deliberation. When she brings deliberation to a close by answering the practical question at issue, she does so in awareness that she is making that answer *her* answer.

Now, that answer might be the one she believes every other rational or reasonable agent would opt to realise in her situation. However, she may be aware that she is only opting to φ because she is bloody angry, but she is angry and she's damn well going to φ (Sects. 4.3.2 and 8.2.3). Or she may be tired of weighing up reasons and therefore go for the first option that occurs to her. Or she may provide a more classical example of akrasia, judging that she should go home and get to bed because she has to be fit early next morning, but nevertheless decide to have another beer. When an agent goes for the particular action in any of these cases, the practical option she picks out is the one she is prepared to see herself as realising. This is what makes deciding to act in a significant sense a matter of self-forging. And that remains true even if, in opting to φ, an agent is simply deciding along lines that she laid down decades ago and has never seriously rethought since. Practical deliberation is the machinery by means of which agents stamp their own practical seal of approval on the contents of optative attitudes that represent their own future behaviour.

The *IC* requirements necessarily hold, I now wish to claim, for agents who necessarily stamp their seal of approval on the actions they intend. Unlike social

[30] This is, I think, a further folk psychological seed of Existentialism (cf. Sect. 8.2.4), i.e. another feature of our self-understanding that Sartre picks up on, but to which he gives overdramatic metaphysical expression. He expresses his version of the point in his famous slogan that we are "condemned to be free" (Sartre 1946, 27).

[31] An overdramatised version of this claim plays a central role in Nietzsche and his followers.

and moral norms, which may contribute to the ontogenesis of intentional agency, but which remain external to the agency thus developed,[32] the *IC* norms are internal to full-blown agency. You can't be an agent in the full sense, that is, a being capable of intentional action, if you aren't prepared to be guided by the intention-consequential requirements.[33]

The claim can be made more precise: rational constraints on postdeliberatively opting are consequences of the two constitutive features of the seal of approval: that it is practical and is the agent's own. Briefly recalling the explanation I gave for the doxastic restrictions on decisional intending may help here. I explained these purely in terms of the practical character of deliberation and decision.

The eminently practical character of decision, as I argued in Section 6.3.3, grounds in the eminently practical character of the deliberation it brings to a close, that is, in deliberation's aim of guiding the agent's behaviour. For this reason, the same negative doxastic conditions are rationally required for both the process and its concluding attitudinal event: an agent's not believing either that he won't φ or that φ-ing is impossible for him (cf. Sect. 6.3.2). Failing to fulfil these doxastic conditions completely undermines the exercise of producing conduct-controlling attitudes. The precise sense of 'undermines' is decisive here. It is not that an agent with a belief in the relevant action's impossibility is pointlessly deliberating or deciding, even in the sense that he will necessarily fail to achieve his aim. Rather, the agent is doing something to which no point or aim can coherently be ascribed. No agent with such a belief could know what he was doing if he went through the motions of 'deliberating' or 'deciding' on the action represented in the relevant belief. For this reason, consciously failing to meet the doxastic conditions whilst deliberating is impossible.

The *IC* requirements' hold on agents grounds in reasons with the same a priori status. However, the explanation of their hold has to focus not only on the practical character of intending, but also on its ownership dimension. That is, it is decisive here that coming to intend is explicitly or implicitly to declare one's willingness to see oneself as the intention's realiser, that is, to make the commitment to the relevant action one's own practical stand on the matter at hand. The *IC* requirements specify the constraints under which agents can understand the individual stands they take on specific practical issues as components of their own practical standpoint, that is, as specifying actions for which they thus take personal responsibility.

Thus, an agent who doesn't act to realise an intention she hasn't abandoned, who fails to form subordinate intentions she believes necessary for the realisation of

[32] But compare Section 10.5.2.

[33] Nicholas Southwood's claim that the requirements of rationality are "the demands of our first-personal standpoints" is, as far as I can tell from his relatively brief elucidatory remarks, fairly close to the position I am advocating. However, his claims that "having a first-personal standpoint is a matter of being accountable to oneself" and that rationality is a matter of "what we owe to ourselves" (Southwood 2008, 28ff.) confound the attributability and accountability dimensions of responsibility. The *IC* requirements don't formulate duties to oneself, but strictures on the intelligibility of our own practical stands and thus on their capacity to anchor attributability.

an unrepudiated initial intention or who forms further intentions whose realisation seem to her incompatible with that of the initial intention she still upholds undermines the coherence of her practical perspective on the action represented in the intention. In Bratman's Frankfurt-style words, with respect to the content of my initial intention, "there is no clear place where I stand" (Bratman 2009b, 236). In the idiom I have been using, the agent's both takes and disclaims responsibility for the original intention: she is both prepared and not prepared to see herself as its realiser.

To repeat: the idea of an agent's practical standpoint is a low-key conception of "practical identity". It is not a question of picking out a subset of the agent's optative attitudes that are "her own" in a strong Frankfurtian or Watsonian – or Bratmanian – sense (cf. Sect. 8.5.1). Correspondingly, there is in standard cases nothing higher-order about the agent's stand – it is not "the fact of having [a] desire" for which we take responsibility (Frankfurt 1987, 170). Importantly, though, although no reflexive attitude to a lower-order attitude need be expressed in an intention, even in a decisional intention, taking personal responsibility does require the Lockean capacity to step back from one's unreflective desires, other wants* and evaluations. But what is then normally opted for is an action, not an attitude. Again, this weak sense in which responsibility is taken corresponds to the fact that there is no implication that the agent who has done so is necessarily accountable for the action in question.

But if intending is constitutive of such a low-key notion of practical identity, what is the force of such a weak basis for the requirements? Can we hope to explain their stringent character in this way? Indeed I think we can. The reason for agents to comply with the *IC* requirements is that without such compliance we lose our ability to understand the practical perspective we take on in intending to do anything. This is a reason that concerns the core of our self-understanding as intentional agents. It is, moreover, a matter of all-or-nothing. We don't just have a reason to comply with the requirements, but we need to do so if we are not to lose our grip on what we are doing whenever we intend. Where we are dealing with decisional intentions, that means: lose our grip on what we are doing when we bring the relevant episode of deliberation to a close.

Note that the claim is not that, in failing to conform to the requirements, we fail to uphold coherence in our practical standpoint, understood, for instance, as the set of all our practically relevant attitudes. Such a claim would provide no explanation as to why an agent should allow his attitude formation and action to be guided by the requirements in individual cases. It would be unclear why a gradual diminishment in coherence should provide a categorically reason-giving consideration. Our overall practical standpoints are certainly not maximally coherent and it is unclear why a little less coherence here or there should be decisive. Rather the claim is that the idea of taking personal responsibility in the individual case – in view of the other beliefs and intentions that are directly relevant for this case – becomes incoherent and thus, for the agent himself, incomprehensible.

How, then, does this intrinsically reason-backed 'ought' line up against the standard reasons we have for our actions – for instance, prudential or moral reasons, such as the fact that a certain action would harm either the agent herself or someone else? Where adhering to the *IC* requirements would lead to imprudent or immoral action if some initial intention is maintained, such unpalatable conclusions can usually be avoided by understanding the requirements as taking wide scope and thus leaving room for the initial intention's revision. But if there are – pace Greenspan and Watson – cases in which an initial intention is genuinely unalterable for the agent, then even within a wide-scope view the *IC* requirements may appear effectively to demand imprudence or immorality.[34] As we have seen, Bratman's solution is to claim that such cases are exceptions in which the requirements lose their hold, as the reason backing them is out of play. If, as I have argued, Bratmanian reasons of self-governance cannot be the source of the purchase of the requirements in the first place, it could appear that, as Broome has suggested might be the case, the requirements cannot be intrinsically reason-giving after all.

However, this need not follow. Although the considerations of rationality modelled by the *IC* requirements represent a form of practical necessity, rationality itself cannot tell us what the status of that rationality-internal necessity is relative to other reason-giving considerations. And indeed, I think that other considerations can override even stringent strictures of rationality. Both thiefs and racists, if they have formed thieving or racist intentions, have, according to my construal, a reason to either perform the immoral acts they have set themselves to performing or to abandon their intention.[35] Obviously, they have other kinds of reasons to opt for the latter. How important the standards of rationality are will vary relative to what else is at stake in relevant situations. However, it seems obvious that a thief's or racist's self-incomprehension is of minimal importance relative to that of their not realising their intentions.

Return again to the unalterability cases: if there are such cases and if "require" were to imply "can" in a relevant sense,[36] then there would be cases in which the requirement of executive consistency picks out the only option left, that is, the theft or the racist action. *EC* would then seem to require a thieving or racist action. However, I wouldn't find this particularly worrying.

To begin with, the requirements should be seen as internal to a particular sets of deontic standards, which can in turn be relativized. Take again the example of etiquette. Etiquette may require certain things categorically, for instance, that men wear a bow tie at the opera. But, even if you buy into etiquette, you can still believe

[34] Note that they would only do so if requirements of rationality only apply to those alternatives they specify that their addressees have the ability to satisfy. Otherwise the option that the agent drop her initial intention would not be removed from the demand's scope simply because it is unrealisable. This is perhaps not as clear as it may at first appear. Does it have to be under the control of the agent whether she has beliefs with the contents specified in the doxastic conditions of *EC*, *SI* and *IBC*?

[35] If the *IC* requirements are reason-providing, this follows from *EC*. For those who, like Broome, believe that rationality necessarily supervenes on the mind, the worry isn't – or at least isn't primarily – reasons for action, but reasons, for instance, for subordinate intentions.

[36] I expressed my doubts about this implication in Section 7.2.3 with respect to cases of sudden permanent forgetting. However, I don't want to insist on the point here. Cf. note 34 above.

that etiquette's categorical demands can be overridden. You might, for instance, think that, if you, for moral reasons, have to remove your bow tie in the middle of an opera performance – perhaps as sign to an accomplice, who then knows that the time has come to act – then etiquette will have to be sacrificed.

Now, there are obvious disanalogies between the requirements of practical rationality and the demands of etiquette. The most salient of these is that, whereas you may or may not buy into etiquette, the *IC* requirements are a priori and therefore not optional for reflective agents. There are, however, two things that importantly do not follow from this disanalogy.

First, it is no consequence of etiquette's optionality that its requirements don't provide reasons. If someone buys into etiquette and is male, then, it seems that he has a reason to wear a bow tie to the opera. Compare the rule in football not to handle the ball. The fact that you can choose whether or not to play the game is neither here nor there. If you are playing the game and you aren't a goalkeeper, you have a reason not to handle the ball. But if you could save someone's life by picking up the ball and running with it, that is a reason that is likely to override the counter-reason generated by the rule internal to the practice.

Conversely, there are two related attributes that should not be confused with, or seen as following from the inescapability of the *IC* requirements. From the fact that they are inescapable it neither follows that they are particularly strong, nor that they cannot be overridden. A requirement can plausibly have a certain level of strength within the sphere or practice, as a result of which it may be non-overridable within that sphere. This fact is in turn only relevant for the strength of the requirement tout court when combined with the importance of the relevant sphere – where importance may not always be determinable outside particular contexts. As neither etiquette nor football are plausibly of particular importance, even non-overridable norms internal to these spheres are, under normal circumstances, not going to generate reasons that need paying much attention to when genuinely serious matters are at stake.

These questions are clearly different from the question of whether a norm is escapable. How important a norm is a completely different question to whether it applies to you, even to whether it applies to you necessarily. It may apply to you, but, because of the grounds on which it demands the things it demands, pale to insignificance when confronted with demands from other sources.[37]

[37] Broome has doubts as to whether etiquette generates reasons at all (Broome 2008, 96). One ground for such doubts may lie in the worry that, if all reasons can be weighed against each other in some quantitative manner, etiquette-based reasons may, if there are enough of them, end up outweighing moral reasons. Applied to my diagnosis of the source of rational requirements: might not enough cases of potential agential incoherence end up justifying a racist action? It's difficult to see how that might work for rationality-based reasons, as these are going to be severely limited for each individual agent. More generally, I don't think that quantitative weighing is an appropriate model for all practical deliberation.

Another ground for such doubts may lie in concerns that the existence of a practice plus some form of subjective endorsement or acceptance is insufficient to make the practice reason-generating, as there are clearly morally bad practices. Buying into a practice of persecution of some minority doesn't generate reasons to persecute anyone. Like the previous point, this is a

Imagine Buck Finn (Sect. 10.3.1) feels a compulsion to turn Jim in and feels himself unable to decide not to do so. However, in spite of not revising his intention, he then doesn't form the subordinate intention to take what he believes to be the only opportunity to realise it. We think he behaves in the way he has most reason to do. I'm claiming that he would also have a reason to form that subordinate intention, a reason he necessarily has because, as a deliberative agent, he needs to make sense of the relationship between his decisional intention and his action. In comparison with what's otherwise at stake, it's a measly reason, but that doesn't mean it's not there.

The reason-giving force of the *IC* requirements grounds in a necessary feature of intentional agency, viz. the necessity of not becoming opaque to oneself in one's deliberation-based opting. There is therefore a sense in which these reasons are more basic than just about any other reasons. No practice based on other reasons could, so it seems, get off the ground if the requirements of agential coherence were not accepted.

10.5.2 Nondecisional Intention and Taking Responsibility

The reason why we are subject to principles such as *EC*, *SI* and *IBC*, then, is that intention formation that doesn't conform to them is unintelligible as a stand by which an agent takes responsibility. Or so I have argued. Clearly, however, the entire argument is dependent on the claim about the status of practical deliberation. It follows that, where intentions are not deliberatively generated, we might appear to be without any upstream genetic feature that could explain why the *IC* requirements remain equally binding. In Section 9.3, I argued that the aetiological difference between the two types of intention means that there are no doxastic conceptual constraints on nondecisional intentions. Were the difference also to result in the inapplicability of the *IC* requirements in nondecisional cases, that would be a decisive argument against the disjunctive theory I have been advocating, as we make no such everyday distinction in the norms' applicability. Fortunately, there is a good explanation – if one with a slightly complicated twist – as to why the strictures in place in the primary, decisional cases should be equally applicable to secondary, nondecisional intentions.

According to the analysis in Chapter 9, nondecisional intentions are conscious action wants* whose action-controlling proclivity is explicitly or implicitly accepted by their bearer. Explicit cases include examples in which the disposition to deliberate, registered for a split second, is overridden or ignored. In implicit cases,

big topic. But even if such a general model for the generation of practice-dependent reasons is plausible, it would be compatible with their being exclusionary conditions for its applicability, such as the moral badness of the practice.

the agent is structured in such a way that her conscious want* to φ simply doesn't trigger a deliberative disposition or deliberative want* (cf. Sect. 9.6). In such cases, according to the disjunctive upstreamist, the agent takes responsibility for her φ-ing by neither rejecting her φ-ing, nor viewing her conscious want* to φ as a matter for deliberation. Not taking either step when prompted by a conscious episode of wanting* expresses the agent's willingness to see herself as the realiser of the relevant action want*. As this is the core feature picked out by the notion of taking responsibility, this, I think we should say, suffices to trigger the applicability of the *IC* requirements.

The downstreamist reaction to this proposal is likely to be sceptical. The move may appear unacceptably ad hoc. After all, the core of the upstreamist case resides in the possibility of picking out an attitudinal event that explains those features of intending and its normative environment that functionalists characterise but don't explain. In paradigmatic cases, that event is the occurrence of a want* token that terminates at least minimal deliberation. In such cases, deliberation is decisive for intending's subjection to the *IC* requirements. In non-paradigmatic cases, the same function is now supposed to be fulfilled not only by deliberation's rejection in the situation, but also by its mere non-occurrence on the agent tokening a conscious want*. But isn't this just to claim that the occurrence of a conscious, motivationally unrivalled want* is sufficient for taking responsibility? And isn't that obviously too little?

The answer to the second question is yes. This is, however, no objection because the answer to the first question is no. I have argued for a conception of intention that is upstreamist relative to intention's causal and normative consequences. The normative dimension of the theory conceives coming to intend as taking responsibility. Upstreamism requires that there be a conceptually decisive event or episode prior to the relevant causal consequences. It also requires that there be a moment at which the agent takes responsibility for her potential action, thus triggering the applicability of the *IC* norms. That moment must be the moment at which the conscious want* occurs. Nevertheless, the responsibility taking is itself not simply the conscious tokening of the want*. Rather, for such an optative occurrence to involve a responsibility taking, it has to take place under specific conditions. The decisive point now is that those conditions are themselves normative.

Specifying the normative context in play here involves naming a normative demand that permeates a culture of intentional agency so extensively, perhaps we should say, so deeply, that it easily goes unnoticed. It is our subscription to that demand that confers on actions that nondeliberatively result from a conscious optative occurrence what I have called the mark of deliberation.

Agents who are socialised into a normative culture are, as I claimed in Section 10.4.2, confronted very early on with practices of being held responsible, a confrontation that feeds into the motivation to practically deliberate. Deliberation takes place with the aim of picking out courses of action one is prepared to see oneself as realising – at least partly in the light of the norms with which one is confronted. As agents become mature deliberators, their deliberative capacity becomes a feature with which other agents don't only reckon; it becomes a feature the exercise of which they positively demand. We assume that persons are generally able to switch into a deliberative mode if they find themselves significantly motivated to perform an action they take to be in some way problematic – where the problem in the offing can derive from the relation to norms or to the self-image of the agent.

As we know that we are all beings with limited time and energy, we obviously don't demand of each other that we always reflect before acting or settling on our intentions. However, what we *do* demand of each other is the willingness to think through the actions we are disposed to carry out, where lack of such thought looks likely to lead to problematic results. Being predominantly motivated to φ is no excuse for φ-ing where φ-ing contravenes some norm or conflicts with one's own self-image. We expect of agents that they exercise their Lockean capacity to step back from their immediate desires – either in the action situation itself or at some point in advance, at which they veto, put on hold or accept action wants* that they believe to be both motivationally strong and to pose a challenge to either social or moral norms or to their individual self-understanding. In sum, deliberative agents live in a social world structured by the normative demand to take personal responsibility.

It is in the light of this demand that either the rejection or non-triggering of an impulse to deliberate on the content of a conscious action want* counts as taking responsibility for its agential realisation. Our normative culture's interest in anchoring attributability in conscious action control is thus plausibly the reason why the concept of intention has the structure it has. Intention is not a normative concept, but nevertheless a concept that, at least in the form it has within our culture, has at least one central normative presupposition.

The upstreamist picture, then, is completed by making it clear that intention ascription, like the ascription of full blown responsibility, is not exhaustively explicable outside a normative framework. Compare the ascription of omissions. Certain non-actions are omissions because the agent has decided against carrying them out; others count as omissions because there is a norm that requires their performance, independently of whether the agent considered that norm. I think that there is a comparable mixture of descriptive and normative components in the reasons why we take episodes of motivationally unrivalled conscious wanting* to be instances of responsibility taking. In a life form in which responsibility is taken by adopting conscious optative stands, these are reasons why certain wants* count as intentions.

To be clear: this is not to say that there are normative criteria for intending, but to say that there is a normative consideration that contributes significantly to fixing the criteria for intending. That consideration, the demand that agents think through whether they are prepared to put their weight behind problematic motivationally unrivalled wants*, is central to a culture that holds agents responsible and, in order to do so, anchors attributability in conscious action control. The structure that mediates such control is the one picked out by *ID* and the attitude that satisfies the conditions *ID* names is intention. The control thus mediated necessarily has a conscious component. In both decisional and explicit non-decisional responsibility takings, that control is synchronic, whereas in implicit non-decisional takings, there is a normative presupposition that such control has been exerted at an earlier point in the process of self-forging.[38] Finally, the subjection of implicit, alongside explicit non-decisional cases to the *IC* requirements results equally from their being covered by that normative presupposition, as a result of which the relevant episode of conscious wanting* counts as a taking of responsibility.

The analogy with omissions only goes some way. First, omissions are frequently failures to act in situations specified in particular norms, whereas the normative demand I have cited is importantly of such generality that it structures our understanding of intentional agency. Second, it is implausible that omissions are necessarily understood in terms of psychological failings: they need only be non-actions.

Now, there is a category whose application plausibly presupposes the failure to instantiate certain normatively required psychological features in the light of requirements on action: the category of negligence. In cases thus classified, we see agents as accountable because they should have thought about the consequences of their behaviour in view of what some norm requires: an agent who doesn't keep to the speed limit cannot excuse himself by pointing out that he didn't attend to it. We think of the negligent actor as implicitly accepting the consequences of his actions, even if he didn't spare those consequences a thought. We justify thinking of him in these terms in the light of the norm to consider those consequences.

The relevant parallel here is to what I have been calling 'implicit responsibility taking' on the part of the non-decisional intender. We think of certain non-decisional intenders as implicitly accepting their motivation to perform some action, although they don't spare the consequences of their actions the barest thought. We justify thinking of them in these terms in the light of the norm to deliberate about our motivation where we have grounds to suspect it might be problematic.

Again, this parallel is limited and, without attention to the level at which it is applicable, could be misleading, as negligence excludes intending the result one has negligently brought about. Intended φ-ing requires a conscious optative stand pro φ-ing, whereas negligently contravening some norm *N* excludes such a stand on the question of whether to contravene *N*. If you've consciously accepted your φ-ing in the light of the belief that it will count as a contravention of *N*, you might be guilty of contravening *N* recklessly, but not of doing so negligently. To repeat: all nondecisional intentions are wants* that have occurred consciously. In cases of explicit responsibility-takings, the agent's disposition to deliberate about whether to satisfy some want* is triggered, but overridden; in cases of implicit responsibility-takings, no such disposition is triggered in the first place.

The parallel with negligence concerns the agent's lack of any further reaction to her conscious wanting* in the latter kinds of case. It thus concerns the reasons

[38] In distinguishing intention from the intentional, Bratman argues that intention has not been "shaped by our concern" with responsibility (Bratman 1987, 125). I am claiming that the source of the concept of intention is exactly such a concern.

for seeing certain wants* as responsibility takings and thus as classifying them as intentions. Whereas negligence is itself a normative category, the parallel is meant to help clarify the way in which intention's full descriptive profile is shaped by a normative presupposition. In this respect, the analogy with omissions is slightly more apt, as omission is plausibly a category with disjunctive conditions, one of which is normative, the other purely descriptive. But here again, the normative feature is a conceptual condition, whereas intention is a purely psychological concept whose full profile can, I have been arguing, only be explained on the basis of a normative presupposition that assigns certain psychological non-occurrences a particular status.

Nondecisional intentions, then, are either attitudes that store practical acceptance of a conscious motivational tendency or attitudes that we see ourselves as justified in treating as if they stored such acceptance. We see cases of the latter kind as resulting from the agent omitting to see to it that she is inclined to deliberate when wants* of a critical motivational strength occur consciously. The description according to which the agent has not taken this step is licensed by the norm that guides the introduction of agents to their developing agential responsibilities during ontogenesis. Whether we are dealing with the explicit rejection of deliberation when prompted by episodes of conscious wanting* or whether the omission in question amounts to the implicit acceptance of a motivational structure that is deliberation-immune whenever such conscious wants* occur, either way nondecisional intentions are also the products of responsibility-takings. That we can understand things in this way is dependent on our otherwise being practical deliberators with the capacity for much more pronounced forms of want* acceptance during the Lockean "suspence [of desire]".

Viewed in this way, both the explicit and the implicit acquisition of nondecisional intentions constitute the kind of stand on the practical question at issue that brings it under the umbrella of what I have been calling "weak ownership". The motivationally unrivalled conscious wants* thus accepted therefore mark the position of the agent relative to an issue in a way that counts as equivalent to taking an explicit decision on the matter. It is for this reason that they are equally subject to the *IC* requirements. The requirements concern the coherence of the agent's step of practical self-forging, that is, of what she sees herself as taking responsibility for in the situation.

The fact that conscious wants* backed by predominant motivation themselves fall under the concept is a consequence of a normative culture that anchors attributability in conscious action control. As I have argued, this is compatible with the facts that the default connection between taking responsibility and being accountable can be broken by defeating conditions and that there are secondary modes of assigning responsibility that do without conscious optative stands. If this is correct, it is conceivable that there could be a human culture that assigns responsibility in a much narrower set of cases, perhaps only where the agent has deliberated on the matter at hand. Were there to be such a culture, it would have a much narrower concept of intention, perhaps one that does without anything like the second disjunct of *ID*'s third condition. On the other hand, a culture that places no premium on conscious action control might anchor attributability in completely different, non-psychological mechanisms. It would be a strange coincidence if such a culture were to have anything like our concept of intention.

10.6 Conclusion: Intention and Normative Culture

I have argued that, although intending is a psychological state for which necessary and sufficient conditions can be provided without reference to anything normative, it is so closely bound up with our normatively structured life form that it may well, for creatures such as us, be inextricable from such normative structures. There are three decisive relations between intending and the deontic.

First, the genesis of our practical deliberation, and thus of our capacity to take our own practical stand on some deliberative issue, may empirically require the confrontation with early forms of being held responsible within a normative culture. Taken on its own, then, this point asserts what may be an empirically necessary condition of deliberation's, and thus of intention's genesis.

Second, the practical stands through which we take personal responsibility for specific actions of ours are themselves subject to the constraints codified in the *IC* requirements. These exclude mutually undermining stands, and mutually undermining stands and actions, as either of these constellations disqualifies the relevant stand-takings from counting as responsibility-takings. Intending, as defined by *ID*, counts for creatures with our life form as the taking of responsibility. It seems doubtful whether the set of conditions *ID* specifies would be soldered together to produce one psychological concept if the culture into which the bearers of the relevant psychological states are socialised didn't set a premium on conscious action control. And it seems equally unlikely that this might be the case outside a normative context which demands an anchor for attributability. In our life form, the role of anchoring attributability via responsibility taking is fulfilled by optative attitudes that satisfy *ID*. It is, however, conceivable that attributability might be anchored in some other way. This is why subjection to the *IC* requirements is not definitive of intention.

A final way in which intention is bound up with the normative concerns the specification of non-decisional intention, that is, with condition 3.2 of *ID*. In the last section, I argued that motivationally unrivalled conscious wants* also count as responsibility-takings in as far as they "bear the mark of deliberation", that is, in so far as they are accessible to deliberation for their bearers. Making clear what "accessibility" means here has involved outlining specific explanations for the satisfaction of *ID* condition 2, that is, of the agent's not seeing her φ-ing as a matter for deliberation, in non-decisional intending.

I distinguished explicit from implicit cases. In the former, the agent overrides an impulse to deliberate about her conscious want*; in the latter, the episode or occurrence of conscious wanting* triggers no such impulse. In cases of this latter kind, the reason why we take the relevant conscious, motivationally unrivalled want* to be a token of responsibility taking is normative. It derives from the demand that agents think through whether they really want to throw their weight behind potentially problematic wants*. The capacity to do so is one of the central constraints of normative address; the demand that we exercise this capacity no doubt supplements the natural motivation to do so in developing agents confronted with social practices of holding responsible. This demand is naturally understood as not calling for deliberation immediately prior to every action, but as also covering reflection on one's general motivational dispositions. This is a specific dimension of the understanding of intentional agency as bound up in a project of self-forging, an understanding which is central not only to personal self-understanding, but also to our normative culture.

It is because of this normative background that we see all episodes of motivationally unrivalled conscious wanting* as instances of responsibility taking. The norm alters the status of the non-occurrence of a deliberative impulse – in the same way that certain non-actions only acquire the status of omissions because they are non-satisfactions of some norm. Without this background, there would be no case for classifying as intentions those conscious wants* that don't trigger deliberative impulses, just as no non-deliberative cases would be thus classified if deliberation didn't play the structural role it plays in our life form.

If these counterfactuals are true, then our concept of intention is dependent on features of a specific life form. This life form may be extensionally equivalent to that of human persons, but could be either broader or narrower. There is every indication that the first of the latter possibilities is not realised as regards the inhabitants of our planet. Whether the second possibility might turn out to be a reality, intending according to *ID* being restricted to certain cultural contexts within the human life form, is a question on which data from cultural anthropology would obviously bear.

The paradigmatic form of intention could, borrowing from Aristotle, be termed "deliberative desire" (NE 1113a10-11). Our deliberative desires, thus understood, and their surrogates, motivationally unrivalled conscious wants*, are our context-specific practical-attitudinal representatives. This seems to have been why Aristotle saw ethics as primarily concerned with patterns of deliberative desire, of "prohairesis". As he remarks (NE 1111b5-6), such patterns "discriminate characters better than actions do". Part of the lesson of this last chapter is that, if virtue is to a significant extent a matter of playing host to patterns of intention that one has most moral reason to have, practical rationality involves playing host to patterns of intention that are in significant measure independent of one's reasons for intending. The relevant patterns ground instead in intending's role as the anchor of attributability, a role it cannot play where there is no answer to the question of what the agent is taking responsibility for.

References

Achtziger, Anja, and Peter M. Gollwitzer. 2008. Motivation and volition in the course of action. In *Motivation and action*, ed. Jutta Heckhausen and Heinz Heckhausen, 272–295. Cambridge: Cambridge University Press.

Achtziger, Anja, Peter M. Gollwitzer, and Paschal Sheeran. 2008. Implementation intentions and shielding goal striving from unwanted thoughts and feelings. *Personality and Social Psychology Bulletin* 34: 381–393.

Adams, Frederick. 1986. Intention and intentional action: The simple view. *Mind and Language* 1: 281–301.

Adams, Frederick, and Alfred Mele. 1989. The role of intention in intentional action. *Canadian Journal of Philosophy* 19: 511–532.

Ainslie, George. 2001. *Breakdown of will*. Cambridge: Cambridge University Press.

Allen, Colin. 1999. Animal concepts revisited: The use of self-monitoring as an empirical approach. *Erkenntnis* 51: 33–40.

Allen, Colin, and Marc Beckoff. 1997. *Species of mind. The philosophy and biology of cognitive ethology*. Cambridge, MA: MIT Press.

Alston, William P. 1967. Expressing. In *Philosophy in America*, ed. Max Black, 15–34. London: George Allen & Unwin.

Alvarez, Maria. 2007. The causalist/Anti-causalist debate in the theory of action: What it is and why it matters. In Leist 2007, 103–123.

Anscombe, G.E.M. 1957. *Intention*. Oxford: Blackwell.

Aquinas, St. Thomas. (S.th. 19). 1967. *Summa Theologiae* vol. 19: *The emotions*, Ia2ae 22–Ia2ae 30. Blackfriars: Eyre and Spottiswoode.

Aristotle (DA). *On the Soul*. Trans. J.A. Smith. In *The complete works of Aristotle*, ed. J. Barnes. Princeton: Princeton University Press 1984, vol. I.

Aristotle (DMA). *Movement of Animals*. Trans. A.S.L. Farquharson. In ed. Barnes, vol. I.

Aristotle (EE). *Eudemian Ethics*. Trans. J. Solomon. In ed. Barnes, vol. II.

Aristotle (MM). *Magna Moralia*. Trans. St.G. Stock. In ed. Barnes, vol. II.

Aristotle (NE). *Nicomachian Ethics*. Trans. W.D. Ross, revised J.O. Urmson. In ed. Barnes, vol. II.

Aristotle (Pol). *Politics*. Trans. B. Jowett. In ed. Barnes, vol. II

Aristotle (Rhet). *Rhetoric*. Trans. W.R. Roberts. In ed. Barnes, vol. II.

Armstrong, D.M. 1968. *A materialist theory of the mind*. London: Routledge & Kegan Paul.

Armstrong, D.M. 1977. The causal theory of mind. In *The nature of mind*. 1981, 16–31. Brighton: Harvester Press.

Arpaly, Nomy. 2003. *Unprincipled virtue. An inquiry into moral agency*. New York: Oxford University Press.

N. Roughley, *Wanting and Intending*, Philosophical Studies Series 123,
DOI 10.1007/978-94-017-7387-4

Astington, Janet W. 1991. Intention in the child's theory of mind. In *Children's theories of mind: mental states and social understanding*, ed. D. Frye and C. Moore, 157–172. Hillsdale, NJ: Erlbaum.

Astington, Janet W. 1993. *The child's discovery of the mind*. Cambridge, MA: Harvard University Press.

Astington, Janet W. 1996. What is theoretical about the child's theory of mind? A Vygotsian view of its development. In *Theories of theories of mind*, ed. P. Carruthers and P.K. Smith, 184–199. Cambridge: Cambridge University Press.

Astington, Janet W. 1999. The language of intention: Three ways of doing it. In *Developing theories of intention. Social understanding and self-control*, ed. P.D. Zelazo, J.W. Astington, and D.R. Olson, 295–315. Mahwah/London: Lawrence Erlbaum.

Astington, Janet W. 2001. The paradox of intention: Assessing children's metarepresentational understanding. In Malle et al. 2001, 85–103.

Astington, Janet. 2006. The developmental interdependence of theory of mind and language. In *Roots of human sociality. Culture, cognition and interaction*, ed. N. Enfield and S.C. Levinson, 179–206. New York: Berg.

Atkinson, John W. 1964. *An introduction to motivation*. Princeton: Van Nostrand.

Audi, Robert. 1973a. The concept of wanting. In Audi 1993, 35–55.

Audi, Robert. 1973b. Intending. In Audi 1993, 56–73.

Audi, Robert. 1986. Intending, intentional action and desire. In Marks 1986, 17–38.

Audi, Robert. 1988. Deliberative intentions and willingness to act: A reply to professor mele. *Philosophia* 18: 243–245.

Audi, Robert. 1991. Intention, cognitive commitment and planning. *Synthese* 86: 361–378.

Audi, Robert. 1993. *Action, intention and reason*. Ithaca/London: Cornell University Press.

Audi, Robert. 1994. Dispositional beliefs and dispositions to believe. *Noûs* 28: 419–434.

Aune, Bruce. 1967. *Knowledge, mind and nature. An introduction to the theory of knowledge and the philosophy of mind*. New York: Random House.

Bach, Kent. 1978. A representational theory of action. *Philosophical Studies* 34: 361–379.

Baier, Annette C. 1991. *A progress of sentiments. Reflections on Hume's treatise*. Cambridge, MA/London: Harvard University Press.

Bargh, John A. 1989. Conditional automaticity: Varieties of automatic influence in social perception and cognition. In *Unintended thought*, ed. J.S. Uleman and J.A. Bargh, 3–51. New York/London: Guildford.

Bargh, John A. 1990. Auto-motives. Preconscious determinants of thought and behavior. In *Handbook of motivation and cognition*, vol. 2, ed. E.T. Higgins and R.M. Sorrentino, 93–130. New York: Guildford.

Bargh, John A., and Kimberly Barndollar. 1996. Automaticity in action. The unconscious as repository of chronic goals and motives. In Gollwitzer, Bargh 1996, 457–481.

Bargh, John A., and Peter M. Gollwitzer. 1994. Environmental control of goal-directed action: Automatic and strategic contingencies between situations and behavior. *Nebraska Symposium on Motivation* 41: 71–124.

Bargh, John A., Peter M. Gollwitzer, Annette Lee-Chai, Kimberly Barndollar, and Roman Trötschel. 2001. The automated will: Nonconscious activation and pursuit of behavioral goals. *Journal of Personality and Social Psychology* 81: 1014–1027.

Bartsch, Karen, and Henry A. Wellman. 1995. *Children talk about the mind*. New York/Oxford: Oxford University Press.

Baumann, Peter. 2000. *Die Autonomie der Person*. Paderborn: Mentis.

Baumeister, Roy F., and Leonard S. Newman. 1994. Self-regulation of cognitive inference and decision processes. *Personality and Social Psychology Bulletin* 20: 3–19.

Bayne, Timothy J., and Neil Levy. 2006. The feeling of doing: Deconstructing the phenomenology of agency. In *Disorders of volition*, ed. N. Sebanz and W. Prinz, 49–68. Cambridge, MA: MIT Press.

Beardsley, Monroe C. 1978. Intending. In *Values and morals*, ed. A.I. Goldman and J. Kim, 163–184. Dordrecht: Reidel.

Beckmann, Jürgen, and Peter M. Gollwitzer. 1987. Deliberative versus implemental states of mind: The issue of impartiality in predecisional and postdecisional information processing. *Social Cognition* 5: 259–279.

Bentham, Jeremy. (PML). 1948. *An introduction to the principles of morals and legislation*. New York: Hafner.

Berlyne, D.E. 1965. *Structure and direction in thinking*. New York: Wiley.

Bishop, John. 1989. *Natural agency. An essay on the causal theory of action*. Cambridge: Cambridge University Press.

Bittner, Ruediger. 2001. *Doing things for reasons*. Oxford: Oxford University Press.

Blackburn, Simon. 1984. *Spreading the word. Groundings in the philosophy of language*. Oxford: Clarendon Press.

Block, Ned. 1980a. *Readings in philosophy of psychology*. Cambridge, MA: Harvard University Press.

Block, Ned. 1980b.Troubles with functionalism. In Block 1980a, 268–305.

Block, Ned. 1994. Functionalism (2). In *A companion to the philosophy of mind*, ed. S. Guttenplan, 323–332. Oxford: Blackwell.

Bobonich, Christopher. 2002. *Plato's Utopia Recast: His later ethics and politics*. Oxford: Oxford University Press.

Boesch, Christophe, and Hedwige Boesch. 1984. Mental map in wild chimpanzees: An analysis of hammer transports for nut cracking. *Primates* 25: 160–170.

Boyle, Matthew, and Douglas Lavin. 2010. Goodness and desire. In Tenenbaum, 2010a. 161–201.

Brand, Myles. 1983. Intending and believing. In Tomberlin. 1983, 171–193.

Brand, Myles. 1984. *Intending and acting. Toward a naturalized action theory*. Cambridge, MA/London: MIT Press.

Brand, Myles. 1986. Intentional actions and plans. *Midwest Studies in Philosophy* X: 212–230.

Brand, Myles. 1989. Proximate causation of action. *Philosophical Perspectives* 3: 423–442.

Brandom, Robert. 1994. *Making it explicit. Reasoning, representing and discursive commitment*. Cambridge, MA/London: Harvard University Press.

Brandt, Richard. 1979. *A theory of the good and the right*. Oxford: Clarendon Press.

Brandt, Richard, and Jaegwon Kim. 1963. Wants as explanations of actions. *Journal of Philosophy* LX: 425–435.

Bratman, Michael E. 1983. *Castañeda's theory of thought and action*, originally in Tomberlin 1983. Reprinted in Bratman 1999a, 225–249.

Bratman, Michael E. 1984. Two faces of intention. In *The philosophy of action*, ed. A. Mele. 1997, 178–203. Oxford/New York: Oxford University Press.

Bratman, Michael E. 1985. Davidson's theory of intention. In Bratman 1999a, 209–224.

Bratman, Michael E. 1987. *Intention, plans and practical reason*. Cambridge, MA/London: Harvard University Press.

Bratman, Michael E. 1991. Cognitivism about practical reason. In Bratman 1999a, 250–264.

Bratman, Michael E. 1995. Planning and temptation. In Bratman 1999a, 35–57.

Bratman, Michael E. 1996. Identification, decision and treating as a reason. In Bratman 1999a, 185–206.

Bratman, Michael E. 1999a. *Faces of intention. Selected essays on intention and agency*. Cambridge: Cambridge University Press.

Bratman, Michael E. 1999b. Introduction: Planning agents in a social world. In Bratman 1999a, 1–12.

Bratman, Michael E. 2000a. Reflection, planning and temporally extended agency. In Bratman 2007a, 21–46.

Bratman, Michael E. 2000b. Valuing and the will. In Bratmann 2007, 47–67.

Bratman, Michael E. 2001. Two problems about human agency. In Bratman 2007a, 89–105.

Bratman, Michael E. 2004. Three theories of self-governance. In Bratman 2007a, 222–253.

Bratman, Michael E. 2005. Nozick, free will and the problem of agential authority. In Bratman 2007a, 127–136.

Bratman, Michael E. 2007a. *Structures of agency: Essays*. Oxford/New York: Oxford University Press.

Bratman, Michael E. 2007b. Introduction. In Bratman 2007a, 3–18.

Bratman, Michael E. 2008. Normative thinking and planning, individual and shared. In Allan Gibbard. *Reconciling our aims. In search of bases for ethics*, ed. B. Stroud, 91–101. Oxford/New York: Oxford University Press.

Bratman, Michael E. 2009a. Intention, belief, practical, theoretical. In *Spheres of reason*, ed. S. Robertson, 28–59. Oxford: Oxford University Press.

Bratman, Michael E. 2009b. Intention rationality. *Philosophical Explorations* 12: 227–241.

Bratman, Michael E. 2009c. Intention, practical rationality and self-governance. *Ethics* 119: 411–443.

Bratman, Michael E. 2012. Time, rationality and self-governance. *Philosophical Issues* 22: 73–88.

Bratman, Michael E. 2014. *Shared agency. A planning theory of acting together*. Oxford: Oxford University Press.

Bricke, John. 1996. *Mind and morality. An examination of Hume's moral psychology*. Oxford: Clarendon Press.

Broad, Charlie Dunbar. 1954. Emotion and sentiment. *Journal of Aesthetics and Art Criticism* XIII: 203–214.

Broad, Charlie Dunbar. 1962. *The mind and its place in nature*. London: Routledge & Kegan Paul.

Broad, Charlie Dunbar. 1985. *Ethics*. Dordrecht: Nijhoff.

Broome, John. 1999. Normative requirements. *Ratio (new series)* XII: 398–419.

Broome, John. 2001. Are intentions reasons? And how should we cope with incommensurable values? In *Practical rationality and preference: Essays for David Gauthier*, ed. C.W. Morris and A. Ripstein, 98–120. Cambridge: Cambridge University Press.

Broome, John. 2002. Practical reasoning. In *Reason and nature. Essays in the theory of rationality*, ed. J.L. Bermùdez and A. Miller, 85–111. Oxford: Clarendon Press.

Broome, John. 2004. Reasons. In Pettit, Scheffler, Smith, Wallace, 2006, 28–55.

Broome, John. 2005. Does rationality give us reasons? *Philosophical Issues* 15: 321–337.

Broome, John. 2007a. Does rationality consist in responding correctly to reasons? *Journal of Moral Philosophy* 4: 349–374.

Broome, John. 2007b. Is rationality normative? *Disputatio* 2(23): 161–178.

Broome, John. 2007c. Wide or narrow scope? *Mind* 116: 359–370.

Broome, John. 2008. Reply to Southwood, Kearns and Star, and Cullity. *Ethics* 119: 96–108.

Broome, John. 2009. The unity of reasoning. In *Spheres of reason. New essays in the philosophy of normativity*, ed. S. Robertson, 62–92. Oxford/New York: Oxford University Press.

Broome, John. 2010. Rationality. In *A companion to the philosophy of action*, ed. T. O'Connor and C. Sandis, 285–292. Chichester: Wiley-Blackwell.

Broome, John. 2013a. *Rationality through reasoning*. Chichester: Wiley Blackwell.

Broome, John. 2013b. Enkrasia. *Organon F* 20: 425–436.

Bruner, Jerome S., and Cecile B. Goodman. 1947. Value and need as organizing factors in perception. In *Beyond the information given. Studies in the psychology of knowing*, ed. J.S. Bruner, 43–56. New York: Norton.

Bruner, Jerome S., and Leo Postman. 1947–1948. Emotional selectivity in perception and reaction. *Journal of Personality* 16: 69–77.

Bruner, Jerome S., and Leo Postman. 1948. Symbolic value as an organizing factor in perception. *Journal of Social Psychology* 27: 203–208.

Brunero, John. 2010. The scope of rational requirements. *The Philosophical Quarterly* 60: 28–49.

Brunero, John. 2012. Instrumental rationality, symmetry and scope. *Philosophical Studies* 157: 125–140.

Brunero, John. 2013. Rational *Akrasia. Organon F* 20: 546–566.

Byrne, Richard. 1995. *The thinking ape. Evolutionary origins of intelligence*. Oxford: Oxford University Press.

Call, Josep, and Michael Tomasello. 2008. Does the chimpanzee have a theory of mind? 30 years later. *Trends in Cognitive Science* 12: 187–192.

Camp, Elisabeth. 2009. Putting thoughts to work: Concepts, systematicity and stimulus-independence. *Philosophy and Phenomenological Research* LXXVIII: 275–311.

Cantril, Hadley, and William A. Hunt. 1932. Emotional effects produced by the injection of adrenalin. *American Journal of Psychology* 44: 300–307.

Carnap, Rudolf. 1935. *Philosophy and logical syntax*. London: Kegal Paul, Trench, Trubner & Co. Ltd.

Carnap, Rudolf. 1963. Replies and systematic expositions. In *The philosophy of Rudolf Carnap*, ed. P.A. Schilpp, 859–1013. La Salle: Open Court.

Carruthers, Peter. 2009. Invertebrate concepts confront the generality constraint (and Win). In Lurz 2009, 89–107.

Castañeda, Hector-Neri. 1967. Indicators and quasi-indicators. *American Philosophical Quarterly* 4: 85–100.

Castañeda, Hector-Neri. 1975. *Thinking and doing. The philosophical foundations of institutions*. Dordrecht: Reidel.

Chartrand, Tanya, and John Bargh. 2002. Nonconscious motivations: Their activation, operation and consequences. In *Self and motivation. Emerging psychological perspectives*, ed. A. Tesser, D.A. Stapel, and J.V. Wood, 13–41. Washington, DC: American Psychological Association.

Chisholm, Roderick M.1966. Freedom and action. In Lehrer 1966, 11–44.

Chisholm, Roderick M. 1970. The structure of intention. *Journal of Philosophy* LXVII: 633–647.

Churchland, Paul M. 1981. Eliminative materialism and the propositional attitudes. *Journal of Philosophy* LXXVIII: 67–90.

Clayton, N.S., J. Dally, J. Gilbert, and A. Dickinson. 2005. Food-caching by western scrub-jays (Aphelocoma californica) is sensitive to the conditions at recovery. *Journal of Experimental Psychology: Animal Behavior Processes* 31: 115–124.

Clayton, N.S., N. Emery, and A. Dickinson. 2006. The rationality of animal memory: Complex caching strategies of western scrub jays. In Hurley/Nudds 2006, 197–216.

Collins, Arthur W. 1969. Unconscious belief. *Journal of Philosophy* 66: 667–680.

Cooper, John M. 1984. Plato's theory of human motivation. *History of Philosophy Quarterly* 1: 3–21.

Cooper, John M. 1999. *Reason and emotion*. Princeton: Princeton University Press.

Copp, David. 1995. *Morality, normativity and society*. New York/Oxford: Oxford University Press.

Copp, David. 2001. Realist-expressivism: A neglected option for moral realism. *Social Philosophy and Policy* 18: 1–43.

Cross, R.C., and A.D. Woozley. 1966. *Plato's republic. A philosophical commentary*. London/New York: MacMillan/St. Martin's Press.

Csikszentmihalyi, Mihaly. 1990. *Flow: The psychology of optimal experience*. New York: Harper-Collins.

Cullity, Garrett. 2008. Decisions, reasons and rationality. *Ethics* 119: 57–95.

Dancy, Jonathon. 1993. *Moral reasons*. Oxford: Blackwell.

Dancy, Jonathon. 2000. *Practical reality*. Oxford: Oxford University Press.

Dancy, Jonathon. 2004. *Ethics without principles*. Oxford: Oxford University Press.

Darwall, Stephen L. forthcoming. Empathy and reciprocating attitudes. In *Forms of fellow feeling. Empathy, sympathy, concern and moral agency*, ed. N. Roughley, T. Schramme. Cambridge: Cambridge University Press.

Davidson, Donald. 1963. Actions, reasons and causes. In Davidson 1980, 3–20.

Davidson, Donald. 1970. How is weakness of the will possible? In Davidson 1980, 21–42.

Davidson, Donald. 1971. Agency. In Davidson 1980, 43–62.

Davidson, Donald. 1975. Thought and talk. In *Inquiries into truth and interpretation*, 155–170. Oxford: Clarendon Press.

Davidson, Donald. 1978. Intending. In Davidson 1980, 83–102

Davidson, Donald. 1980. *Essays on actions and events*. Oxford: Clarendon Press.

Davidson, Donald. 1982a. Rational animals. In Davidson 2001, 95–105.

Davidson, Donald. 1982b. Paradoxes of irrationality. In *Philosophical essays on Freud*, ed. R. Wollheim and J. Hopkins, 289–305. Cambridge: Cambridge University Press.

Davidson, Donald. 1985. Incoherence and irrationality. *Dialectica* 39: 345–354.

Davidson, Donald. 1997. The emergence of thought. In Davidson 2001, 123–134.

Davidson, Donald. 2001. *Subjective, intersubjective, objective*. Oxford: Clarendon Press.

Davis, Wayne A. 1982. A causal theory of enjoyment. *Mind* XCI: 240–256.

Davis, Wayne A. 1984. A causal theory of intending. In *The philosophy of action*, ed. A.R. Mele. 1997, 131–148. Oxford/New York: Oxford University Press.

Davis, Wayne A. 1986. Two senses of desire. In Marks. 1986, 63–82.

Deigh, John. 2010. Concepts of emotion. In *The Oxford handbook of philosophy of emotion*, ed. P. Goldie, 17–40. Oxford: Oxford University Press.

del Corral, Miranda. 2013. Against normative judgement internalism. *Organon F* 20: 567–587.

Dennett, Daniel C. 1984. *Elbow room. The varieties of free will worth wanting*. Cambridge, MA: MIT Press.

de Sousa, Ronald. 1974. The good and the true. *Mind* LXXXIII: 534–551.

de Sousa, Ronald. 1987. *The rationality of emotion*. Cambridge, MA/London: MIT Press.

Donagan, Alan. 1987. *Choice*. London/New York: Routledge & Kegan Paul.

Dretske, Fred. 1988. *Explaining behavior. Reasons in a world of causes*. Cambridge, MA/London: MIT Press.

Dretske, Fred. 1995. *Naturalizing the mind*. Cambridge, MA/London: MIT Press.

Duffy, Elizabeth. 1941. The conceptual categories of psychology: A suggestion for revision. *Psychological Review* 48: 177–203.

Duffy, Elizabeth. 1951. The concept of energy mobilization. *Psychological Review* 58: 30–40.

Dummett, Michael. 1972. Can analytic philosophy be systematic, and ought it to be? In *Truth and other enigmas*, 437–458. London: Duckworth.

Dummett, Michael. 1993. *The origins of analytic philosophy*. Cambridge, MA: Harvard University Press.

Duncker, Karl. 1940–1941. On pleasure, emotion and striving. *Philosophy and Phenomenological Research* I: 391–430.

Elster, Jon. 1985. The nature and scope of rational-choice explanation. In Lepore, McLaughlin 1985, 60–72.

Esken, Frank. 2012. Early forms of metacognition in human children. In *Foundations of metacognition*, ed. Michael J. Beran et al., 134–145. Oxford: Oxford University Press.

Fehige, Christoph. 2001. Instrumentalism. In *Varieties of practical reasoning*, ed. E. Millgram, 49–76. Cambridge, MA: MIT Press.

Feinberg, Joel. 1968. Action and responsibility. In *The philosophy of action*, ed. A.R. White, 95–119. Oxford: Oxford University Press.

Ferguson, Eva Dreikurs. 2000. *Motivation. A biosocial and cognitive integration of motivation and emotions*. New York/Oxford: Oxford University Press.

Ferrero, Luca. 2012. Diachronic constraints of practical rationality. *Philosophical Issues* 22: 144–164.

Fink, J., ed. 2013. Special issue on *The nature of the enkratic requirement of rationality. Organon F* 20 (4): 422–631.

Finkelstein, David H. 1999. On the distinction between conscious and unconscious states of mind. *American Philosophical Quarterly* 36: 79–100.

Finkelstein, David H. 2003. *Expression and the inner*. Cambridge, MA: Harvard University Press.

Finlay, S., and M. Schroeder. 2012. Reasons for action: Internal vs external. *Stanford Encyclopedia of Philosophy*. Last accessed 8 Apr 2015: http://plato.stanford.edu/entries/reasons-internal-external/

Fishbein, M., and I. Ajzen. 1975. *Belief, attitude, intention and behaviour. An introduction to theory and research*. Reading: Addison-Wesley.

Fletcher, George P. 1971. The theory of criminal negligence: A comparative analysis. *University of Pennsylvania Law Review* 119: 401–438.

Fodor, Jerry A. 1978. Propositional attitudes. *The Monist* 61: 501–523.

Fodor, Jerry A. 1987. *Psychosemantics. The problem of meaning in the philosophy of mind*. Cambridge, MA: MIT Press.

Franken, Robert E. 1994. *Human motivation*. Belmont: Wadsworth.

Frankfurt, Harry G. 1971. Freedom of the will and the concept of a person. In Frankfurt 1988, 11–25.

Frankfurt, Harry G. 1976. Identification and externality. In *ibid.*, 58–68.

Frankfurt, Harry G. 1987. Identification and wholeheartedness. In *ibid.*, 159–176.

Frankfurt, Harry G. 1988. *The importance of what we care about: Philosophical essays*. Cambridge: Cambridge University Press.

Frankfurt, Harry G. 1992. The faintest passion. In *Necessity, volition and love*, 95–107. Cambridge: Cambridge University Press 1999.

Frege, Gottlob. 1879. *Conceptual notation and related articles*, ed. T.W. Bynum. 1972, 101–208.

Frege, Gottlob. 1918–1919. Thoughts. In *Collected papers on mathematics, logic and philosophy*, ed. Brian McGuinness. 1984, 351–372. Oxford: Blackwell.

Frege, Gottlob. 1979. Separating a thought from its trappings. In *Posthumous writings*, ed. H. Hermes, F. Kambartel, and F. Kaulbach, 138–149. Oxford: Blackwell.

Geach, P.T. 1957. *Mental acts*. London: Routledge & Kegan Paul.

Geach, P.T. 1965. Assertion. *The Philosophical Review* LXXIV: 449–465.

Geach, P.T. 1966. Dr. Kenny on practical inference. *Analysis* 26: 76–79.

Gibbs, Raymond W. Jr. 2001. Intentions as emergent products of social interactions. In Malle et al 2001, 105–122.

Gilbert, Margaret. 1989. *On social facts*. Princeton, NJ: Princeton University Press.

Ginet, Carl. 1970. Can the will be caused? In *Determinism, free will and moral responsibility*, ed. R. Dworkin, 119–126. Englewood Cliffs: Prentice-Hall.

Glock, Hans-Johann. 1999. Animal minds: Conceptual problems. *Evolution and Cognition* 5: 174–188.

Glock, Hans-Johann. 2000. Animals, thoughts and concepts. *Synthese* 123: 35–64.

Glock, Hans-Johann. 2010. Can animals judge? *Dialectica* 64: 11–33.

Goldie, Peter. 2000. *The emotions. A philosophical exploration*. Oxford: Clarendon.

Goldman, Alvin. 1970. *A theory of human action*. Englewood Cliffs: Prentice-Hall.

Goldman, Alvin. 1976. The volitional theory revisited. In *Action theory*, eds. M. Brand, and D. Walton. 1976, 67–84. Dordrecht: Reidel.

Gollwitzer, Peter M. 1990. Action phases and mind-sets. In *Handbook of motivation and cognition*, vol. 2, ed. E.T. Higgins and R.M. Sorrentino, 52–92. New York/London: Guildford Press.

Gollwitzer, Peter M. 1991. *Abwägen und Planen. Bewusstseinslagen in verschiedenen Handlungsphasen*. Göttingen/Toronto/Zurich: Hogrefe.

Gollwitzer, Peter M. 1996. The volitional benefits of planning. In Gollwitzer, Bargh 1996, 287–312.

Gollwitzer, Peter M. 1999. Implementation intentions. Strong effects of simple plans. *American Psychologist* 54: 493–503.

Gollwitzer, Peter M. 2003. Why we thought that action mind-sets affect illusions of control. *Psychological Inquiry* 14: 261–269.

Gollwitzer, Peter M., and J. Bargh. 1996. *The psychology of action: Linking cognition and motivation to behavior*. New York: Guilford Press.

Gollwitzer, Peter M., and Ute Bayer. 1999. Deliberative versus implemental mindsets in the control of action. In *Dual-process theories in social psychology*, ed. S. Chaiken and Y. Trope, 403–422. New York: Guildford Press.

Gollwitzer, Peter M., and Ronald F. Kinney. 1989. Effects of deliberative and implemental mindsets on illusion of control. *Journal of Personality and Social Psychology* 56: 531–542.

Gollwitzer, Peter M., and Bernd Schaal. 1998. Metacognition in action: The importance of implementation intentions. *Personality and Social Psychology Review* 2: 124–136.

Gollwitzer, Peter M., and Paschal Sheeran. 2008. Implementation intentions. In *Health behavior constructs: Theory, measurement and research.* National Institutes of Health, National Cancer Institute.

Gollwitzer, Peter M., Heinz Heckhausen, and Birgit Steller. 1990. Deliberative versus implemental mind-sets: Cognitive tuning toward congruous thoughts and information. *Journal of Personality and Social Psychology* 59: 1119–1127.

Gollwitzer, Peter M., Juan D. Delius, and Gabriele Oettingen. 2000. Motivation. In *International handbook of psychology*, ed. K. Pawlik and M.R. Rosenzweig, 196–206. London: Sage.

Gosling, J.C.B. 1969. *Pleasure and desire. The case for hedonism reviewed*. Oxford: Clarendon Press.

Greenspan, Patricia S. 1978. Behavior control and freedom of action. *The Philosophical Review* 87: 225–240.

Grice, H.P. 1967/87. Logic and conversation. In *Studies in the way of words*. Cambridge, MA: Harvard University Press 1989, 3–143.

Grice, H.P. 1971. Intention and uncertainty. *Proceedings of the British Academy* LVII: 263–279.

Griffiths, Paul E. 1997. *What emotions really are. The problem of psychological categories*. Chicago/London: University of Chicago Press.

Groves, P.M., and G.V. Rebec. 1988. *Introduction to biological psychology*. Dubuque: Wm. C. Brown Publishers.

Haddock, Adrian, and Fiona Macpherson. 2008. Introduction: Varieties of disjunctivism. In *Disjunctivism: Perception, action, knowledge*, ed. A. Haddock and F. Macpherson, 1–24. Oxford: Oxford University Press.

Hale, Courtney Melinda, and Helen Tager-Flussberg. 2003. The influence of language on theory of mind: A training study. *Developmental Science* 6: 346–359.

Hampshire, Stuart. 1970. *Thought and action*. London: Chatto & Windus.

Hampshire, Stuart. 1975. *Freedom of the individual*. London: Chatto & Windus.

Hampshire, Stuart, and H.L.A. Hart. 1958. Decision, intention and certainty. *Mind* LXVII: 1–12.

Hanks, Peter W. 2007. The content-force distinction. *Philosophical Studies* 134: 141–164.

Hare, Richard. 1952. *The language of morals*. Oxford: Clarendon Press.

Hare, Richard. 1963. *Freedom and reason*. Oxford: Clarendon Press.

Hare, Richard. 1968. Wanting: Some pitfalls. In *Practical inferences*. 1971, 44–58. London/Basingstoke: MacMillan.

Hare, Richard. 1970. Meaning and speech acts. *Philosophical Review* 79: 3–24.

Hare, B., J. Call, B. Agnetta, and M. Tomasello. 2000. Chimpanzees know what conspecifics do and do not see. *Animal Behaviour* 59: 771–785.

Hare, B., J. Call, and M. Tomasello. 2001. Do chimpanzees know what conspecifics know? *Animal Behaviour* 61: 139–151.

Harman, Gilbert. 1975/76. Practical reasoning. *Review of Metaphysics* 29: 431–463.

Harman, Gilbert. 1986a. *Change in view. Principles of reasoning*. Cambridge, MA/London: MIT Press.

Harman, Gilbert. 1986b. Willing and intending. In *Philosophical grounds of rationality. Intentions, categories, ends*, ed. R.E. Grandy and R. Warner, 363–380. Oxford: Clarendon Press.

Hebb, D.O. 1955. Drives and the C.N.S (Conceptual Nervous System). *Psychological Review* 62: 243–254.

Heckhausen, Heinz. 1987. Perspektiven einer Psychologie des Wollens. In *Jenseits des Rubikon. Der Wille in den Humanwissenschaften*, ed. H. Heckhausen, P.M. Gollwitzer, and F.E. Weinert, 121–142. Berlin: Springer.

Heckhausen, Heinz. 1991. *Motivation and action*. Berlin: Springer.

Heckhausen, Heinz, and Juergen Beckmann. 1990. Intentional action and action slips. *Psychological Review* 97: 36–48.

Heckhausen, Heinz, and Peter M. Gollwitzer. 1986. Information processing before and after the formation of an intent. In *Human memory and cognitive capabilities*, ed. F. Klix and H. Hagendorf, 1071–1082. North-Holland: Elsevier.

Heckhausen, Heinz, and Peter M. Gollwitzer. 1987. Thought contents and cognitive functioning in motivational versus volitional states of mind. *Motivation and Emotion* 11: 101–120.

Heckhausen, Heinz, and Julius Kuhl. 1985. From wishes to action: The dead ends and short cuts or the long way to action. In *Goal-directed behavior: The concept of action in psychology*, ed. M. Frese and J. Sabini, 134–159. Hillsdale/London: Erlbaum.

Hempel, Carl G. 1935. The logical analysis of psychology. In Block 1980, 14–23.

Higgins, Raymond L. 1990. Self-handicapping. Historical roots and contemporary branches. In *Self-handicapping. The paradox that isn't*, eds. R.L. Higgins, C.R. Snyder, and S. Berglas, 1–35. New York/London: Plenum Press 1990.

Higgins, E.T. 1996. Knowledge activation: Accessibility, applicability and salience. In *Social psychology. Handbook of basic principles*, ed. E.T. Higgins and A.W. Kruglanski, 133–168. New York/London: The Guildford Press.

Hill, Thomas E., Jr. 1986. Weakness of will and character. In *Autonomy and self-respect*, 118–137. Cambridge: Cambridge University Press 1991.

Hobbes, Thomas. (L). 1977. *Leviathan*. New York/London: Norton.

Hobbes, Thomas. (DH). 1993. *De Homine*. In *Man and citizen*. Indianapolis: Hackett.

Hobbes, Thomas. (HN). 1994. *Of human nature*. In *The elements of law, natural and politic*. Oxford: Oxford University Press.

Hofstadter, Albert, and J.C.C. McKinsey. 1939. On the logic of imperatives. *Philosophy of Science* VI: 446–457.

Holton, Richard. 1999. Intention and weakness of will. *Journal of Philosophy* 46: 241–262.

Holton, Richard. 2009. *Willing, wanting, waiting*. Oxford: Clarendon Press.

Holton, Richard. 2015. Primitive self-ascription: Lewis on the de se. In *A comparison to David Lewis*, ed. B. Loewer and J. Schaffer, 399–411. Malden, MA: Wiley Blackwell.

Hornsby, Jennifer. 1980. *Actions*. London: Routledge & Kegan Paul.

Hornsby, Jennifer. 2008. A disjunctive conception of acting for reasons. In *Disjunctivism: Perception, action, knowledge*, ed. A. Haddock and F. Macpherson, 244–261. Oxford: Oxford University Press.

Hornsby, Jennifer. 2010. The standard story of action: An exchange (2). In *Causing human action*, ed. J.H. Aguilar and A.A. Buckareff, 57–68. Cambridge, MA: MIT Press.

Hull, Clark L. 1930. Knowledge and purpose as habit mechanisms. *Psychological Review* 37: 511–525.

Hull, Clark L. 1943a. *Principles of behavior. An introduction to behavior theory*. New York: Appleton-Century-Crofts.

Hull, Clark L. 1943b. The problem of intervening variables in molar behavior theory. *Psychological Review* 50: 273–291.

Humberstone, I.L. 1992. Direction of fit. *Mind* 101: 59–83.

Hume, David. (T). 1978. *A treatise of human nature*. Oxford: Clarendon.

Humle, Tatyana. 2003. *Culture and variation in wild Chimpanzee behaviour: A study of three communities in West Africa.* Ph.D. thesis, University of Stirling. www.greencorridor.info/data/Culture_and_variation_in_wild_chimanzee_behaviour__a_study_of_three_communities_in_West_Africa.pdf. Last accessed May 2012.

Hurley, Susan. 2006. Making sense of animals. In Hurley/Nudds 2006, 139–171.

Hurley, Susan, and Mathew Nudds (eds.). 2006. *Rational animals?* Oxford: Oxford University Press.

Irwin, Terence. 1977. *Plato's moral theory. The early and middle dialogues*. Oxford: Clarendon Press.

Irwin, Terence. 1995. *Plato's ethics*. New York/Oxford: Oxford University Press.

Jackson, Frank. 1984. Weakness of will. *Mind* XCIII: 1–18.

James, William. 1890.*The principles of psychology*. Cambridge, MA/London: Harvard University Press 1981.

Kane, Robert. 1996. *The significance of free will*. New York/Oxford: Oxford University Press.

Kant, Immanuel. (GMS). 1911. *Grundlegung zur Metaphysik der Sitten*, Prussian Academy Edition vol. IV. Berlin: Reimer.

Kaufman, Arnold S. 1966. Practical decision. *Mind* LXXV: 25–44.

Kavka, Gregory S. 1983. The toxin puzzle. *Analysis* 43: 33–36.

Kenny, Anthony. 1963. *Action, emotion and will*. Bristol: Thoemmes 1994.

Kenny, Anthony. 1973. *The anatomy of the soul. Historical essays in the philosophy of mind*. Oxford: Blackwell.

Kenny, Anthony. 1975. *Will, freedom and power*. Oxford: Blackwell.

Kenny, Anthony. 1989. *The metaphysics of mind*. Oxford: Oxford University Press.

Kenward, Ben. 2012. Over-imitating preschoolers believe unnecessary actions are normative and enforce their performance by a third party. *Journal of Experimental Child Psychology* 112: 195–207.

Keupp, Stefanie, Tanya Behne, and Hannes Rakoczy. 2013. Why do children overimitate? Normativity is crucial. *Journal of Experimental Child Psychology* 116: 392–406.

Kim, Jaegwon. 1976. Intention and practical inference. In *Essays on explanation and understanding*, ed. J. Manninen and R. Tuomela, 249–269. Dordrecht/Boston: Reidel.

Kim, Jaegwon. 1993. Psychophysical supervenience. In *Supervenience and mind. Selected philosophical essays*, 175–193. Cambridge: Cambridge University Press.

Kim, Jaegwon. 1996. *Philosophy of mind*. Boulder: Westview.

Kolodny, Niko. 2005. Why be rational? *Mind* 114: 509–563.

Kolodny, Niko. 2007. State or process requirements? *Mind* 116: 371–385.

Korsgaard, Christine. 1996. *The sources of normativity*. Cambridge: Cambridge University Press.

Korsgaard, Christine. 1997. The normativity of instrumental reason. In *Ethics and practical reason*, ed. G. Cullity and B. Gaut, 215–254. Oxford: Clarendon Press.

Lacey, John I. 1967. Somatic response patterning and stress: Some revisions of activation theory. In *Psychological stress. Issues in research*, ed. M.H. Appley and R. Trumbull, 14–37. New York: Appleton Crofts.

Langton, Rae. 2004. Intention as faith. In *Agency and action*, ed. J. Hyman and H. Steward, 243–258. Cambridge: Cambridge University Press.

Lehrer, Keith (ed.). 1966. *Freedom and determinism*. New York: Random House.

Leist, Anton (ed.). 2007. *Action in context*. Berlin/New York: de Gruyter.

Lepore, E., and B.P. McLaughlin. 1985. *Action and events. Perspectives on the philosophy of Donald Davidson*. Oxford: Blackwell.

Lewis, David. 1966. An argument for the identity theory. *Journal of Philosophy* LXIII: 17–25.

Lewis, David. 1972. Psychophysical and theoretical identifications. *Australasian Journal of Philosophy* 50: 249–258.

Lewis, David. 1978. Mad pain and Martian pain. In *Philosophical Papers*. 1983, 122–130. New York/Oxford: Oxford University Press.

Libet, Benjamin. 1985. Unconscious cerebral initiative and the role of conscious will in voluntary action. *The Behavioural and Brain Sciences* 8: 529–566.

Locke, John. (E). 1975. *An essay concerning human understanding*. Oxford: Clarendon.

Lockery, Shawn, and Stephen Stich. 1989. Prospects for animal models of mental representation. *The International Journal of Comparative Psychology* 2: 157–173.

Ludwig, Kirk. 1992. Impossible doings. *Philosophical Studies* 65: 257–281.

Luebbe, Hermann. 1965. Zur Theorie der Entscheidung. In *Theorie und Entscheidung. Studien zum Primat praktischer Vernunft*. 1971. Freiburg: Rombach.

Lumer, Christoph. 2005. Intentions are optimality beliefs – But optimizing what? *Erkenntnis* 62: 235–262.

Lumer, Christoph. 2007. An empirical theory of practical reasons and its use for practical philosophy. In Lumer, Nannini 2007, 157–186.

Lumer, Christoph, and Sandro Nannini (eds.). 2007. *Intentionality, deliberation and autonomy. The action-theoretic basis of political philosophy*. Aldershot: Ashgate.

Lurz, Robert W. 2009. *The philosophy of animal minds*. Cambridge: Cambridge University Press.

MacCorquodale, Kenneth, and Paul E. Meehl. 1948. On a distinction between hypothetical constructs and intervening variables. *Psychological Review* 55: 95–107.

Magill, Kevin. 1997. *Freedom and experience. Self-determination without illusions*. Basingstoke: Macmillan.

Malle, B.F., L.J. Moses, and D.A. Baldwin (eds.). 2001. *Intentions and intentionality. Foundations of social cognition*. Cambridge, MA/London: MIT Press.

Marks, J. 1986. *The ways of desire. New essays in philosophical psychology on the concept of wanting*. Chicago: Precedent Publishing.

Martin, C.B. 1994. Dispositions and conditionals. *The Philosophical Quarterly* 44: 1–8.

May, Joshua, and Richard Holton. 2012. What in the world is weakness of will? *Philosophical Studies* 157: 341–360.

McArthur, L.Z. 1981. What grabs you? The role of attention in impression formation and causal attribution. In *Social cognition: The Ontario symposium*, ed. E.T. Higgins, C.P. Herman, and M.P. Zanna, 201–246. Hillsdale: Erlbaum.

McCann, Hugh J. 1975. Trying, paralysis and volition. In McCann 1998, 94–109.

McCann, Hugh J. 1986a. Rationality and the range of intention. *Midwest Studies in Philosophy* X: 191–211.

McCann, Hugh J. 1986b. Intrinsic intentionality. In McCann 1998, 127–146.

McCann, Hugh J. 1991. Settled objects and rational constraints. In *ibid.*, 195–212.

McCann, Hugh J. 1998. *The works of agency. On human action, will and freedom*. Ithaca/London: Cornell University Press.

McClelland, David C., and John W. Atkinson. 1948. The projective expression of needs: I. The effect of different intensities of the hunger drive on perception. *Journal of Psychology* 25: 205–222.

McClelland, David C., J.W. Atkinson, R. Clark, and E.L. Lowell. 1976. *The achievement motive*. New York: Irvington.

McDowell, John. 1978. Are moral requirements hypothetical imperatives? In McDowell 1998, 77–94.

McDowell, John. 1981. Non-cognitivism and rule-following. In McDowell 1998, 198–218.

McDowell, John. 1998. *Mind, value and reality*. Cambridge, MA/London: Harvard University Press.

McNaughton, David. 1988. *Moral vision. An introduction to ethics*. Oxford: Blackwell.

Meiland, Jack W. 1970. *The nature of intention*. London: Methuen.

Meinong, Alexius. 1894. *Psychologisch-ethische Untersuchungen*. In *Abhandlungen zur Werttheorie*, *Gesamtausgabe* vol. III. Graz: Akademischer Druck- und Verlagsanstalt 1968.

Melden, A.I. 1960. Willing. *Philosophical Review* 69: 475–484.

Mele, Alfred R. 1984a. Aristotle on the proximate efficient cause of action. *Canadian Journal of Philosophy*. Supp. Vol. X: 133–155.

Mele, Alfred R. 1984b. Aristotle's wish. *Journal of the History of Philosophy* XXII: 139–156.

Mele, Alfred R. 1989. She intends to try. *Philosophical Studies* 55: 101–106.

Mele, Alfred R. 1992a. *Springs of action. Understanding intentional behavior*. New York/Oxford: Oxford University Press.

Mele, Alfred R. 1992b. Recent work on intentional action. *American Philosophical Quarterly* 29: 199–217.

Mele, Alfred R. 1995a. *Autonomous agents. From self-control to autonomy*. New York/Oxford: Oxford University Press.

Mele, Alfred R. 1995b. Motivation: Essentially motivation-constituting attitudes. *The Philosophical Review* 104: 387–423.

Mele, Alfred R. 1998. Motivational strength. *Noûs* 32: 23–36.

Mele, Alfred R. 1999. Motivation, self-control and the agglomeration of desires. *Facta Philosophica* 1: 77–86.

Mele, Alfred R. 2000. Deciding to act. *Philosophical Studies* 100: 81–108.

Mele, Alfred R. 2003a. *Motivation and agency*. New York: Oxford University Press.

Mele, Alfred R. 2003b. Intending and trying. Tuomela vs. Bratman at the video arcade. In *Realism in action*, ed. K. Miller, M. Sintonen, and P. Ylikoski, 129–135. Norwell: Kluwer Academic Pub.

Mele, Alfred R. 2007. Reasonology and false beliefs. *Philosophical Papers* 56: 91–118.

Mele, Alfred R. 2010. Weakness of will and akrasia. *Philosophical Studies* 150: 391–404.

Meltzoff, Andrew N. 1995. Understanding the intentions of others: Re-enactment of intended acts by 18-month-old children. *Developmental Psychology* 31: 838–850.

Melzack, Ronald. 1973. *The puzzle of pain*. New York: Basic Books.

Mill, John Stuart. (U). 1991. *Utilitarianism*. In *On liberty and other essays*. Oxford/New York: Oxford University Press.

Mook, Douglas G. 1996. *Motivation. The organization of action*. New York/London: Norton.

Moore, G.E. 1942. A reply to my critics. In *The philosophy of G.E. Moore*, ed. P.A. Schilpp, 533–688. La Salle: Open Court.

Moore, G.E. 1944. Russell's "theory of descriptions". In *The philosophy of Bertrand Russell*, ed. P.A. Schilpp, 177–225. La Salle: Open Court.

Moran, Richard. 2001. *Authority and estrangement. An essay on self-knowledge*. Princeton: Princeton University Press.

Moses, Louis J. 2001. Some thoughts on ascribing complex intentional concepts to young children. In Malle et al. 2001, 69–83.

Mulcahy, Nicholas, and Josep Call. 2006. Apes save tools for future use. *Science* 312: 1038–1040.

Nagel, Thomas. 1970. *The possibility of altruism*. Princeton: Princeton University Press.

Neal, David T., and Wendy Wood. 2009. Automaticity in situ and in the lab: The nature of habit in daily life. In *Oxford handbook of human action*, ed. E. Morsella, J.A. Bargh, and P.M. Gollwitzer, 442–457. Oxford: Oxford University Press.

Nielsen, Karen Margrethe. 2012. The will: Origins of the notion in Aristotle's thought. *Antiquorum philosophia* 6: 47–68.

Nietzsche, Friedrich. (FW). *Die Froehliche Wissenschaft*. In *Werke*. 1969, vol. II. Frankfurt am Main/Berlin/Vienna: Ullstein.

Nisbett, Richard E., and Timothy DeCamp Wilson. 1977. Telling more than we can know: Verbal reports on mental processes. *Psychological Review* 84: 231–259.

Norman, D.A. 1981. Categorization of action slips. *Psychological Review* 88: 1–15.

Nowell-Smith, P.H. 1957a. *Ethics*. Oxford: Blackwell.

Nowell-Smith, P.H. 1957b. Choosing, deciding and doing. *Analysis* 18: 63–69.

O'Connor, Timothy. 1995. Agent causation. In *Agents, causes & events. Essays on indeterminism and free will*, ed. T. O'Connor, 173–200. New York/Oxford: Oxford University Press.

O'Connor, Timothy. 2000. *Persons and causes. The metaphysics of free will*. New York: Oxford University Press.

Orellana-Damacella, Lucía E., et al. 2000. Decisional and behavioral procrastination. How they relate to self-discrepancies. *Journal of Social Behavior and Personality* 15: 225–238.

Orlick, Terry. 1990. *In pursuit of excellence. How to win in sport and life through mental training*. Champaign: Leisure Press.

O'Shaughnessy, Brian. 1980. *The will. A dual aspect theory*, 2 vols. Cambridge: Cambridge University Press.

O'Shaughnessy, Brian. 1997. Trying (as the mental 'pineal gland'). In *The philosophy of action*, ed. Alfred R. Mele, 53–74. Oxford: Oxford University Press.

Palmer, F.R. 1987. *The English verb*. London/New York: Longman.

Papineau, David, Cecilia Heyes. 2006. Rational or associative? Imitation in Japanese quail. In Hurley/Nudds 2006, 188–195.

Parfit, Derek. 1984. *Reasons and persons*. Oxford: Clarendon Press.

Peacocke, Christopher. 1985. Intention and Akrasia. In Vermazen, Hintikka 1985, 51–73.

Pears, David. 1985. Intention and belief. In Vermazen, Hintikka 1985, 75–88.

Pendlebury, Michael. 1986. Against the power of force: Reflections on the meaning of mood. *Mind* 95: 361–372.

Perry, Ralph Barton. 1967. *General theory of value its meaning and basic principles construed in terms of interest*. Cambridge, MA: Harvard University Press.

Perry, John. 1979. The problem of the essential indexical. *Noûs* 13: 3–21.

Peters, R.S. 1961–1962. Emotions and the category of passivity. *Proceedings of the Aristotelian Society* 62: 117–134.

Peters, R.S., and H. Tajfel. 1958. Hobbes and Hull – Metaphysicians of behaviour. *The British Journal for the Philosophy of Science* VIII: 30–44.

Pettit, Philip. 1991. Decision theory and folk psychology. In *Foundation of decision theory. Issues and advances*, ed. M. Bacharach and S. Hurley, 147–175. Oxford/Cambridge, MA: Basil Blackwell.

Pettit, Philip. 1996a. Three aspects of rational explanation. In *Rules, reasons, and norms. Selected essays*, 177–191. Oxford: Clarendon Press 2002.

Pettit, Philip. 1996b. *The common mind. An essay on psychology, society and politics*. Oxford: Oxford University Press.

Pettit, Philip, Samuel Scheffler, Smith Michael, and R. Jay Wallace. 2006. *Reason and value. Themes from moral philosophy of Joseph Raz*. Oxford: Clarendon Press.

Piller, Christian. 2001. Normative practical reasoning. *Proceedings of the Aristotelian Society, Supplementary Volume* 75: 195–216.

Pink, Thomas. 1996. *The psychology of freedom*. Cambridge: Cambridge University Press.

Plato. (Rep). 1970. *The Republic*. Trans. F.M. Cornford. Oxford: Oxford University Press.

Plato. (Phil). 1975. *Philebus*. Trans. J.C. Gosling. Oxford: Clarendon Press.

Plato. (Prot). 1995. *Protagoras*. Trans. C.C.W. Taylor. Oxford: Clarendon Press.

Platts, Mark de. Bretton. 1979. *Ways of meaning. An introduction to the philosophy of language*. London/Henley/Boston: Routledge & Kegan Paul.

Postman, Leo, and Geoffrey Leytham. 1950–1951. Perceptual selectivity and ambivalence of stimuli. *Journal of Personality* 19: 390–405.

Postman, Leo, Jerome S. Bruner, and Elliot McGinnies. 1948. Personal values as selective factors in perception. *Journal of Abnormal and Social Psychology* 43: 142–154.

Prinz, Jesse. 2004. *Gut reactions. A perceptual theory of emotions*. Oxford/New York: Oxford University Press.

Prinz, Wolfgang. 1987. Ideo-motor action. In *Perspectives on perception and action*, ed. H. Heuer and A.F. Sanders, 47–76. Hillsdale: Erlbaum.

Prinz, Wolfgang. 1990. A common coding approach to perception and action. In *Relationships between perception and action. Current approaches*, ed. O. Neumann and W. Prinz, 167–201. Berlin: Springer.

Quinn, Warren. 1993. Putting rationality in its place. In *Morality and action*, 228–255. Cambridge: Cambridge University Press.

Quirk, Randolph, S. Greenbaum, G. Leech, and J. Svartvik. 1985. *A comprehensive grammar of the English language*. London/New York: Longman.

Raby, C.R., D.M. Alexis, A. Dickinson, and N.S. Clayton. 2007. Planning for the future by western scrub jays. *Nature* 445: 919–921.

Rakoczy, Hannes. 2010. Executive function and the development of belief-desire psychology. *Developmental Science* 13: 648–661.

Rakoczy, Hannes, and Michael Tomasello. 2007. The ontogeny of social ontology: Steps to shared intentionality and status functions. In *Intentional acts and institutional facts*, ed. S.L. Tsohatzidis, 113–137. Dordrecht: Springer.

Rakoczy, Hannes, Felix Warneken, and Michael Tomasello. 2008. The sources of normativity: Young children's awareness of the normative structure of games. *Developmental Psychology* 2008: 875–881.

Raz, Joseph. 1978. Reasons for actions, decisions and norms. In *Practical reasoning*, ed. J. Raz, 128–143. Oxford: Oxford University Press.

Raz, Joseph. 1999. Explaining normativity: On rationality and the justification of reason. *Ratio* 12: 354–379.

Reason, J. 1979. Actions not as planned: The price of automatization. In *Aspects of conciousness*, vol. 1, ed. G. Underwood and R. Stevens, 67–89. London: Academic Press.

Reeve, C.D.C. 2012. *Action, contemplation and happiness: An essay on Aristotle*. Cambridge, MA: Harvard University Press.

Repacholi, Betty M., and Alison Gopnik. 1997. Early reasoning about desires: Evidence from 14- and 18-month-olds. *Developmental Psychology* 33: 12–21.

Ridge, Michael. 1998. Humean intentions. *American Philosophical Quarterly* 35: 157–177.

Roberts, Robert C. 2009. The sophistication of non-human emotion. In Lurz 2009, 218–236.

Rosenthal, David M. 1989. Intentionality. In *Rerepresentation. Readings in the philosophy of mental representation*, ed. S. Silvers, 311–339. Dordrecht/Boston/London: Kluwer.

Rosenthal, David M. 1993. Thinking that one thinks. In *Consciousness. Psychological and philosophical essays*, ed. M. Davis and G.W. Humphreys, 197–222. Oxford/Cambridge, MA: Blackwell Pub.

Roughley, Neil. 1999. Mögen und Wünschen. Zur volitiven Theorie des Hedonischen. In *Die Zukunft des Wissens. XVIII. Deutscher Kongress fuer Philosophie*, ed. J. Mittelstrass, 336–343. Konstanz: Universitätsverlag Konstanz.

Roughley, Neil. 2001. Review of Michael Bratman, "faces of intention". *International Journal of Philosophical Studies* 9: 265–270.

Roughley, Neil. 2007a. Hilberts Krawatte, Ryles Clown und Gehlens Schlüssel. Zur Analyse von Gewohnheitshandlungen. *Zeitschrift für Philosophische Forschung* 61: 188–206.

Roughley, Neil. 2007b. On the ways and usings of intending. Lessons from Velleman's Bratman critique. In Leist 2007, 216–230.

Roughley, Neil. 2007c. The double failure of double effect. In Lumer, Nannini 2007, 91–116.

Roughley, Neil. 2008a. Willensschwäche und Personsein. In *Personalität*, ed. H. Tegtmeyer and F. Kannetzky, 144–161. Leipzig: Universitätsverlag.

Roughley, Neil. 2008b. Das irrationale Tier. In *Der Ort der Vernunft in einer natürlichen Welt. Logische und anthropologische Ortsbestimmungen*, ed. W.-J. Cramm and G. Keil, 216–233. Weilerswist: Velbrück.

Roughley, Neil. 2010. Intrinsisch/extrinsisch. In *Enzykopaedie Philosophie und Wissenschaftstheorie*, vol. 4, ed. J. Mittelstraß, 52–58. Stuttgart/Weimar: Metzler.

Roughley, Neil. unpublished a. Representation, hedonic experience and desire.

Roughley, Neil. unpublished b. Pains for representationalists.

Roughley, Neil. unpublished c. Intentional action: An eliminable concept? Talk given at the conference "Blame in Action", Potsdam University 2010.

Sabini, John, and Maury Silver. 1982. *Moralities of everyday life*. Oxford: Oxford University Press.

Sachs, Oliver. 1995. *An anthropologist on Mars*. London: Macmillan.

Saidel, Eric. 2009. Attributing mental representations to animals. In Lurz 2009, 35–51.

Sartre, Jean-Paul. 1943. *Being and nothingness. An essay in phenomenological ontology*. New York: Philosophical Library 1956.

Sartre, Jean-Paul. 1946. *Existentialism and humanism*, 1948. London: Methuen & Co.

Sauvé Meyer, Susan. 1998. Moral responsibility: Aristotle and after. In *Companions to ancient thought 4: Ethics*, ed. Stephen Everson, 221–240. Cambridge: Cambridge University Press.

Savage-Rumbaugh, Sue, and Roger Lewin. 1994. *Kanzi. The ape at the brink of the human mind*. New York: John Wiley and Sons.

Scanlon, T.M. 1998. *What we owe to each other*. Cambridge, MA/London: Harvard University Press.

Scanlon, T.M. 2004. Reasons: A puzzling duality? In Pettit, Scheffler, Smith, Wallace 2006, 231–246.

Scanlon, T.M. 2007. Structural irrationality. In *Common minds. Themes from the philosophy of Philip Pettit*, ed. Geoffrey Brennan, Robert Goodin, and Michael Smith, 84–103. Oxford: Clarendon Press.

Schachter, Stanley, and Jerome E. Singer. 1962. Cognitive, social and psychological determinants of emotional state. *Psychological Review* 69: 379–399.

Schechtman, Marya. 2004. Self-expression and self-control. *Ratio* (new series) XVII: 409–427.

Schiffer, Stephen. 1976. A paradox of desire. *American Philosophical Quarterly* 13: 195–203.

Schlick, Moritz. 1930. *Fragen der Ethik*. Frankfurt am Main: Suhrkamp 1984.

Schmidt, Marco, Hannes Rakoczy, and Michael Tomasello. 2011. Young children attribute normativity to novel actions without pedagogy or normative language. *Developmental Science* 14: 530–539.

Schouwenberg, Henri C. 1995. Academic procrastination. Theoretical notions, measurement and research. In *Procrastination and task avoidance. Theory, research and treatment*, ed. Joseph R. Ferrari et al., 71–96. New York/London: Plenum Press.

Schroeder, Mark. 2004. The scope of instrumental reason. *Philosophical Perspectives* 18: 337–364.

Schroeder, Mark. 2007. *Slaves of the passions*. Oxford: Oxford University Press.

Schroeder, Mark. 2009. Means-end coherence, stringency and subjective reasons. *Philosophical Studies* 143: 223–248.

Schroeder, Severin. 2001. The concept of trying. *Philosophical Investigations* 24: 213–227.

Schroeder, Timothy. 2001. Pleasure, displeasure and representation. *Canadian Journal of Philosophy* 31: 507–530.

Schroeder, Timothy. 2004. *Three faces of desire*. Oxford/New York: Oxford University Press.

Schueler, G.F. 1995. *Desire. Its role in practical reasoning and the explanation of action*. Cambridge, MA/London: MIT Press.

Schueler, G.F. 2003. *Reasons and purposes. Human rationality and teleological explanation of action*. Oxford: Oxford University Press.

Schult, Carolyn A. 2002. Children's understanding of the distinction between intentions and desires. *Child Development* 73: 1727–1747.

Scruton, Roger. 1974. *Art and imagination. A study in the philosophy of mind*. South Bend: St. Augustine's Press 1998.

Searle, John. 1969. *Speech acts. An essay in the philosophy of language*. Cambridge: Cambridge University Press.

Searle, John. 1979. *Expression and meaning. Studies in the theory of speech acts*. Cambridge: Cambridge University Press.

Searle, John. 1983. *Intentionality. An essay in the philosophy of mind*. Cambridge: Cambridge University Press.

Searle, John. 1990. Consciousness, explanatory inversion and cognitive science. *Behavioral and Brain Sciences* 13: 585–596.

Searle, John. 1992. *The rediscovery of mind*. Cambridge, MA/London: MIT Press.

Searle, John. 1994a. The connection principle and the ontology of the unconscious. A reply to Fodor and Lepore. *Philosophy and Phenomenological Research* LIV: 847–855.

Searle, John. 1994b. Animal minds. *Midwest Studies in Philosophy* XIX: 206–219.

Searle, John. 2001. *Rationality in action*. Cambridge, MA/London: MIT Press.

Seebass, Gottfried. 1981–1982. Mediation theory and the problem of psychological discourse on 'inner' events, parts I & II. *Ratio* vol. 23: 81–97; vol. 24, 29–43.

Seebass, Gottfried. 1993. *Wollen*. Klostermann: Frankfurt am Main.

Seebass, Gottfried. 2006. *Handlung und Freiheit*. Tübingen: Mohr Siebeck.

Sellars, Wilfrid. 1956. Empiricism and the philosophy of mind. In *Science, perception and reality.* 1964, 127–196. London: Routledge & Kegan Paul.

Sellars, Wilfrid. 1966. Thought and action. In Lehrer 1966, 105–139.

Setiya, Kieran. 2004. Hume on practical reason. *Philosophical Perspectives* 18: 365–389.

Setiya, Kieran. 2007. Cognitivism about instrumental reason. *Ethics* 117: 649–673.

Shore, Bradd. 2000. Human diversity and human nature. The life and times of a false dichotomy. In *Being humans. Anthropological universality and particularity in transdisciplinary perspectives*, ed. N. Roughley, 81–103. Berlin/New York: de Gruyter.

Sidgwick, Henry. (ME). 1981. *The Methods of Ethics*. Indianapolis: Hackett.

Skinner, B.F. 1953. *Science and human behavior*. New York: MacMillan.

Smith, Michael. 1987. The humean theory of motivation. *Mind* XCVI: 36–61.

Smith, Michael. 1994. *The moral problem*. Oxford/Malden: Blackwell.

Smith, Michael. 2010. The standard story of action: An exchange (1). In *Causing human action*, ed. J.H. Aguilar and A.A. Buckareff, 45–56. Cambridge, MA: MIT Press.

Sobel, Jordan Howard. 1994. Useful intuitions. In *Taking chances. Essays on rational choice*, 237–254. Cambridge: Cambridge University Press.

Sorell, Tom. 1986. *Hobbes*. London: Routledge & Kegan Paul.

Southwood, Nicholas. 2008. Vindicating the norms of rationality. *Ethics* 119: 9–30.

Stagner, Ross. 1977. Homeostasis, discrepancy, dissonance. A theory of motives and motivation. *Motivation and Emotion* 1: 103–138.

Stampe, Dennis. 1987. The authority of desire. *Philosophical Review* XCVI: 335–381.

Stevenson, Charles L. 1944. *Ethics and language*. New Haven/London: Yale University Press 1965.

Stevenson, Charles L. 1948. The nature of ethical disagreement. In *Facts and values. Studies in ethical analysis*. 1964, 1–9. New Haven/London: Yale University Press.

Stich, Stephen P. 1979. Do animals have beliefs? *Australasian Journal of Philosophy* 57: 15–28.

Stocker, Michael. 1979. Desiring the bad: An essay in moral psychology. *Journal of Philosophy* LXXVI: 738–753.

Stoutland, Frederick. 2007. Reasons for action and psychological states. In Leist 2007, 75–94.

Strawson, Galen. 1994. *Mental reality*. Cambridge, MA/London: MIT Press.

Taylor, Richard. 1964. Deliberation and foreknowledge. *American Philosophical Quarterly* 1: 73–80.

Taylor, Richard. 1966. *Action and purpose*. Englewood Cliffs: Prentice-Hall.

Taylor, S.E., and S.T. Fiske. 1978. Salience, attention and attribution: Top of the head phenomena. In *Advances in experimental and social psychology*, vol. 11, ed. L. Berkowitz, 249–288. New York: Academic Press.

Taylor, Shelley, and Peter M. Gollwitzer. 1995. Effects of mindset on positive illusions. *Journal of Personality and Social Psychology* 69: 213–226.

Tenenbaum, Sergio. 2003. Accidie, evaluation and motivation. In *Weakness of will and practical irrationality*, ed. S. Stroud and C. Tappolet, 147–171. Oxford/New York: Oxford University Press.

Tenenbaum, Sergio (ed.). 2010a. *Desire, practical reason and the good*. Oxford/New York: Oxford University Press.

Tenenbaum, Sergio. 2010b. Introduction. In Tenenbaum 2010a, 3–5.

Thalberg, Irving. 1962. Intending the impossible. *Australasian Journal of Philosophy* XL: 49–56.

Thalberg, Irving. 1972. How can we distinguish between doing and undergoing? In *Enigmas of agency. Studies in the philosophy of human action*, 48–72. London: Allen & Unwin.

Tinbergen, Nikolaas. 1989. *The study of instinct*. Oxford: Clarendon Press.

Toates, Frederick. 1986. *Motivational systems*. Cambridge: Cambridge University Press.

Tolman, E.C. 1938. The determiners of behavior at a choice point. *Psychological Review* 45: 1–41.

Tomasello, Michael, and Josep Call. 2006. Do chimpanzees know what others see – or only what they are looking at? In Hurley/Nudds 2006, 371–384.

Tomasello, M., M. Carpenter, J. Call, T. Behne, and H. Moll. 2005. Understanding and sharing intentions. The origins of cultural cognition. *Behavioral and Brain Sciences* 28: 675–735.

Tomberlin, J.E. 1983. *Agent, language and structure of the world. Essays presented to Hector-Neri Castañeda with his replies*. Indianapolis: Hackett.

Tugendhat, Ernst. 1976. *Vorlesungen zur Einführung in die sprachanalytische Philosophie*. Suhrkamp: Frankfurt am Main.

Tuozzo, Thomas M. 1994. Conceptualized and unconceptualized desire in Aristotle. *Journal of the History of Philosophy* 32(4): 525–549.

Tye, Michael. 1985. *Ten problems of consciousness. A representational theory of the phenomenal mind*. Cambridge, MA/London: MIT Press.

Ullmann-Margalit, Edna, and Sidney Morgenbesser. 1977. Picking and choosing. *Social Research* 44: 757–785.

Van Hees, M., and O. Roy. 2009. Intentions, decisions and rationality. In *Economics, rational choice and normative philosophy*, ed. T. Boylan and R. Gekker, 56–72. London/New York: Routledge.

Van Inwagen, Peter. 1989. When is the will free? *Philosophical Perspectives* 3: 399–422.

Velleman, J. David. 1989. *Practical reflection*. Princeton: Princeton University Press.

Velleman, J. David. 1992a. The guise of the good. In Velleman 2000a, 99–122.

Velleman, J. David. 1992b. What happens when someone acts? In Velleman 2000a, 123–143.

Velleman, J. David. 1996. The possibility of practical reason. In Velleman 2000a, 170–199.

Velleman, J. David. 2000a. *The possibility of practical reason*. Oxford: Clarendon Press.

Velleman, J. David. 2000b. Introduction. In Velleman 2000a, 1–31.

Velleman, J. David. 2007. What good is a will? In Leist 2007, 193–215.

Vendler, Zeno. 1972. *Res cogitans. An essay in rational psychology*. Ithaca/ London: Cornell University Press.

Vermazen, Bruce, and Meril B. Hintikka. 1985. *Essays on Davidson. Actions and events*. Oxford: Clarendon Press.

Vogt, Katja. unpublished. Desiring the good: A socratic reading of Aristotle. http://katjavogt.com/pdf/katja_vogt_motivation.pdf. Last accessed Aug 2014.

Vollmer, Fred. 1993. Intentional action and unconscious reasons. *Journal for the Theory of Social Behaviour* 23: 315–326.

von Wright, G.H. 1971. *Explanation and understanding*. London: Routledge.

von Wright, G.H. 1972. On so-called practical inference. *Acta Sociologica* 15: 39–53.

Wagner, Hugh. 1999. *The psychobiology of human motivation*. London/New York: Routledge.

Wallace, R. Jay. 1999. Addiction as a defect of the will. In Wallace 2006b, 165–189.

Wallace, R. Jay. 2001. Normativity, commitment and instrumental reason. In Wallace 2006b, 82–111.

Wallace, R. Jay. 2006a. Postscript to Chapter 5. In Wallace 2006b, 111–120.

Wallace, R. Jay. 2006b. *Normativity and the will. Selected essays on moral psychology and practical reason*. Oxford: Oxford University Press.

Warneken, Felix, and Michael Tomasello. 2006. Altruistic helping in human infants and young chimpanzees. *Science* 311: 1301–1303.

Watson, Gary. 1996. Two faces of responsibility. In Watson 2004, 260–288.

Watson, Gary. 1999. Disordered appetites: Addiction, compulsion and dependence. In Watson 2004, 59–87.

Watson, Gary. 2003. The work of the will. In Watson 2004, 123–157.

Watson, Gary. 2004. *Agency and answerability. Selected essays*. Oxford: Clarendon Press.

Watson, John B. 1913. Psychology as the behaviorist views it. *The Psychological Review* XX: 158–177.

Watson, John B. 1924. *Behaviorism*. Chicago/London: University of Chicago Press 1966.

Way, Jonathan. 2011. The symmetry of rational requirements. *Philosophical Studies* 155: 227–239.

Wedgwood, Ralph. 2007. *The nature of normativity*. Oxford: Clarendon Press.

Wegner, Daniel M. 1994. Ironic processes of mental control. *Psychological Review* 101: 34–52.

Weir, Alex A.S., Jackie Chappell, and Alex Kacelnik. 2002. Shaping of hooks in New Caledonian crows. *Science* 297: 981.

Weir, Alex A.S., and Alex Kacelnik. 2006. A New Caledonian crow (Corvus moneduloides) creatively re-designs tools by bending or unbending aluminium strips. *Animal Cognition* 9: 317–334.

Wellman, Henry M. 1992. *The child's theory of the mind*. Cambridge, MA/London: MIT Press.

Wellman, Henry, M. David Cross, and Julanne Watson. 2001. Meta-analysis of theory-of-mind development: The truth about false belief. *Child Development* 72: 655–684.

Whiten, Andrew, and Richard W. Byrne. 1988. The manipulation of attention in primate tactical deception. In *Machiavellian intelligence. Social expertise and the evolution of intellect in monkeys, apes and humans*, ed. R.W. Byrne and A. Whiten, 211–223. Oxford: Clarendon Press.

Williams, Bernard. 1965. Ethical consistency. In Williams 1973, 166–186.

Williams, Bernard. 1970. Deciding to believe. In *ibid.*, 136–151.

Williams, Bernard. 1973. *Problems of the self. Philosophical papers 1956–1972*. Cambridge: Cambridge University Press.

Williams, Bernard. 1980. Internal and external reasons. In *Moral luck. Philosophical papers, 1973–1980*. 1981, 101–113. Cambridge: Cambridge University Press.

Williams, Bernard. 1985. *Ethics and the limits of philosophy*. Glasgow: Fontana.

Wittgenstein, Ludwig (Z). 1981. *Zettel*. Berkeley: University of California Press.

Woodfield, Andrew. 1981–1982. Desire, intentional content and teleological explanation. *Proceedings of the Aristotelian Society* LXXXII: 69–87.

Woodfield, Andrew. 1982. On specifying the contents of thoughts. In *Thought and object. Essays on intentionality*, ed. A. Woodfield, 258–297. Oxford: Clarendon.

Wright, Larry. 1976. *Teleological explanation*. Berkeley/Los Angeles: University of California Press.

Young, Paul Thomas. 1955. The role of hedonic processes in motivation. In *The Nebraska symposium on motivation*, ed. M.R. Jones, 193–238. Lincoln: University of Nebraska Press.

Young, Paul Thomas. 1961. *Motivation and emotion. A survey of the determinants of human and animal activity*. New York/London/Sydney: Wiley.

Zangwill, Nick. 1998. Direction of fit and normative functionalism. *Philosophical Studies* 91: 173–203.

Author Index

N. Roughley, *Wanting and Intending*, Philosophical Studies Series 123,
DOI 10.1007/978-94-017-7387-4

Subject Index

N. Roughley, *Wanting and Intending*, Philosophical Studies Series 123,
DOI 10.1007/978-94-017-7387-4