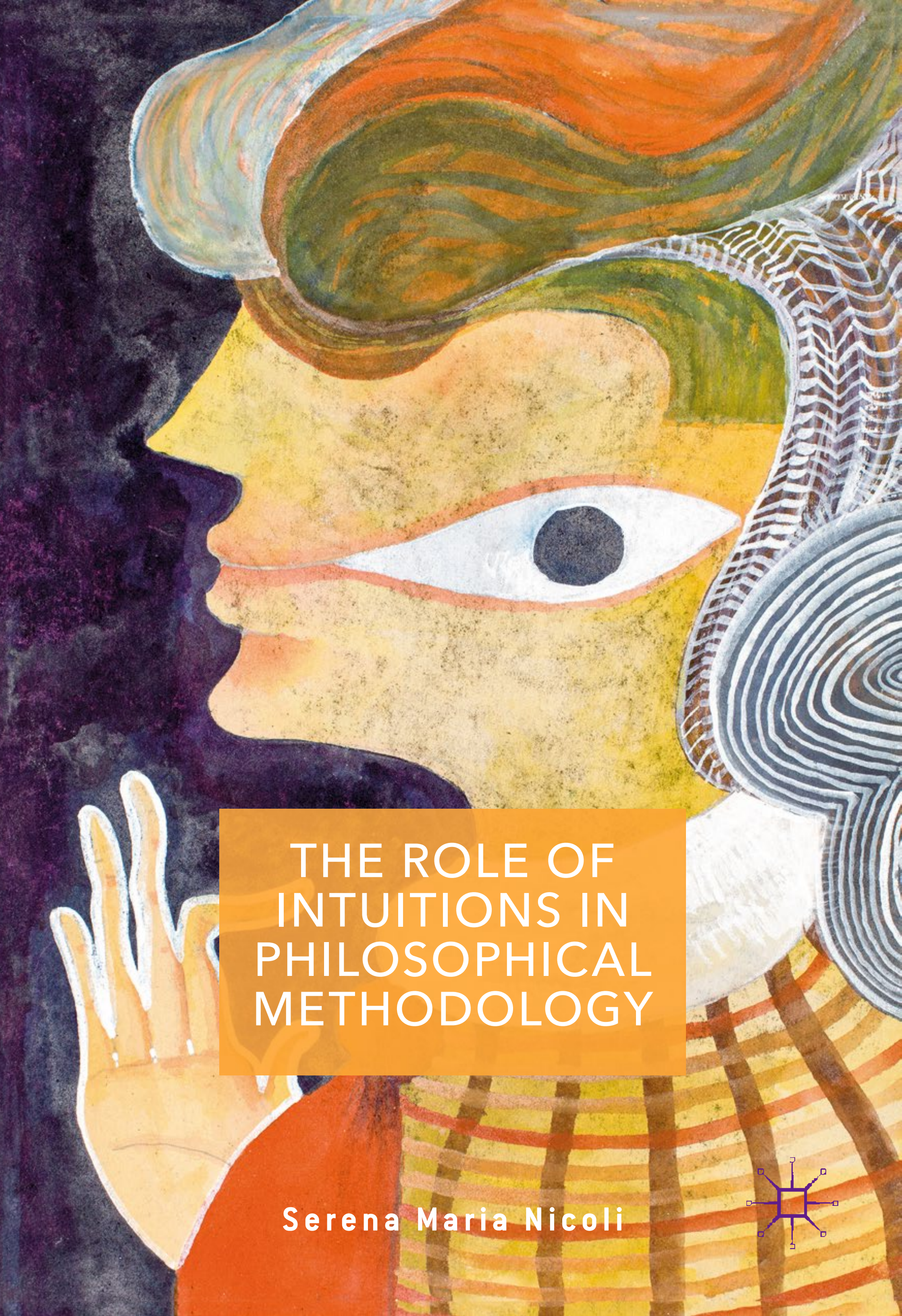

THE ROLE OF INTUITIONS IN PHILOSOPHICAL METHODOLOGY

Serena Maria Nicoli

The Role of Intuitions in Philosophical Methodology

Serena Maria Nicoli

The Role of Intuitions in Philosophical Methodology

palgrave
macmillan

Serena Maria Nicoli
Torino, Italy

ISBN 978-1-137-56714-7 ISBN 978-1-137-56715-4 (eBook)
DOI 10.1057/978-1-137-56715-4

Library of Congress Control Number: 2016958143

© The Editor(s) (if applicable) and The Author(s) 2016
The author(s) has/have asserted their right(s) to be identified as the author(s) of this work in accordance with the Copyright, Designs and Patents Act 1988.
This work is subject to copyright. All rights are solely and exclusively licensed by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.
The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.
The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made.

Cover Image: © Paul-Daniel Florea, iStock / Getty Images Plus
Cover Design: Fatima Jamadar

Printed on acid-free paper

This Palgrave Macmillan imprint is published by Springer Nature
The registered company is Macmillan Publishers Ltd. London
The registered company address is: The Campus, 4 Crinan Street, London, N1 9XW, United Kingdom

To Matteo

Acknowledgements

I would like to express my gratitude to Professor Diego Marconi for his insight and key input to this study. I am also indebted to Professor Alfredo Paternoster for his advice. Last but not least, I would like to thank all the friends who have contributed in their own personal way to my work: Laura Celani, Fabrizio Calzavarini, Jennifer Cooke, and Giacomo Melis.

Contents

Part I

The Method of Philosophy

1

Introduction to Part I

1.1 What Is X? Intuitions: A Basis for Theory Construction and Justification

One of the main worries of philosophers from Plato's dialogues to today's inquiries is investigating the so-called 'Socratic questions', in other words, questions of the form 'What is X?'. Answering questions like 'What is knowledge?', 'What is justice?', 'What is reference?' is arguably the main motive for most philosophers' work. Hence, it is important to clarify the methodology of investigation of Socratic questions.

Let us consider one of Plato's most famous dialogues, *The Republic.* In Book I, Plato asks what justice is. The question emerges through a dialogue between Socrates and Cephalus. First, Socrates asks Cephalus what being old and rich is like. Cephalus answers

> The great blessing of riches, I do not say to every man, but to a good man, is, that he has had no occasion to deceive or to defraud others, either intentionally or unintentionally; and when he departs to the world below he is not in any apprehension about offerings due to the gods or debts which he

© The Author(s) 2016

S.M. Nicoli, *The Role of Intuitions in Philosophical Methodology*,
DOI 10.1057/978-1-137-56715-4_1

> owes to men. Now to this peace of mind the possession of wealth greatly contributes; and therefore I say, that, setting one thing against another, of the many advantages which wealth has to give, to a man of sense this is in my opinion the greatest. (Plato 1901, pp. 330e–331a)

Cephalus's answer prompts new questions and, in fact, Socrates replies:

> But as concerning justice, what is it?—to speak the truth and to pay your debts—no more than this? And even to this are there not exceptions? Suppose that a friend when in his right mind has deposited arms with me and he asks for them when he is not in his right mind, ought I to give them back to him? No one would say that I ought or that I should be right in doing so, any more than they would say that I ought always to speak the truth to one who is in his condition.
>
> You are quite right, he replied.
>
> But then, I said, speaking the truth and paying your debts is not a correct definition of justice.
>
> Quite correct, Socrates. (Plato 1901, p. 331b–331d)

What is Socrates' teaching? He teaches Cephalus that justice doesn't amount to telling the truth and paying one's own debts. How does he teach that? By presenting a case in which we would say that telling the truth and paying one's own debts is unjust.

The strategy adopted by the philosopher is part of a broader practice called 'the method of cases': a philosopher investigating a problem X (justice, knowledge, and so on) considers a series of situations in which we would say that a certain action is just or unjust, a certain belief is or isn't knowledge, and tries to find a theory on X accounting for what we would say in such and such circumstances.

The case discussed above is appropriate to reject a specific account of justice: justice doesn't amount to telling the truth and paying one's own debts, since, in the case of the crazed friend, we would say that telling the truth and paying one's own debts is unjust. The case in question is an imaginary counterexample to a thesis or, as participants in recent debate on the philosophical method generally call it, a 'thought experiment' (TE).

In the last century, Rawls, a political philosopher who was interested, like Plato, in the problem of justice, provided a detailed description of the method of cases. He called 'intuitions' the judgements expressing what we would say in such and such circumstances, and 'reflective equilibrium' (RE) the method that uses such judgements to build and justify theories proposed as answers to Socratic problems.

According to Rawls (1951), intuitions are the result of an inquiry into the facts concerning the cases being judged and reflections on the possible effects of different decisions. More precisely, a (moral) judgement is intuitive as it is not explicitly theoretical, that is, not determined by a systematic and conscious application of (ethical) principles. In general and non-strictly moral terms, his idea can be reconstructed as follows: an intuition on X (justice, knowledge, reference) expresses what we would say about X in real or counterfactual circumstances without appealing to any specific theory about X (justice, knowledge, reference).

What about the method of philosophy—the method that Rawls (1971) dubs 'reflective equilibrium' (RE)?

Suppose we are wondering what X (justice, knowledge, and so on) is. Presumptively, we would begin by considering a bunch of situations in which we say that a certain action is just or unjust, a certain belief is or is not knowledge. Hence, based on the initial set of intuitive verdicts (let's call this set I_1), we would try to elaborate a response to our question, that is, develop a theory T_1 by an inference to the best explanation (IBE) from the set I_1.

Let's suppose that we have managed to produce a theory, T_1, that is a consistent framework accounting for I_1. T_1 will imply a set of consequences. Inside this set there will be consequences agreeing with I_1. However, there will plausibly be other consequences unexpectedly contradicting some other intuitions of ours (let's call the set of these consequences, I_2). Namely, we may find that there are cases—real or hypothetical—in which we judge the opposite of what the theory predicts (case A). T_1 is satisfactory insofar as it presupposes (and explains) I_1, dubious insofar as it entails I_2. If (case A′) the consequences I_2 are felt to be irreparably counterintuitive, then T_1 has to be abandoned and a new theory has to be built. The constraint for the new theory, T_2, is to presuppose I_1, as T_1 did, and not to entail I_2.

Let's suppose that T_2 actually explains I_1, and that its consequences do not contradict the intuitions T_1 contradicted. Still, T_2 might also happen to bear some unwanted consequences (I_3). Again, one should go back to the theory and try to adjust it in the light of this new evidence. The search for a new theory that doesn't entail intuitions we aren't willing to accept may continue indefinitely; however, let's assume that one manages to find a theory T_2 that actually explains I_1, that does not disagree with the intuitions T_1 contradicted and doesn't bear further unwanted consequences (I_3). Then the result is mutual support between theory and intuitions, that is, achievement of the so-called reflective equilibrium (RE).

It is worth pointing out that, according to Rawls (1971) and to the other theorists of this method (Goodman 1955; Lewis 1973, 1980a; Daniels 1979b, 1980a), the opposite to what happens in case A′ can also occur: the theory can prevail on our intuitions. Namely, a theory could be strong enough to prompt us to a *Gestalt switch* by persuading us to judge the relevant case according to the theory, and against the originally problematic intuition. If so, we decide to preserve the theory and revise the set of intuitions: we abandon our original intuitions and embrace the consequences of the theory (I_3) as new intuitions (case A″).

In essence, in the face of problematic scenarios the theorist can either modify the theory on the basis of intuitions, or adjust the set of intuitions on the basis of the theory. Either way, the goal is the same: reconcile our theoretical convictions with our intuitive judgements. That is to say, we seek an agreement between a theory's predictions (what the theory says is correct to judge in such and such circumstances) and what we deem correct to say in the same circumstances regardless of the theory.

In Chap. 3, I will say more on reflective equilibrium. At this point I will use a (non-moral) example to illustrate the reflective dynamics in which intuitions are involved. Let us imagine that we are interested in understanding what knowledge is, and let us consider the following scenarios. A subject, *S*, states it is two o'clock. Does *S* know it is two o'clock? Intuitively, if *S* knows it is two o'clock, then (1) it is two o'clock, and (2) *S* believes it is two o'clock. Let us take (1): if it were not two o'clock, would we say that *S* knows it is two o'clock? Obviously, we would not: if it were not two o'clock, *S* would not know that it is two o'clock, rather *S* would merely presume to know it is two o'clock. Let us then take (2): *S* says 'it is two o'clock, but I do not believe it is two o'clock': would we say that *S* knows it is two o'clock? Not really: even if it were two o'clock, we would never say that *S* does not believe it is two o'clock and nevertheless knows it. To be considered knowledge, a proposition has to be believed. Hence, have we defined what knowledge is? Let us imagine S says it is two o'clock, believes it is two o'clock, but is merely taking a wild guess. Would we say that *S* knows it is two o'clock? Obviously not: for *S* to know it is two o'clock, *S* should have an adequate justification. So, let us suppose that *S* looks at a watch and the watch shows two o'clock. Would we say that *S* knows it is two o'clock? In this case, we would. Hence, based on these considerations we may conclude that knowledge is justified true belief. However, let us imagine this rare—but concretely possible—scenario: *S* looks at the watch at two o'clock, the watch shows two o'clock but is broken. Fortuitously the minute and hour hands are oriented so as to show two o'clock. Does *S* know it is two o'clock? No, we would not say that *S* knows.

This last instance—the case of the broken clock showing fortuitously the right time—is a version of the sort of cases introduced by Gettier in 'Is justified true belief knowledge?' (1963). In this famous paper Gettier refuted the theory of knowledge that was at the time accepted by the community of epistemologists: the theory according to which knowledge is justified true belief (JTB theory of knowledge). How did he manage to do this? By showing that there are cases in which a certain belief is justified true belief but not knowledge. Let us take the case of the broken watch: if the JTB theory of knowledge were true, it would follow that *S* would know that it is two o' clock, but, clearly, we would not say that *S* does.

In conclusion, intuitive verdicts seem to play a crucial role in the investigation of Socratic problems. They precede and govern the elaboration of theories and offer criteria for their acceptance: we can claim a theory is justified by showing it matches with our intuitions about cases; or we can attack a theory by showing that it does not account for our intuitions in a series of cases. According to a widely shared thesis, intuitions are the evidence philosophers appeal to.

1.2 What Are Intuitions?

Yet, what are intuitions exactly? Can they be legitimately used to support or attack philosophical theories? Over the last decades, a renewed interest in metaphilosophical issues has prompted many philosophers in the analytic tradition to investigate the nature of intuitions. Generally, when explaining what intuitions are, participants in the debate introduce cases of the sort considered in Section 1.1. Philosophical intuitions are presented through TEs because TEs are the 'loci' where the appeal to intuitions is usually acknowledged and intuitive judgements are easy to identify. Aside from judgements expressing what we would say in the hypothetical circumstances described in TEs, also verdicts like 'Torturing an innocent for fun is wrong' or 'It is impossible for a square to have five sides' are presented as intuitions (Pust 2012). Yet, can we give a definition of what a philosophical intuition is?

In the last few years, a wide range of answers has been provided. In his entry to the *Stanford Encyclopaedia of Philosophy*, 'Intuition', Pust notices that psychologists and philosophers with naturalistic inclinations (Gopnik and Schwitzgebel 1998; Kornblith 1998; Devitt 2006) tend to attribute a special aetiology to intuitive judgements: an intuition is a belief that is not consciously inferred from some other belief. By contrast, tradition-oriented philosophers opt for more parsimonious answers. For example, Lewis claims that 'intuitions are simply opinions' (Lewis 1983a, p. x). Similarly, for Van Inwagen 'our "intuitions" are simply our beliefs—or perhaps, in some cases, the tendencies that make certain beliefs attractive to us, that "move" us in the direction of accepting certain propositions without taking us all the way to acceptance' (Van Inwagen 1997, p. 309).

One problem for parsimonious characterizations is that they do not account for the fact that 'one can have an intuition that *p* while one does not believe that *p* and one can have a belief that *p* without having an intuition that *p*' (Pust 2012, p. 3). To illustrate this problem Pust uses the analogy with perceptive illusions. He mentions Müller–Lyer arrows, an optical illusion in which two lines (arrows) of identical length are perceived to be different in length because of the different orientation of the fins (on one arrow the fins are oriented inwards and on the other one they are oriented outwards): even though we have measured the two lines and we know that they are equal, we still see one line longer than the other. The same 'resistance' seems to characterize (some) philosophical intuitions: although the theory we embrace seems to tell us better, we keep assessing the cases the way we did (that is, according to our intuition and contrarily to the theory). Secondly, there are beliefs that obviously do not count as intuitions: one can believe that *p* (that the cat is on the table, that one is thirsty) without having the intuition that *p* (that the cat is on the table, that one is thirsty).

More restrictive solutions were proposed. Sosa (1998), for instance, holds that intuitions are beliefs but imposes a further constraint: a judgement *p* counts as intuitive if the subject having an intuition that *p* is disposed to believe *p* merely on the basis of understanding that *p*. Ludwig (2007) restricts intuitions to beliefs solely based on competence. Other advocates of the role of intuitions claim that intuitive judgements have a specific phenomenology: 'that peculiar form of phenomenology with which we are all well acquainted, but which I can't describe in any way other than as the phenomenology that goes with seeing that such proposition is true' (Plantinga 1993, pp. 105–106). Finally, there are philosophers who do not characterize intuitions in terms of beliefs, but describe them as *sui generis* propositional attitudes. According to Bealer (1998), for instance, intuitions are intellectual seemings.

It has also been noticed (Lycan 1988) that saying that intuitions are treated as evidence is somehow ambiguous: is the fact (psychological state) of having an intuition that *p* (the *intuiting*), or is it the content of the intuition (*p*, the *intuited*) to be treated as evidence? And if intuitions are psychological states or events, is there a faculty of intuition 'producing' them?

Last, there are few participants in the debate on philosophical method (Hintikka 1999; Williamson 2007; Cappelen 2012) who believe that the term 'intuition' is misleading and we would be better not using it. Hintikka (1999) claims philosophers make a naïve use of the term in the attempt to attribute to their work the scientific status of linguists' work. According to Williamson (2007), the use of the term 'intuition' amounts to an improper psychologization of the evidence philosophers appeal to. Cappelen (2012) purports that the term 'intuition' simply amounts to a verbal tick or virus. This use became popular within the philosophical community about thirty years ago, and did not really affect the results of first-order philosophy, but caused metaphilosophers many pseudo-problems: they started believing that 'special' judgements called 'intuitions' played a central role in the philosophical practice, while—Cappelen argues—in fact they do not.

Hence, looking at the metaphilosophical debate, there does not seem to be an agreement on how intuitions should be characterized, nor on whether it is appropriate to qualify the kind of judgements philosophers appeal to as 'intuitive'. One may wonder whether philosophers should stop talking about intuitions altogether.

As it happens, the issue has attracted the attention of scholars from several areas (metaphysics, philosophy of language, epistemology, and psychology), and it has pushed many philosophers to further investigate closely related matters (for instance, the notion of *a priori* and the plausibility of the idea that philosophical knowledge can be gained from the armchair). It remains an important concern for those interested in philosophical methodology. Furthermore, in spite of the disagreement on the characterization of 'intuition', a common ground for those interested in methodological questions may be identified. What they all acknowledge is that (1) verdicts expressed at the end of the so-called thought experiments (TEs) are examples of the kind of judgements we are interested in, and that (2) philosophers, whether legitimately or not, do in fact use 'judgements about cases' to build, support or attack theories.

Therefore, in what follows, I will focus on thought experiments (TEs) and reflective equilibrium (RE)—the two methods in which the appeal to judgements about cases is explicit. TEs are first introduced to present the idea of philosophical intuition (Chap. 2). RE is discussed at length, as it involves TEs (Chap. 3).

I will focus preliminarily on the *role* of judgements about cases in the construction and justification of theories answering Socratic questions. Namely, understanding their role in the philosophical practice will help us shed light on the nature of the aims and results of philosophical inquiries and, consequently, on the *nature* of intuitions.

References

Alvin Plantinga (1993) Warrant and Proper Function, Oxford University Press.

Bealer, G. 1998. Intuition and the Autonomy of Philosophy. In *Rethinking Intuition: The Psychology of Intuition and Its Role in Philosophical Inquiry*, eds. M. DePaul and W. Ramsey, 201–239. Lanham, MD: Rowman and Littlefield.

Cappelen, H. 2012. *Philosophy without Intuitions*. Oxford: Oxford University Press.

Devitt, M. 2006. Intuitions in Linguistics. *British Journal for the Philosophy of Science* 57: 481–513.

Gettier, E. 1963. Is Justified True Belief Knowledge? *Analysis* 23: 121–123.

Gopnik, A., and E. Schwitzgebel.1998. Whose Concepts are they, Anyway?: The Role of Philosophical Intuition in Empirical Psychology. In *Rethinking Intuition: The Psychology of Intuition and Its Role in Philosophical Inquiry*, eds. M. DePaul and W. Ramsey. Lanham, MD: Rowman & Littlefield.

N. Goodman (1955), Fact, Fiction, and Forecast, MA: Harvard University Press, Cambridge.

Hintikka, J. 1999. The emperor's new intuitions. *Journal of Philosophy* 96(3): 127–147.

Kornblith, H. 1998. The Role of Intuition in Philosophical Inquiry: An Account with No Unnatural Ingredients. In *Rethinking Intuition: The Psychology of Intuition and Its Role in Philosophical Inquiry*, eds. M. DePaul and W. Ramsey. Lanham, MD: Rowman & Littlefield.

Lewis, D. 1973. *Counterfactuals*. Oxford/Cambridge: Blackwell Publishers/ Harvard University Press (reprinted with revisions, 1986).

Lewis, D. 1983a. *Philosophical Papers*, vol I. Oxford: Oxford University Press.

Ludwig, K. 2007. The Epistemology of Thought Experiments: First Person versus Third Person Approaches. *Midwest Studies in Philosophy* 31: 128–159.

Lycan, W. (1988). Moral facts and moral knowledge. In W. Lycan (Ed.), Judgment and justification. Cambridge: Cambridge University Press.

Plato. 1901. *The Republic*, Translation of Benjamin Jowett, web edition http://studymore.org.uk/xpla0.htm Date Accessed 28 December 2015. Stephanus page numbers used.

Pust, J. 2012. Intuition. *The Stanford Encyclopedia of Philosophy* (Fall 2014 Edition), ed. Edward N. Zalta, URL = http://plato.stanford.edu/archives/fall2014/entries/intuition/. Date accessed 28 December 2015.

Rawls, J. 1951. Outline of a Decision Procedure for Ethics. *Philosophical Review*, 60(2): 177–97; reprinted in Rawls, J. 1999. *Collected Papers*, 1–19. Cambridge, MA: Harvard University Press.

Rawls, J. 1971. *A Theory of Justice*. Cambridge, MA: Harvard University Press; edition used: Rawls, J. 1972. *A Theory of Justice*. Oxford: Clarendon Press.

Sosa, E. 1998. Minimal Intuition. In *Rethinking Intuition: The Psychology of Intuition and Its Role in Philosophical Inquiry*, eds. M. DePaul and W. Ramsey. Lanham, MD: Rowman & Littlefield.

Van Inwagen, P. 1997. Materialism and the Psychological-Continuity Account of Personal Identity. *Philosophical Perspectives* 11: 305–319.

Williamson, T. 2007. *The Philosophy of Philosophy*. Malden, MA: Blackwell.

2

Thought Experiments

In the 'Introduction to Part I', I said that philosophical intuitions are usually presented through thought experiments (TEs), because TEs are the 'loci' where the appeal to intuitions is generally acknowledged and intuitive judgements are easy to identify. First, I introduced a case presented by Plato in *The Republic*. Then, I reconstructed the arguments taking a reasoner from the question 'What is knowledge?' to the JTB theory of knowledge and mentioned Gettier counterexamples. In this chapter, I report Gettier's version of one of these cases. Then, I introduce another example recurring in the literature on TEs: Thomson's ailing violinist case. The goal is to discuss the aspects TEs have in common and to highlight the role they have inside the reflective dynamics.

2.1 Popular Cases: Gettier Counterexamples and the Ailing Violinist Case

Here is one of the cases introduced by Gettier (1963) to show that knowledge is not justified true belief. The verdict refuting the theory is given in bold type.

© The Author(s) 2016

S.M. Nicoli, *The Role of Intuitions in Philosophical Methodology*,
DOI 10.1057/978-1-137-56715-4_2

> Suppose that Smith and Jones have applied for a certain job. And suppose that Smith has strong evidence for the following conjunctive proposition: (1) Jones is the man who will get the job, and Jones has ten coins in his pocket. Smith's evidence for (1) might be that the president of the company assured him that Jones would in the end be selected, and that he, Smith, had counted the coins in Jones's pocket ten minutes ago. Proposition (1) entails: (2) The man who will get the job has ten coins in his pocket. Let us suppose that Smith sees the entailment from (1) to (2), and accepts (2) on the grounds of (1), for which he has strong evidence. In this case, Smith is clearly justified in believing that (2) is true. But imagine, further, that unknown to Smith, he himself, not Jones, will get the job. And, also, unknown to Smith, he himself has ten coins in his pocket. Proposition (2) is then true, though proposition (1), from which Smith inferred (2), is false. In our example, then, all of the following are true: (i) (1) is true; (ii) Smith believes that (2) is true, and (iii) Smith is justified in believing that (2) is true. But it is equally clear that **Smith does not know that (2) is true**. (Gettier 1963, p. 122)

If the JTB theory of knowledge were true, it would follow that the subject in the Gettier case would know that (2), but it is clear that he doesn't, or, better, it is clear that we wouldn't say that he does: given the way we use 'knowledge' (or given our concept of knowledge), (2) wouldn't be called 'knowledge' (wouldn't be categorized as *knowledge*). By reflecting on the Gettier case epistemologists realized that conditions (i), (ii) and (iii) are necessary but not sufficient for saying that, in that situation, there is knowledge. What is the extra condition, or what are the extra conditions, one should make explicit to draw up a list of characteristics sufficient to define it?

In the years following the presentation of the cases, different solutions have been proposed. The strategies consisted in adding a fourth condition to the three listed by the JTB theory of knowledge (JTB + *Y*),[1] or in strengthening the justification condition so to exclude Gettier cases from

[1] Examples are the so-called 'no false lemmas' (Armstrong 1973; Clark 1963), 'sensitivity' (Nozick 1981); 'safety' (Sosa 1999), 'relevant alternatives' (Stine 1976; Goldman 1976; Dretske 1981) conditions.

cases of justified belief.[2] Either by integrating or by modifying it, the goal of each strategy was to ‘immunize’ the JTB theory of knowledge from examples of Gettier’s kind. Actually, many of these ‘new’ accounts met this requirement; however, they also turned out to be vulnerable to other counterexamples. To date, understanding what the complete analysis of knowledge is remains an open problem. I will not delve into these questions (*see* Ichikawa and Steup 2012). In fact, concerning the role played by Gettier cases, the important issue is this: the intuitive evaluation we give of the described scenarios contradicts the predictions of the JTB theory of knowledge.

I said that, by reflecting on Gettier cases, epistemologists realized that conditions (i), (ii) and (iii) were not sufficient for saying that, in such and such situation, there is knowledge. However, by thinking about cases, one could also conclude that a certain theory is mistaken and calls for a proper revision. Distinguishing intuitions elicited by cases of Gettier’s kind from those elicited by cases of this last kind, Pust (2000) writes:

> A robust intuition that the analysandum term does not apply to a particular case even though the case has all the features specified in the analysans is usually taken to show that the analysis fails to provide sufficient conditions. A robust intuition that the analysandum term applies to a particular case but that the case lacks one of the features specified in the analysans is usually taken to show that the features the analysis holds to be necessary are not in fact necessary. (Pust 2000, p. 3)

Hence, there are cases showing that the conditions stated by a theory are not necessary, and cases showing that the conditions stated by a theory are not sufficient. However, as Pust himself grants, not all TEs can be described in this manner. This is the case of many TEs that describe moral scenarios.

Let us consider, for instance, Thomson’s violinist case. In her popular article, ‘A defence of abortion’ (1971), Thomson asks the reader to imagine the case of an ailing and unconscious violinist and to assess whether

[2] For instance, the justification condition in the JTB theory is replaced either with the reliability condition or with a condition requiring a causal connection between the belief and the fact believed (Goldman 1967, 1976).

it is legitimate (or not) to prevent a person who was kidnapped and whose circulatory system was (forcedly) plugged into the violinist's to unplug him or her, causing the death of the violinist. Thomson's prediction is that one would recognize the right of the unwilling donor to be unplugged from the violinist and that one would accept the analogy with the case of abortion; if so, one would also be ready to acknowledge that the thesis according to which the violinist/foetus right to life outweighs the donor's/woman's right to decide what happens to his or her body is unacceptable. Let us see her argument in detail.

After having taken for granted that the foetus is a person from the moment of conception, Thomson enunciates anti-abortionists' thesis (a foetus's right to life outweighs a woman's right to decide what happens to her body) and then the case of the ailing violist:

> Let me ask you to imagine this. You wake up in the morning and find yourself back to back in bed with an unconscious violinist. A famous unconscious violinist. He has been found to have a fatal kidney ailment, and the Society of Music Lovers has canvassed all the available medical records and found that you alone have the right blood type to help. They have therefore kidnapped you, and last night the violinist's circulatory system was plugged into yours, so that your kidneys can be used to extract poisons from his blood as well as your own. The director of the hospital now tells you, 'Look, we're sorry the Society of Music Lovers did this to you—we would never have permitted it if we had known. But still, they did it, and the violinist is now plugged into you. To unplug you would be to kill him. But never mind, it's only for nine months. By then he will have recovered from his ailment, and can safely be unplugged from you.' Is it morally incumbent on you to accede to this situation? No doubt it would be very nice of you if you did, a great kindness. But do you have to accede to it? What if it were not nine months, but nine years? Or longer still? What if the director of the hospital says: 'Tough luck, I agree. But now you've got to stay in bed, with the violinist plugged into you, for the rest of your life. Because remember this: All persons have a right to life, and violinists are persons. Granted you have a right to decide what happens in and to your body, but a person's right to life outweighs your right to decide what happens in and to your body. So you cannot ever be unplugged from him.' I imagine you would regard this as outrageous, which suggests that

> something really is wrong with that plausible-sounding argument I mentioned a moment ago. (Thomson 1971, pp. 70–71)

What do the three cases (Gettier counterexamples, the Platonic dialogue and the Ailing Violinist case) have in common?

They all reveal a mismatch between the way in which we judge and the way in which a thesis or theory would require us to judge. More precisely, there is a theory, thesis or definition prescribing that, in the relevant case, the correct evaluation to give is *H*: if the JTB theory of knowledge were true, the subject in the Gettier case would know that the man who will get the job has ten coins in his pocket; if Cephalus's definition of justice were true, then telling the truth and giving back the arms to a crazed friend would be the right thing to do; if a foetus's right to life outweighed the woman's right to decide what happens in her body, then the doctor would be right to refuse to unplug the unwilling donor from the violinist. However, in each of these cases, we would say the opposite (that non-*H*): in the Gettier case we would say that Smith doesn't know that the man who will get the job doesn't have ten coins in his pocket; in the case of the crazed friend we would say that telling the truth and giving back the arms is the wrong thing to do; in the case of the ailing violinist we would say that denying the unwilling donor the right to unplug from the violinist is unjust (outrageous). The theory, thesis or definition is therefore incorrect (or somehow incomplete). If one wants to carry on with the inquiry on that topic, one is then committed to go back to the theory, try to adjust it, or build a new one.

In sum, TEs work in a way analogous to actual counterexamples: a thesis or theory predicts that, in a certain situation, the correct evaluation to give is *H*. In that situation, however, we would say that non-*H*. The correctness of the theory is therefore questionable. But, how is it possible that imaginary counterexamples (TEs) and actual counterexamples are equally effective?

In order to answer this question, I will introduce the analysis of the logical structure of TEs presented by Williamson (2007). His discussion on the epistemology of modal thinking and counterfactual thinking is complex and, to a certain extent, controversial. In the next section, I am not going to examine it in detail. Rather, I will highlight the aspects of his analysis that are useful in answering the question just raised.

2.2 Imaginary Counterexamples

The imaginary case used by Williamson to analyse the logical structure of TEs is a Gettier case. The first step consists in giving a description of the generalization Gettier cases aim to attack: 'necessarily, for any subject *x* and proposition *p*, *x* knows *p* if and only if *x* has a justified true belief in *p*' (Williamson 2007, p. 183). In symbols:

I. $\Box \forall x \forall p\ (K(x, p) \leftrightarrow \text{JTB}\ (x, p))$

The second step is to describe a particular Gettier case (GC). First of all, the details of Gettier's narrative are replaced by variables and the story is given by the neutral description GC (*x*, *p*), where variable *x* is the person who believes (Smith) and variable *p* the believed proposition, that is, the content of the true and justified belief (the man who will get the job has ten coins in his pocket). Then, Williamson identifies the two premises the objection is grounded on:

- a premise of possibility: a subject *x* could have stood as described to a proposition *p*. In symbols:

II. $\Diamond \exists \times \exists p\ \text{GC}\ (x,p)$

- a premise expressed by means of a counterfactual or subjunctive conditional: if someone stood to a proposition in the described relation, then one would have had justified true belief without knowledge in respect of that proposition. In symbols:

III. $\exists \times \exists p\ \text{GC}(x,p) \;\Box\!\!\rightarrow \forall \times \forall p\ (\text{GC}(x,p) \supset (\text{JTB}(x,p)\ \&\ \neg K(x,p))$

III has the structure of a counterfactual conditional , '$q \Box\!\!\rightarrow r$', that is, of that particular conditional in which the consequent (*r*) describes what would have happened if the situation described in the antecedent (*q*) had occurred.

Given II and III the conclusion follows: someone, *x*, could have had justified true belief without knowledge in respect of some proposition, *p*. In symbols:

IV. $\Diamond \exists \times \exists p \; (\mathrm{JTB}(x,p) \;\&\; \neg K(x,p))$

So given that IV is inconsistent with I (in particular with its right-to-left direction), then the three conditions settled by the traditional analysis are not sufficient for knowledge.

Let us then go back to our question: how is it possible that an imaginary counterexample is as effective as a real one? Details of Williamson's analysis aside: given that the generalization at the basis of the traditional analysis of knowledge states that *necessarily*, for any subject *x* and proposition *p*, *x* knows *p* if and only if *x* has a justified true belief in *p*, then, in order to refute it, it is sufficient to show that it is *possible* that someone has a justified true belief that *p* without knowing that *p*. Gettier's TE proves this.

In general, this seems to be true for any TE aimed at disproving a philosophical thesis: given that the generalizations at the basis of philosophical analysis are in the form of *necessarily H*, then showing that it is *possible that not H* is enough to reject them.

Obviously, this does not amount to saying that TEs are all equally apt to show that it is possible that non-*H*. Williamson himself underlines this. Speaking about premises II and III above, he argues the following. First, premise II seems to be less problematic than premise III: II says only that the Gettier case is possible; and in fact, II is usually uncontested: even if rare, 'Gettier cases are not far-out science fictions but mundane practical and physical possibilities' (Williamson 2005, p. 10). However, there are also cases where the truthfulness of II cannot be taken for granted. Williamson mentions far-out science fictions as, for instance, the brain in the vat, a case in which we are asked to imagine a disembodied brain stimulated so to have conscious experiences such as those of a normal perceiver; or Chalmers' zombies, creatures being physical duplicates of human beings but lacking conscious experience (*see* Williamson 2007, p. 189). As regards these TEs, many doubt II as well.

2.3 Exceptionality and Persuasion

Once it has been clarified in which sense TEs can play their role as successfully as actual counterexamples, we may consider a second kind of question: how can it be that a fully fledged theory is called into question by the simple exhibition of a case? In particular: how is it possible that, for a certain (sometimes long) period of time, a person or a group of people (the community of experts) embraces a theory without taking into account the case or the class of cases the experiment describes?

Let us focus on Gettier cases again. Formerly, I described the content of Gettier's intuition in the following way: given the way we use 'knowledge' (or given our concept of knowledge), the Gettierized belief wouldn't be called 'knowledge' (wouldn't be categorized as *knowledge*). This is a verdict epistemologists come to on the basis of their semantic (conceptual) competence, that is to say their capacity to use 'knowledge' in a way that reflects the way that word is used in the community. Looking at it from this perspective, the judgement expressed is just like any other judgment epistemologists made in advancing the JTB theory of knowledge; that is, the theory whose incompleteness is shown by the judgments prompted by Gettier cases. Hence, one may wonder why the set of intuitive judgements epistemologists built the JTB theory on does not include Gettier's judgement. The most obvious answer to this question is that Gettier cases describe unusual situations, that is, situations we rarely come across or happen to think of.

Hence, let us go back to the initial question: why would the consideration of a new (real or imaginary) case call into question a thesis we thought sound? The reason seems to be that by thinking about new cases, and in particular unusual cases, we become aware of our opinion on X in those cases. In particular, thanks to the examples described in TEs, we find our opinions to be inconsistent with our—perhaps laboriously earned—theoretical convictions. Contrary to what we had supposed, our theory on X is not the best explication of our opinion on X.

The idea that the possibility of acquiring new knowledge of the object of our inquiry is due to the rarity of the cases allows me to introduce the thesis of exceptionality. This thesis has been endorsed by Gendler: 'When the contemplation of an imaginary scenario brings us to new knowledge, it does so by forcing us to make sense of an exceptional case' (Gendler 2000, p. 1). I will not explain here what Gendler exactly means by 'exceptional case', nor consider how she thinks exceptional cases should be handled. In what follows, I will just see how the concept of exceptionality is useful in answering our question.

First, the concept of exceptionality is applicable to TEs in general. It is clear the sense in which cases described in TEs are exceptional relatively to theories they refute: the cases described refute the theory precisely because the case is an exception in comparison to the ones that have been considered in the stage of theory development, that is, the cases that the theory explicates. Exceptionality, however, does not only regard this patent aspect. Cases provided by TEs are exceptional in the sense in which Gettier examples or the Platonic case are: by describing unusual situations. This second sense of exceptional is interesting: in fact, it explains the enduring conviction that our theory is nevertheless satisfactory, the difficulties in coming to terms with cases of this kind, and the shocking impact they have on the person (or on the community of people) embracing the theory.

I have mentioned Gettier examples and the case Plato describes in *The Republic*. What about the case of the ailing violinist? Like Gettier examples and the Platonic case, the case of the ailing violinist is exceptional: it describes an extraordinary situation, or better, a situation on the verge of utter unlikelihood. However, the case of the ailing violinist is also dissimilar from the other two: namely, whereas Gettier's imaginary cases and the imaginary case described by Plato are categorical to the theories they refute, Thomson's imaginary case is conditional in respect to the anti-abortionists' thesis. What I mean by this last claim is that the violinist case counts as a counterexample to the claim it aims to refute *provided* the analogy with the case of the woman and foetus is accepted—*provided* that between the two cases (the imaginary case of the violinist and the donor and the real case of the foetus and the woman) there are no relevant differences impairing judgement. Let us put ourselves in the shoes of an anti-abortionist: what could be his or her reaction to Thomson's conclusion? The anti-abortionist could insist on the differences between the two cases and refuse the analogy. Hence, for the purpose of a criticism of the anti-abortionist's perspective, the analogy is both problematic and useful.

How is the analogy useful? Re-describing the foetus–mother relationship in terms of a relationship between an ailing violinist and an unwilling donor allows to consider the situation that is the object of the controversy between abortionists and anti-abortionists from a new standpoint: from a perspective free from explicit allusions to the foetus and to the woman. This should enable judges to evaluate the case judge on the basis of their pre-theoretical convictions (according to their moral capacity), in lieu of their theoretical (religious, moral) convictions.

This idea could also be expressed in a slightly different way. One could argue that the re-description of a problematic situation in an unprecedented manner has the goal of providing the interlocutor with resources enabling him or her to make a perspective shift. This is the position that Gendler (2007) defends. Speaking about the role of moral and political philosophy, Gendler stresses the importance to design cases (she speaks of 'images') that make perspective shifts possible (Gendler 2007, p. 83). Apropos, she mentions Thomson's violinist example. Gendler claims that the experiment can work 'if it brings out a reframing of the subject's attitudes in the domain it is intended to illuminate—if it comes, either reflectively or unreflectively, to represent the question of the fetus–mother relationship in ways akin to those that he represents the violinist–patient relationship' (Gendler 2007, p. 86). If the subject accepts the analogy, then the perspective shift occurred. The probability of Thomson persuading her interlocutor is now greater than before.

In general, Gendler claims, TEs 'evoke responses that run counter those evoked by alternative presentations of relevantly similar content [...] When thought experiments succeed as devices of persuasion, it is because the evoked response becomes dominant, so that the subject comes (either reflectively or unreflectively) to represent relevant non-thought experimental content in light of the thought-experimental conclusion' (Gendler 2007, p. 86).

To conclude, the exceptionality thesis helped us to answer the question of how a previously accepted theory or thesis could be questioned by the simple exhibition of a case: by thinking about unusual cases or by considering alternative presentations of cases close to home, one becomes aware of his or her opinion about a certain topic in such cases. Since, contrary to what one had supposed, one's opinion turns out to be inconsistent with one's previous 'theoretical' convictions, then one is committed to go back to the theory to try to adjust it, or explain that judgement away. However, why should one try to account for intuitive judgements disagreeing with theoretical beliefs? According to the supporters of reflective equilibrium method (RE), the answer is connected the reflective equilibrium requirement.

To deepen our understanding of these dynamics, in the next chapter I will analyse the version of reflective equilibrium method proposed by Goodman, Rawls, and Lewis, and I will compare them with Carnap's method of explication.

References

Armstrong, D.M. 1973. *Belief, Truth, and Knowledge*. Cambridge: Cambridge University Press.

Clark, M. 1963. Knowledge and Grounds. A Comment on Mr. Gettier's Paper. *Analysis* 24: 46–48.

Dretske, F. 1981. *Knowledge and the Flow of Information*. Cambridge: MIT Press.

Gendler, T.S. 2000. *Thought experiment: on the powers and limits of imaginary cases*. London: Garland Press, now Routledge.

Gendler, T.S. 2007. Philosophical Thought Experiments, Intuitions, and Cognitive Equilibrium. *Midwest Studies in Philosophy of Science* 31: 68–89.

Gettier, E. 1963. Is Justified True Belief Knowledge? *Analysis* 23: 121–123.

Goldman, A. 1967. A Causal Theory of Knowing. *The Journal of Philosophy* 64(12): 357–372.

Goldman, A. 1976. Discrimination and Perceptual Knowledge. *The Journal of Philosophy* 73: 771–791.

J. J. Ichikawa, M. Steup (2012) 'The Analysis of Knowledge', The Stanford Encyclopedia of Philosophy (Spring 2014 Edition), Edward N. Zalta (ed.), URL = . Date accessed 28 December 2015.

Nozick, R. 1981. *Philosophical Explanations*. Cambridge, MA: Harvard University Press.

Pust, J. 2000. *Intuitions as evidence*. New York: Garland Publishing.

Sosa, E. 1999. How to Defeat Opposition to Moore. *Noûs* 33(supplement: Philosophical Perspectives, 13, Epistemology): 141–153.

Stine, G. 1976. Skepticism, Relevant Alternatives, and Deductive Closure. *Philosophical Studies* 29: 249–261.

Thomson, J.J. 1971. A defence of abortion. *Philosophy and Public Affairs* 1: 47–66.

Williamson, T 2005. Armchair Philosophy, Metaphysical Modality and Counterfactual Thinking. *Proceedings of the Aristotelian Society* 105: 1–23.

———2007. *The Philosophy of Philosophy*. Malden, MA: Blackwell.

3

Reflective Equilibrium

3.1 A Method for Logic, Ethics and Much More: Goodman's and Rawls's Reflective Equilibrium

The expression 'reflective equilibrium' (RE) was introduced for the first time by Rawls in *A Theory of Justice* (1971) and it was intended to describe the procedure regulating the formulation of acceptable principles of justice. Rawls finds the origins of his method in a proposal advanced by Goodman in his article 'The new riddle of induction' (1955). In this classic text, Goodman deals with a long-debated problem in the history of philosophy: the legitimacy of the rules of inductive logic—also known as 'Hume's riddle'. Goodman claims that (both deductive and inductive) inferential rules can be justified by their conformity with accepted inferential practice. More precisely: in order to decide whether a rule is justified or not, one should determine if that rule yields the particular inferences we actually make and sanction, that is, whether it agrees with our judgements to the effect that the inference is (or isn't) correct. According to Rawls, this approach to the justification of rules is valid for moral rules

© The Author(s) 2016

S.M. Nicoli, *The Role of Intuitions in Philosophical Methodology*,
DOI 10.1057/978-1-137-56715-4_3

as well: proposed principles of justice are correct if they can adequately account for our intuitive judgements of 'just' or 'unjust' about cases.

Let us consider 'The new riddle of induction'. Goodman proposes to deal with the problem of induction within the broader issue of figuring out what legitimates the adoption of any inferential rule. So, Goodman's first move is to look at deductive inferences and ask what a justification amounts to in the case of deductive reasoning.

Let us take this reasoning: 'If it is daytime, then there is light. It is daytime. Therefore, there is light'. How is it justified? By showing that it is an instance of *Modus Ponens*: 'IF *P*, then *Q*. *P*. Therefore *Q*'. This reasoning is a deductive inference and it is valid because it is an instance of a valid deductive rule. According to Goodman, the criterion used to justify this particular inference—conformity to a valid rule—is the one used to justify deductive and inductive inferences in general.

Once established that an argument (deductive or inductive) is justified when it is a correct application of valid logical principles, Goodman asks: What justifies a rule?

Goodman refuses to advance hypotheses as, for instance, Aristotle's and Descartes' idea of the existence of self-evident axioms, or Kant's hypothesis of rules grounded in the very nature of the human mind, arguing that a satisfactory answer can be found 'much nearer the surface' (Goodman 1955, p. 67). In short, as a criterion for the acceptance of deductive and inductive principles, Goodman proposes conformity to the accepted inferential practice: accordance with the set of the particular inferences 'we actually make and sanction' or, better yet, with 'judgements rejecting or accepting particular inferences' (Goodman 1955, p. 67). So, generally speaking, Goodman's thesis is the following: rules are justified by the set of inferences that we judge to be correct, that is, by our intuitive judgements regarding which inferences are valid and which are not.

However, how exactly do judgements of validity support a rule? On this point, Goodman is not explicit. It is however possible to suppose that he drawing an inference to the best explanation (IBE): to infer the rule underlying the supposedly valid inferences, we consider said inferences as complying with a unique model, that is, a (proposed) rule. The rule is the best explanation of those inferences.

Hence, Goodman's solution to the problem of justification can be summarized as follows: a particular inference is valid when it complies with a valid rule (theory); that rule (theory) is legitimate if it produces individual inferences we usually assess as valid.

The first impression is that this is a circular argument. The objection of circularity does not elude Goodman. However, he does not consider his argument to be circular in a 'bad way'. On the contrary, he claims that the viciousness, typical of justificatory circles, is in this case absent:

> But this circle is a virtuous one. The point is that rules and particular inferences alike are justified by being brought into agreement with each other. *A rule is amended if it yields an inference we are unwilling to accept; an inference is rejected if it violates a rule we are unwilling to amend.* The process of justification is the delicate one of making mutual adjustments between rules and accepted inferences; and in the agreement achieved lies the only justification needed for either. (Goodman 1955, p. 67)

Thus, Goodman suggests that the virtuousness of the circle can be clarified by highlighting the procedure leading to agreement between rules and accepted inferences, and that this solves the problem of the justification of both individual inferences and principles. But then, what is the structure of the 'delicate process' having the important role of producing the agreement of rules and judgements of validity—an agreement that is necessary and sufficient for the justification of both? We know what the relation between rules and inferences is once the agreement is reached: the individual inferences we sanction support the rule by IBE. In particular, if the rule is conceived as an infinite class of inferences of the same form, then the rule justifies all inferences belonging to it. Hence, a rule *R* justifies an inference *I*, if *I* belongs to the class of inferences specified by *R*. However, concerning the nature of the procedure of 'bringing them into agreement' very little is known: Goodman says that when an inconsistency between rules and judgements occurs, adjustments and corrections are allowed on both sides. Yet, we completely ignore the details of the process, that is, the way in which (and the extent to which) adjustments of the rules and the judgements of acceptability/unacceptability are (can be) made.

In what follows, I shall provide two possible interpretations of the balancing process leading to final equilibrium. The first interpretation postulates the existence of two stable cores (a subset of the set of judgements and a subset of the set of rules) governing the procedure and determining its outputs. The second interpretation describes reflective dynamics as a 'free game' in which the balancing procedure between theory and intuitions has no preferential direction and adjustments are wide-ranging. I will argue in favour of this second interpretation. Finally, I will consider how vicious circularity can be prevented.

3.1.1 A Free Game

How would the process of balancing the rules (the theory) and the accepted inferential practice (judgements of validity about single inferences) appear, if some rules and some judgements were considered as fixed, that is, non-adjustable?

According to Goodman, the reciprocal adjustments of rules and inferences can occur either by adjusting or rejecting one or more rules, or by rejecting one or more judgements of validity. Now, supposing that, over and beyond such reflective dynamics, there is a group of inderogably valid rules and a set of non-negotiable intuitions, this operation will be understood as follows: a (derogable) rule is amended if it yields inferences contradicting some non-negotiable intuitions; a (negotiable) intuitive judgement is abandoned if it contradicts an inderogable rule.

Thus the first interpretation of the balancing process leading to equilibrium postulates, from the beginning, a clear-cut distinction between aspects that can be modified (or, if necessary, amended) by reflective dynamics and aspects directing such adjustments. Namely, inside the set of rules and the set of judgements there are, respectively, two subsets: the subset of the rules we think are in any case correct, and that of the rules we are ready to negotiate; the subset of the non-negotiable judgements, and that of the judgements we are ready to question.

During the reflective process the 'strong' theoretical core (constituted by the rules we regard as inderogable) and the stable intuitive core (made of non-negotiable verdicts) are held fixed, and one uses them as a guide to modify the other two subsets. Consequently, the process of bringing theories and intuitions into agreement does not appear to be an operation of wide-ranging adjustments and balancing, but rather a 'pruning' work, governed by those theoretical and intuitive aspects we assess to be incorrigible.

Does this interpretation of reflective equilibrium account for what Goodman meant?

The interpretation has the merit of giving a precise sense to the ambivalence of the adjustments (both the theory and the intuitions can be equally modified) and of making the direction of these adjustments explicit (in case of a conflict between theory and intuitions, it is easy to decide what should be preserved and what can be revised). However, it presupposes the existence of two fixed cores (which have never been explicitly theorized by Goodman) excluded from the adjustments and governing the dynamics.

Furthermore, there is a feature of Goodman's suggestion that the first interpretation misses: circularity. An argument is circular when it derives *P* from *Q* and then *Q* from *P*, that is to say when *Q* justifies *P* and *P* justifies *Q*. No matter whether such circularity will come out to be virtuous or not, it is evident that, in the first interpretation, the feature of circularity itself is absent: negotiable inferences and rules are justified as being in agreement with non-negotiable inferences and inderogable rules, whereas the latter do not need any justification.

Another interpretation is needed. Let us consider a theory, T_1, which has been developed by an inference to the best explanation from an initial set of intuitions, I_1. T_1 entails a set of consequences. Inside this set there are consequences that agree with the initial intuitions, I_1, and others that unexpectedly contradict some other intuitions of ours. This is, for example, the case of a logical rule that explains a considerable part of the inferences we usually make and sanction, but also entails a certain kind of reasoning we would never consider as valid. T_1 is satisfactory insofar as it pre-supposes I_1, dubious insofar as it entails I_2. If the consequences, I_2, are felt to be irreparably counterintuitive, then T_1 has to be abandoned and a new theory has to be built. The constraint for the new theory, T_2, is to presuppose I_1, as T_1 did, and not to entail I_2. Let us suppose that T_2 actually explains I_1 and that its consequences do not contradict the intuitions T_1 contradicted. Still, it can happen that T_2 also bears some unwanted consequences (I_3). I_3 conflicts with some of our intuitions, but in this case we find T_2 persuasive. So we decide to preserve the theory and revise the set of intuitions: we remove the intuitions that are in conflict with the consequences of the theory and we accept the consequences I_3 as new intuitions. On the side of intuitions, this transformation can be seen as a *Gestalt switch*.

This description of reflective dynamics differs from the previous one because it allows adjustments and corrections to be wide-ranging: there are no intuitions or rules that are, as a matter of principle, unmodifiable. Furthermore, adjustments and corrections occur on both sides: one does not only bring the theory in accordance with intuitions, but one also modifies one's opinions on what is acceptable and what is not, on the basis of the theory. In this case, an enhancement is obtained both on the theory's and the intuitions' side.

In virtue of the complex balancing mechanism established between theory and intuitions, the justificatory circle can be seen as a virtuous one. In fact, the relation of coherence that is established between the theory, such as it appears at the end of the procedure, and the modified intuitions is highly articulated and much more complex than a relation of mere circularity. Based on this description of the method, it is insufficient to say that T_2 is justified because it explicates I_3 and I_1, and that I_3 and I_1 are justified because they are entailed by T_2. Of course T_2 is the best explanation inferred from I_3 and I_1, but it has further merits: aside from entailing I_3 and I_1, we know that T_2 has a negative relation with I_2. With respect to intuitions, it is true that I_3 and I_1 are explained by T_2, but we cannot properly say they are justified by the theory. The pre-theoretical set (I_1) is made of intuitions that are already accepted and needed to define the field of our research. Something similar happens with *I*3: the new set of intuitions becomes—in a certain way—independent: once we are convinced that the new intuitions are correct, we believe them regardless of the fact that they are consequences that could be inferred from T_2. The idea is that intuitions as such have their own solidity (we come to judge in a way that is in fact consistent with the theory, without appealing to the theory), and their justification depends primarily on their acceptance.

In conclusion, the structure emerging from this second analysis is a complex network describing how intuitions and theory support each other, and making the initial concern on vicious circularity disappear. Furthermore, while the first interpretation of reflective dynamics (the interpretation based on fixed cores) was not adequate to explain what Goodman probably meant by 'virtuous' circularity, this new interpretation is.

3.1.2 *A Theory of Justice*, an Example of Philosophical Negotiation

In *A Theory of Justice* (1971), Rawls explicitly introduces the idea of reflective equilibrium, identifying the origins of the method in Goodman (1955). However, the most significant anticipation of the methodological proposal presented in *A Theory of Justice* is not to be found in 'The new riddle of induction', but rather in 'Outline of a decision procedure for ethics', an article written by Rawls in 1951, in which 'a reasonable method for validating and invalidating proposed moral rules' (Rawls 1951, p. 177) is investigated.

In the following section, I will present the thesis of reflective equilibrium as it is described in *A Theory of Justice*. References to the method defined in 'Outline of a decision procedure for ethics' will be made whenever important analogies with the theses presented in *A Theory of Justice* emerge.

In order to avoid the risk of being too generic in engaging with question 'What is Justice?', Rawls provides in *A Theory of Justice* a detailed account of the research field from the very beginning:

> Many different kinds of things are said to be just or unjust: not only laws, institutions, and social systems, but also particular actions of many kinds, including decisions, judgments, and imputations. We also call the attitudes and dispositions of persons, and persons themselves, just or unjust. Our topic, however, is that of social justice. For us the primary subject of justice is the basic structure of society, or more exactly, the way in which major social institutions distribute fundamental rights and duties and determine the division of advantages of social cooperation. (Rawls 1972, p. 7)

Once established that the object of a theory of justice is society, and in particular a well-ordered society, the question 'What is justice?' is rephrased in the following way: what is the conception of justice which would rule a well-ordered society?

Let us start from the concept of society. Rawls defines society as an association of persons cooperating and acting in accordance with certain recognized (and binding) rules of conduct. Rawls points out that cooperation with others is useful for the individual 'since social cooperation makes possible a better life for all than any would have if each were to live solely by his own efforts' (Rawls 1972, p. 4). However, cooperation is also characterized by conflict, since each member of society tends to prefer, for himself or herself, a larger part of the benefits resulting from cooperation. So, the cooperative venture presents two aspects: identity and conflict of interests. Identity and conflict of interests (the so-called 'circumstances of justice') motivate the necessity of a public conception of justice, that is, the need to reach a good level of agreement on the standards (on the principles) regulating the proper distributive shares.

Rawls proceeds by pointing out that existing societies are rarely regulated, effectively, by a public *conception* of justice. On the contrary, within different communities there is a certain amount of disagreement about the principles that are supposed to define the appropriate distribution of social benefits and burdens. However, it seems possible to identify a shared *concept* of justice and describe its requirements:

> Those who hold different conceptions of justice can, then, still agree that institutions are just when no arbitrary distinctions are made between persons in the assigning of basic rights and duties and when the rules determine a proper balance between competing claims to the advantages of social life. (Rawls 1972, p. 6)

In short, the concept of justice is widely accepted because it does not assign any precise value to expressions such as 'non-arbitrary distinctions' and 'proper balance': supporters of different conceptions are free to assign different meanings to these two. But, whilst the concept of justice as described by Rawls cannot be an adequate response to the question of what is justice, it constitutes a prerequisite for its formulation: namely, the concept identifies the parameters a correct conception of justice will give precise values to. So, the core of the theory of justice is going to be constituted by two principles: the first determining in a precise manner the non-arbitrary way in which rights and duties are to be assigned; the second making explicit the criteria for an appropriate distribution of resources.

The next question Rawls raises is: how can one identify the two principles and assess their correctness? Rawls's idea is that the principles of justice are the object of an original agreement: 'they are the principles that free and rational persons concerned to further their own interests would accept in an initial position of equality' (Rawls 1972, p. 11). Namely, if the situation in which the principles are chosen is fair, then the principles themselves will be fair as well.

The idea that correctness of the theory depends on the decision procedure that is at the basis of its choice is already present in 'Outline of a decision procedure for ethics': if the procedural standards are reasonable, so will the principles be.

Therefore, the challenge is to describe, in an appropriate manner, the conditions characterizing the initial situation. In 'Outline of a decision procedure for ethics', Rawls defines the characteristics of the deciders (that is, of competent moral judges) and the conditions under which they express their judgements. In *A Theory of Justice*, he introduces the notion of 'original position' and finds in the rationality of the contracting parts and in the veil of ignorance its specific characteristics. These two constraints (the rationality of the subjects who decide and the veil of ignorance), together with the circumstances of justice (the conditions of moderate scarcity) and with the formal constraints of the concept of right (generality, universality, publicity, ordinality, and finality), form the set of the conditions an adequate theory of justice should meet.

The veil of ignorance is a device that Rawls uses to nullify the effects of arbitrary conditions on the choice of principles: namely, if the result is supposed to be a fair and unanimous theory, then contracting parties should abstract from all specific information that could lead them to tailor principles to their advantage. Therefore, subjects under the veil of ignorance have no knowledge of the specific natural and social circumstances that put men at odds: 'no one knows his place in society, his class position or social status, nor does he know his fortune in the distribution of natural assets and abilities, his intelligence, strength, and the like' (Rawls 1972, p. 137). Beyond this, no one knows 'his conception of the good, the particulars of his rational plan of life, or even the special features of his psychology such as his aversion to risk or liability to optimism or pessimism' (Rawls 1972, p. 137). Finally, contractors ignore the particular historical facts about the society they will live in. However, they have no restrictions on knowledge of general facts about human society: 'they understand political affairs and the principles of economic theory; they know the basis of social organization and the laws of human psychology' (Rawls 1972, pp. 137–138), and so on. Generally, there are no limitations on general information.

In addition to being under the veil of ignorance, the subject in the original position is rational and promotes his or her interests: he or she tends to choose such principles as would allow him or her to fulfil his or her rational plan of life. But, how can a subject favour his or her ends and desires, given that the veil of ignorance prevents him or her from knowing such details? Rawls's idea is the following: ignoring the details of one's plan of life, the rational subject would be inclined to prefer for himself or herself more primary social goods (than less). Namely, thanks to one's knowledge of general facts, the subject knows that, in order to promote his or her aims (whatever they may be), he or she would have to access primary social goods. In other terms, one is aware of having, as a person, some basic interests that have to be satisfied regardless of the place in society one will then occupy.

Starting from this situation (a situation that Rawls judges as perfectly fair), it is possible to decide which are the correct values one should give to the parameters of the concept of justice, that is, to determine the principles that would rule a well-ordered society.

Rawls states two principles and stipulates the priority of the first principle over the second.

> First: each person is to have an equal right to the most extensive scheme of equal basic liberties compatible with a similar scheme of liberties for others.

> Second: social and economic inequalities are to be arranged so that they are both (a) reasonably expected to be to everyone's advantage, and (b) attached to positions and offices open to all. (Rawls 1972, pp. 60–61)

The description of the original position and the two principles form the core of Rawls's theory and conception of justice. In particular, the conditions stated in the initial situation determine which kind of principles will be chosen. Rawls writes: 'as the circumstances are presented in different ways, correspondingly different principles are accepted' (Rawls 1972, p. 18). The dynamics binding the original position with Rawls's principles are identical in any set of principles and any possible version of the initial situation. Rawls further specifies that 'there are [...] many possible interpretations of the initial situation. [...] In this sense, there are many different contract theories. Justice as fairness is but one of these' (Rawls 1972, p. 121); and again: 'we may conjecture that for each traditional conception of justice there exists an interpretation of the initial situation in which its principles are the preferred solution' (Rawls 1972, p. 121).

But then, what does it make of a specific description of the initial situation, and in particular of this specific description of the initial situation (that is, the original position), a description that is preferable to any other?

To answer this question, Rawls introduces the reflective equilibrium argument. In short, the method of reflective equilibrium consists in '[checking] an interpretation of the initial situation, then, by the capacity of its principles to accommodate our firmest convictions and to provide guidance where guidance is needed' (Rawls 1972, p. 20). So, the problem of the justification of a conception of justice 'is settled [...] by showing that there is one interpretation of the initial situation which best expresses the conditions that are widely thought reasonable to impose on the choice of principles yet which, at the same time, leads to a conception that characterizes our considered judgements in reflective equilibrium' (Rawls 1972, p. 121).

Let us see in detail how the argument is articulated. First, one should describe the initial situation 'so that it represents generally shared and preferably weak conditions' (Rawls 1972, p. 20). The second step is to see if this version of the initial situation is strong enough to yield a set of principles. If it is not, then one should re-describe it.

However, let us suppose one comes to a description of the initial situation yielding principles. Once such principles are displayed, and premises yielding them appear to be entirely plausible, can one consider one's enterprise as concluded? Not really: the principles could fail to match the considered judgements of 'just' and 'unjust' we express when assessing particular cases. One ought to check whether this is the case or not.

Presumably, some discrepancies will be found, that is, one realizes that the chosen principles contradict some of our considered judgements. What should one do in that case? According to Rawls, there are two possibilities: going back to the initial situation and re-describing it, so that it generates new principles; or else, modifying the intuitive core, rejecting the judgements the theory conflicts with.

Hence, the procedure yielding the formulation of principles and the description of the original position is not an easy and linear process, but rather a complicated structure of comparisons and reciprocal adjustments between our conception of justice (conditions of the original position plus principles) and our intuitive judgements on what is just and unjust. Moreover, the procedure is never exhausted in just a single 'move' (description of the initial situation, formulation of the principles and assessment of their intuitive plausibility), but moves back and forth between the theoretical pole and the pole of considered judgements: one modifies the theory and the intuitive core in turn, until the agreement between the two is reached. When this happens, we reach the reflective equilibrium, that is, that stage of our reflection on a certain problem when the theory we embrace can be said to be the best explanation of our actual convictions concerning the topic under inquiry (in this case, about justice). Obviously, this equilibrium can once more be called into question by new considerations or new intuitions contradicting the theory. However, until the equilibrium is maintained, the justification of our conception of justice remains stable.

In the last part of this section, I will spend a few words on a method described by Rawls in 'Outline of a decision procedure for ethics' that is generally regarded as an anticipation of the method of reflective equilibrium.

In this article, Rawls asks what is the test one should use to determine whether a judgement in a particular case is rational. His answer—identical to Goodman's—is: 'a judgment in a particular case is evidenced to be rational by showing that, given the facts and the conflicting interests of the case, the judgment is capable of being explicated by a justifiable principle (or set of principles)'. (Rawls 1951, p. 10). However, how do we know that a set of principle is justifiable? Rawls's answer is articulated in four points:

1. The principles are acceptable insofar as they constitute a successful explication of our 'moral insight' as it is expressed in considered judgements representing 'the mature convictions of competent men as they have been worked out under the most favourable existing conditions' (Rawls 1951, p. 10). In other terms: principles should be the result of the explication of the total range (or at least of a significant range) of the intuitive judgements expressed by competent moral judges. The fact that a specific set of principles is a good explication can be determined by applying (in an explicit and conscious way) the principles to the same cases competent judges expressed their considered judgements about: if the outputs of a conscious application of the principles to the cases are identical to (or are, at least, a good approximation of) competent judges' intuitive judgements, then the explication can be regarded as successful and the principles as satisfactory.
2. The principles are reasonable when their merits are weighted by criticism and open discussion, without being falsified and, eventually, implementing 'a gradual convergence of uncoerced opinion' (Rawls 1972, p. 11).
3. Besides being explicative, the principle should be fruitful: 'able to resolve moral perplexities which existed at the time of its formulation and which will exist in the future' (Rawls 1972, p. 11).
4. 'Finally, the reasonableness of a principle is tested by seeing whether it shows a capacity to hold its own (that is, to continue to be felt reasonable), against a subclass of the considered judgments of competent judges, as this fact may be evidenced by our intuitive conviction that the considered judgments are incorrect rather than the principle, when we confront them with the principle' (Rawls 1972, p. 11).

> In other terms: despite having ascertained that principles fail to explicate some judgements, competent judges decide to hold the principles as correct anyway. Why? Because the principles are so convincing that they have changed the way judges think is correct to judge.

With this fourth condition, Rawls explains in which sense the conditions of the explication are not so restrictive as to ask for a complete 'restitution' of the initial material: principles have to account for a significant part of initial judgements, not for the totality of them. Or, better, principles cannot account for the totality of initial judgements. As I will extensively argue in Sect. 3.2.1 on Carnap, there are specific reasons why the outcome of an explication should, on the one hand, account for as many judgements as possible, but cannot, on the other hand, account for them all.

But let us stick to Rawls's idea of explication a little longer. What about explication in *A Theory of Justice*?

In *A Theory of Justice* the relation between principles and judgements in reflective equilibrium is of an explicative nature as well: 'what is required is a formulation of a set of principles which, when conjoined to our beliefs and knowledge of the circumstances, would lead us to make these judgments with their supporting reasons were we to apply these principles conscientiously and intelligently. A conception of justice characterizes our moral sensibility when the everyday judgments we do make are in accordance with its principles' (Rawls 1972, p. 46). Nonetheless, unlike in 'Outline of a decision procedure for ethics' (where Rawls deals with the question only marginally), in *A Theory of Justice* the procedure at the basis of the successful explicative relation connecting judgements and principles acquires a great importance. Through the analysis of those dynamics, one comes to conceive a good explication as the result of a series of adjustments occurring both on theory and on intuitions. In this sense, the final version of judgements and principles in reflective equilibrium diverges from the original intuition core and initial theoretical options: the theory that we finally accept has been preceded by different theoretical hypotheses, first advanced and then adjusted or rejected in the light of the reflective process. Also, the set of judgements this theory coheres with differs from the pre-theoretical set. In particular, what is important for the justification of a theoretical picture is not considered judgements in general, but the considered judgements we embrace at the end of the reflective process.

3.1.3 Considered, Intuitive, and Pre-theoretical Judgements

So far I have used the terms ‘intuitions’, ‘considered judgements’, and ‘pre-theoretical convictions’ as roughly synonymous but, in the light of the foregoing considerations, some clarification is in order.

First of all, Rawls claims that verdicts composing the original set of judgements must be *considered*, that is, ‘rendered under conditions favourable to the exercise of the sense of justice, and therefore in circumstances where the more common excuses and explanations for making a mistake do not obtain’ (Rawls 1972, p. 11). One has to discard ‘judgments made with hesitation, or in which we have little confidence. Similarly, those given when we are upset or frightened, or when we stand to gain one way or the other can be left aside’ (Rawls 1972, p. 11). Hence, considered judgements are judgements of competent moral judges (that is to say rational and adult people), formulated under the best conditions possible for the decision.

According to what Rawls claims in *A Theory of Justice*, only a limited proportion of these judgements will ‘survive’ the reflective process and will be, for that reason, designated as justifying the principles. Nevertheless, speaking about considered judgements has great importance: imposing constraints on people expressing the judgements and on situations in which these judgements are formulated helps make a good pre-selection of our verdicts of ‘just’ and ‘unjust’.

The concept of considered judgement is present in ‘Outline of a decision procedure for ethics’, as well. Here Rawls defines both the class of competent judges and the conditions under which judgements should be elaborated—the so-called ‘circumstances of the judger’. In addition, Rawls provides a direct characterization of *intuitive* judgments.

> By the term ‘intuitive’ I do not mean the same as that expressed by the terms ‘impulsive’ and ‘instinctive’. An intuitive judgment may be consequent to a

> thorough inquiry into the facts of the case, and it may follow a series of reflections on the possible effects of different decisions, and even the application of a common-sense rule, e.g., promises ought to be kept. What is required is that the judgment not be determined by a systematic and conscious use of ethical principles. The reason for this restriction will be evident if one keeps in mind the aim of the present inquiry, namely, to describe a decision procedure whereby principles, by means of which we may justify specific moral decisions, may themselves be shown to be justifiable [...] It is clear that if we allowed these judgments to be determined by a conscious and systematic application of these principles, then the method is threatened with circularity. (Rawls 1951, pp. 7–8)

In sum, by presenting judgements as *intuitive*, Rawls does not refer to judgement-specific features (to the fact judgements have a particular phenomenology, a certain aetiology, and so on). Rawls is interested in describing the role they play in the justification dynamics: judgements can justify the theory as they are intuitive, that is, not determined by a systematic and conscious use of ethical principles. This is true of all judgements one appeals to in order to justify the theory. Ergo, also of those judgements which do not appear in the initial set of intuitions and which we end up embracing at the end of the process, that is, judgements that are the outcome of the decisions we have taken in the course of the reflective dynamics. In what follows, I will argue in favour of the intuitiveness of these judgements as well. However, before doing this, I shall make the distinction between pre-theoretical and post-theoretical judgements clear.

In both ‘Outline of a decision procedure for ethics’ and in *A Theory of Justice*, judgements play a controlling role: one checks principles against them. However, judgements are not indefeasible: they can be corrected in light of the theory. Hence, the conformity of the theory to our intuitions does not amount to a complete continuity with our initial considered judgements, but rather to a relation of coherence between our theory and our current judgements. These judgements have to correspond to the initial ones to an important extent, but not completely. So, verdicts justifying the theory are not considered judgements, but considered *and* reflected judgements, that is, judgements we embrace at the end of the reflective process.

Post-theoretical judgements should be intuitive as well: if we want them to justify the theory, then they should not be the result of the conscious and systematic application of the latter. But post-theoretical judgements are not just the pre-theoretical judgements that have 'survived' the reflective process: they are also those that we came to embrace during (and thanks to) the reflective dynamics.

So, one could ask: are judgements that we came to embrace during (and thanks to) the reflective process really intuitive? If so, why? In the light of the previous discussion on Goodman, I suggest that the judgements we get at the end of the reflective process are intuitive in so far as we take them to be correct independently of the theory. The theory may have contributed to their rising, but our judging the cases in the way we do—in a way that is consistent with the theory—is not the result of a conscious and systematic application of it. Suppose that we consider a new case, and that our judgment is consistent with the theory we came to embrace and inconsistent with the way in which we would have judged previously. In and of itself, this does not mean that we judged the case by applying the theory in a conscious and systematic manner. The theory could be so strong and persuasive to have caused (in us) a sort of *Gestalt switch*, prompting us to 'see' the situation in a way which agrees with it and is different from the way in which we would have seen it before.

But then, what does justify theoretical and intuitive judgements in RE?

The idea is that both in moral philosophy and in logic, the correctness of the rules depends on instances of intuitive nature: considered judgements to the effect that an action is (or isn't) just and verdicts to the effect that an inference is (or isn't) correct. In turn, intuitions (moral considered judgements and verdicts of validity) are assessed as rational for being explicated by ethical principles and inductive criteria. Is this account circular? Let us see how this question—already debated in Sect. 3.1 on Goodman—can be answered in the light of our discussion of Rawls.

Let us start from the principles. Principles of justice are confirmed when they match with our duly pruned intuitions. So, principles have the merit not only of accounting for our actual intuitions, but also of having been sufficiently strong to 'prevail' over other pre-theoretical convictions.

In addition, principles are considered correct not only because of their conformity with our intuitions, but also on the basis of the (initial) situation that originated them. This situation incorporates fair and correct procedural standards and is therefore thought to generate fair and correct results. Hence, as regards the justification of principles, the appeal is twofold: on the one hand, one can refer to the positive upshot of the judgements-test; on the other hand, one can refer to the correctness of the original position.

One could object that intuitions play a great role in the choice of the conditions of the original position as well. In fact, we know that the final version of the original position depends, largely, on the development of reflective dynamics: namely, inconsistencies between principles and judgements can be solved by modifying the description of the initial situation so that it could generate principles accounting for our intuitions. Does this mean that standards are determined *ad hoc* on the basis of the judgements we want to justify? Not really. The conditions we choose for the initial situation have a certain independence from the other elements of the procedure. Let us consider the description of the original position: it involves formal (principles of publicity), motivational (mutual disinterest), and knowledge (the veil of ignorance) constraints on contractors and principles. They are the standards we think are more reasonable to impose on the choice of principles, that is, the standards we in fact accept, or we can be persuaded to accept through philosophical considerations. The same holds for the ideal conditions in which considered judgements should be elaborated: they are standards of epistemic nature that Rawls describes as 'favourable for deliberation and judgement in general' (Rawls 1972, p. 48) and are distinct from both theoretical and intuitive beliefs that will play a role in the related process.

One of the major supporters of Rawls's reflective equilibrium, Daniels (1979a, b, 1980a, b, 1985) refers to the kind of theories I have just mentioned using the expression 'background theories and beliefs'. According to Daniels, background theories and beliefs play a central role in the method: they are the third element involved in the reflective dynamics, besides principles and intuitions.

Other examples of background beliefs and theories used by Rawls to develop his conception of justice are theories about the concept of a person, the role of morality and justice in society and general notions for the social theories that are at judges' disposal under the veil of ignorance. One could complete the list of theories that can be used to build a moral theory by adding elements such as assumptions on human rationality, biology notions (or scientific notions in general), epistemological principles, and so on.

Hence, background theories intervene in two phases of the reflective process, other than in the original position and in the process of selection of considered judgements: first, during the choice of the principles in the original position; secondly, while checking the coherence between principles and intuitions.

In sum: the justification of the principles depends on considered judgements and on the correctness of the initial situation, which is justified, in turn, by background theories and beliefs. In general, background theories play a central role in the elaboration of moral theories and provide a justification of the principles that is independent from that offered by considered judgements.

Finally, let us focus on intuitions. As I claimed in Sect. 3.1 on Goodman, the correctness of intuitions is somehow independent of the principles: a theory explaining intuitions can help them emerge and can motivate their acceptance. However, it does not really justify them. The strength of intuitions lies in the very fact of being accepted—of being the judgements that we in fact express or think is correct to express in front of the cases. So, according to this perspective, the theory does not really justify intuitions; or better, it does not justify intuitions in the same way in which the theory is justified by intuitions. Simply, it produces a framework that can be used to testify the rationality of their acceptance.

This thesis is similar to that espoused by Gutting in *What Philosophers Know* (2009). Gutting speaks of intuitions in the terms of pre-philosophical opinions (or convictions). According to Gutting, Rawls moves from pre-philosophical convictions on what is just or unjust, 'that have no need for justification by philosophical argument; that, in other words […] do not require philosophical foundations' (Gutting 2009, p. 225). The reason is that they are already grounded: namely, 'they express convictions rooted in lived experience and connected to our fundamental self-understanding' (Gutting 2009, p. 191). 'But, although convictions do not require philosophical justification, they do require philosophical maintenance' (Gutting 2009, 225): they require to pass through the conservative and, at the same time, corrective and enhancing process of reflective equilibrium.

> We are intellectual creatures and cannot avoid thinking about our convictions—about what they mean, how we can defend them against challenges, and so on. To adapt Levi-Strauss' well-known terminology: we need practices (and so convictions) that are *good-for-living*. But as intelligent humans, what we find good-for-living must also be *good-for-thinking*, and the continual probing and refining of our convictions through philosophical reflection is the way we ensure that they remain good for thinking. (Gutting 2009, p. 225)

Hence, according to Gutting, reflecting (philosophically) on our convictions is an 'inevitable aspect of developed human life': as thinking beings, we need to reflect on our convictions, to give them a systematic and consistent exposition, to make their consequences explicit, to defend them from objections and, if necessary, to adjust or even reject them.

3.2 An Anticipation of the Method: Carnap's Explication

In the previous section we saw that the method of reflective equilibrium appeals to explications. But what exactly is an explication? And what is its relation to reflective equilibrium?

Rawls ascribes to Goodman the invention of the approach he dubs 'reflective equilibrium'. However, it is possible that the methodological theses stated in 'The new riddle of induction' (1953), in *A Theory of Justice* (1971), and in 'Outline of a decision procedure for Ethics' (1951) draw on *Logical Foundations of Probability* (1950), a text in which Carnap introduced the explication method. Rawls's methodological debt to Carnap emerges clearly in 'Outline of a decision procedure for Ethics' where the building and justification of moral theory is described in terms of explication. In addition, Carnap's influence on Rawls's and Goodman's ideas is shown by important similarities between the method of explication and that of reflective equilibrium: both are strategies for theory elaboration, both move from a Socratic problem (questions in the form 'What is X?') and both establish a precise, though not absolute, constraint on theory construction.

3.2.1 Fish, *Piscis*, *Piscis**

Carnap's goal in *Logical Foundations of Probability* is to answer the question 'What is probability?' in scientifically rigorous terms. At the time Carnap engaged with the issue, the probability calculus and the concept of probability were largely in use, both among ordinary people and scientists; yet, a theory defining in an exact and unambiguous manner the nature of probability was missing. While Carnap does not face this problem directly, in the prefatory chapter he illustrates the procedure by means of which the application of the concept of probability (or the use of the term 'probability') can be made rigorous. This procedure, which applies in general to the conditions of application of any concept and to the conditions of use of any word, is the method of explication.

The task of explication consists in transforming an ambiguous or imprecise concept (or term), which is expressed in everyday language or in a previous stage in the development of scientific language, into a precise concept (or term), which is expressed in 'a well-constructed system of scientific either logico-mathematical or empirical concepts' (Carnap 1950, p. 4).

Before describing what explication is, Carnap underlines that the explication should not be confused with the explanation. The explanation makes the explication possible, but does not coincide with it. Carnap claims that explanation is part of the process of the formulation of the problem, 'not yet to the construction of an answer' (Carnap 1950, p. 5). By means of the explanation, one gives examples of the uses one deems as exemplary and examples of the uses that one does not wish to take into account. Sometimes the explanation may involve spelling out information of a more general character. Let us consider, for instance, the question 'What is truth?' In giving an explanation of the concept of truth, one would say that one is 'looking for an explication of the term "true", not as used in phrases like "a true democracy", "a true friend", etc., but as used in everyday life, in legal proceedings, in logic, and in science, in about the sense of "correct", "accurate", "veridical", "not false", "neither error or lie", as applied to statements, assertions, reports, stories, etc.' (Carnap 1950, p. 6). Hence, through the explanation one neither defines the term 'true', nor gives a theory of truth; rather, one just clarifies what the object of theorization will be. On the contrary, 'to explicate' means to give an answer to the question 'What is truth?'. An explication of 'truth' is, for instance, the definition of 'true' given by Tarski.

Let us now consider the description of explication. Carnap writes: 'if a concept is given as *explicandum*, the task consists in finding another concept as its *explicatum* which fulfils the following requirements to a sufficient degree: similarity, exactness, fruitfulness, and simplicity' (Carnap 1950, p. 8). A brief discussion of the requirements of similarity and fruitfulness will help to understand what explication is.

Let us start with similarity. First of all, 'the explicatum is to be similar to the explicandum in such a way that, in most cases in which the explicandum has so far been used, the explicatum can be used' (Carnap 1950, p. 8).

To explain what kind of similarity is required between the explicandum and the explicatum, Carnap introduces the following example, where the pre-scientific concept *fish* and the scientific concept *Piscis* are, respectively, explicandum and explicatum. The result of the explication of *fish* (that is, the new concept *Piscis*) is still applicable to a large part of the objects which were subsumed under the concept *fish*, but it does not even approximately coincide with the pre-scientific concept *fish*: while the term 'fish' generically means 'animal living in water', the term '*Piscis*', still applied to animals living in water, refers to some of those animals which show distinctive properties. Carnap describes this transformation as a change in the rules of language. Furthermore, he claims that this change is motivated by factual discoveries: first, zoologists find that aquatic animals presenting certain peculiarities (cold-blooded vertebrates, with gills throughout life) have many more features in common than the aquatic animals in general; therefore, they replace the inexact and generic concept *fish* with the exact and more fruitful concept *Piscis*. As a consequence of this change, some objects which were subsumed under the concept *fish* are no longer subsumed under the concept *Piscis*: for example, whales and seals have the property of being animals living in water, which the concept *fish* designates, but they do not fall under the concept *Piscis*.

Hence, a close similarity between explicatum and explicandum is not required. The new concept only has to apply to a *significant* part of the objects the old concept applied to. What are the reasons for the continuity and the discontinuity between the explicatum and explicandum? Let us begin from the continuity. The reason why the explicated term has to include a large, or at least a significant part, of the uses of the initial term (the 'central' uses of the explicandum) is easy to guess. Let us imagine that the new term does not apply to a great part of the old uses—in particular, to the uses we believe are more significant. What would we say of this result of the explication? Plausibly, we would say that it is not the explication of *that* term (is not the explicatum of that explicandum). Hence, if we want our explication to be satisfactory, that is, to be considered an adequate answer to the initial question, then we cannot avoid outlining rules for the use of the explicatum that account for a substantial part of the uses the rules for the use of the explicandum previously accounted for.

What about discontinuity? The reason why the explicatum cannot (or better, should not) account for the totality of the uses the explicandum accounted for is due to the constitutive imprecision of the explicandum: as the new concept has to be precise, it cannot reproduce the imprecision of the old one. If an explicatum reflected the imprecision of its explicandum we would say that the alleged explicatum is not really an explicatum after all.

This last argument help us introduce the second of the four requirements for a good explication, that is, exactness: 'The characterization of the explicatum, that is, the rules of its use (for instance, in the form of a definition), is to be given in an exact form, so as to introduce the explicatum into a well-connected system of scientific concepts' (Carnap 1950, p. 8).

Thirdly, 'the explicatum is to be a fruitful concept, that is, used for the formulation of many universal statements (empirical laws in the case of a nomological concept, logical theorems in the case of a logical concept)' (Carnap 1950, p. 8).

With the introduction of the constraint of fruitfulness, it becomes clear why the procedure of explication goes much further than mere conventional stipulation. Let us go back to our example. It has been said that the new concept *Piscis* is narrower than that of *fish*. One could ask whether the agents who performed the explication could have proceeded in a different way. For example, 'instead of the concept Piscis they could have chosen another concept—let us use for it the term "piscis*"—which would likewise be exactly defined but which would be much more similar to the prescientific concept Fish by not excluding whales, seals, etc.' (Carnap 1950, p. 7). To this doubt, Carnap answers that this construction would be of little use for scientific purposes: the concept *Piscis* is preferable to any other concept more similar to fish (for instance, *Piscis**), since it can be used for the formulation of general laws. Fruitfulness justifies discontinuity of the explicatum with the original domain. In other words, the criterion of fruitfulness is more important than that of similarity.

3.2.2 Answering the Question 'What Is X?' Comparing Explication with Reflective Equilibrium

We have seen that, in his discussion of explication, Carnap highlights the aspects of continuity and discontinuity that characterize the relation that the new scientific concept maintains with the pre-scientific one: the new concept has to account for ordinary use of the pre-scientific concept, but it can also (or better, it should) diverge from it. Metaphorically, the explicated concept has to 'capture' the way we use a term, not 'photograph' it.

The underlying structure of the explication procedure described by Carnap is similar to what Goodman suggests about the legitimation of induction principles, and what Rawls suggests about the construction and justification of moral principles. Just like the relation between the explicatum and the explicandum, the relation which is determined between the theory of valid reasoning and the accepted inferential practice (singular intuitions of validity about particular inferences), and the one between the conception of justice (principles of justice and original position) and considered judgements, can be described in terms of continuity and discontinuity: the theory must account for the way we think it is correct to reason or to judge, though it can set aside some of our initial intuitions.

In detail, when comparing the explication and the reflective equilibrium methods, three important analogies can be found.

(1) Despite the different objects they analyse, Carnap, Goodman, and Rawls share a similar purpose: the construction of a theory starting from a Socratic problem, that is to say, a question of the form 'What is X?'. Carnap explicitly moves from the question 'What is probability?', Rawls from the question 'What is justice?', and Goodman's investigation pre-supposes the question 'What is valid reasoning?'.

(2) The theory-building method involves aspects that precede and govern the elaboration of the theory and that work as criteria for its acceptance. Rawls's and Goodman's method is based on intuitions: the theory of inductive and deductive reasoning is built on and justified by judgements of validity about single inferences; the theory of justice is built on judgements of 'just' and 'unjust' about morally salient situations. Carnap's method is based on language use.

Hence, intuitions and uses of the relevant terms guide the theorizing. Furthermore, intuitions and uses work as constraints for theories that are already formed. Namely, when taking into account a theory or a result of an explication, one has to check whether the starting point has been respected or not. In particular, as an explicatum—in so far as it can be considered a good explicatum—has to account, to a significant extent, for the use of the explicandum, a theory—in so far as it can be considered a good theory—has to agree, to a significant extent, with our intuitions. Otherwise, the explicatum is not the explicatum of that explicandum and the theory is not the theory of what we wanted to make a theory: it is not a satisfying answer to the initial question.

(3) Yet, neither the constraint of intuitions, nor the constraint of use is absolute: the theory resulting from the reflective procedure and the explicatum can be partially disjointed from the original set of intuitions and from the original set of ordinary uses. These differences are usually due to the systematizations a theory requires: the initial sets are often inaccurate and incoherent; consequently, a scientific theory cannot

respect them entirely. Moreover, differences can be due to some intuitions which arise during the process of theory-building and prevail on some pre-theoretical intuitions. As already explained, it can be that the theory we have produced agrees with a considerable part of our intuitions, but involves some unexpected consequences which disagree with previous pre-theoretical intuitions. If we find the theory persuasive, we can decide to hold on to it and to abandon the set of intuitions that disagree with it. In such a case, judgements entailed by the theory prevail over the previous intuitions.

Clearly, in the perspective of RE, the opposite could occur as well: we could decide to modify the theory in favour of intuitions. The RE method establishes that the original set of intuitions has to be revised when incoherencies with the theory we accept occur: intuitions disagreeing with the consequences of the theory are removed and the set of intuitions is integrated with the theory's unexpected consequences, so that the mutual support between theory and intuitive judgements is achieved. Hence, in Rawls's and Goodman's method, a loss is always evened-up by a gain: the method is productive and enhancing, not merely corrective. On the contrary, in the case of the explication, a divergence between the application of the old and that of the new concept does not compel us to make any kind of adjustment: the theory catches from the pre-scientific use what is significant to catch; uses which are not significant are simply abandoned.

Hence, despite the similarities, there is also an important difference between the methods. Indeed, they differ in how one should handle the aspects constraining the theory—uses and intuitions—when, in the case of an incongruence between them and theory, one decides to hold the theory. Whereas in the case of reflective equilibrium pre-theoretical judgements must be adjusted (one has to remove from the set of intuitions the pre-theoretical judgements disagreeing with the theory and replace them with the new, unexpected consequences of the theory), in the case of explication coming back to ordinary use, and modifying it, is not necessary.

The reason for this disanalogy is ascribable to the different goals of the two methods. Whereas explication aims to create a scientific concept of X; reflective equilibrium aims to answer the question 'What is X?' without restrictions. In the first case, when an answer is given, the answer is normative only with respect to the scientific use of 'X': it is a convention relative to a specific domain. As concern reflective equilibrium, the ordinary use should conform to it as the answer is (at least, aims to be) normative with respect to the use of 'X'. This is also the reason why the agent engaging with explication moves from the ordinary use but does not aim to replace or to modify it: the goal is to come up with an exact and fruitful concept which is exploitable for scientific goals; not to analyse and reform an ordinary concept.

As Nagel writes: 'we may use the scientific meaning of the term in contexts where precision is required while keeping the original explicandum for everyday use; for example, we may agree to use the term "or" only in a non-exclusive sense in our logic texts while leaving it ambiguous in ordinary language' (Nagel 2007, p. 795). In general, in explication there is not the 'back-and-forth' dynamics that is typical of the reflective method.

3.3 Clarifications of the Method: Lewis's Reflective Equilibrium

3.3.1 Credibility and Systematicity

In this last section on RE, I will handle the version of the method proposed by Lewis in *Counterfactuals* (1973), in *On the Plurality of Worlds* (1986), and in the introduction to the *Philosophical Papers* (1983a). Lewis's discussion on the method is interesting as it helps to better explain a series of aspects: (1) the relation between two important features a philosophical theory should have in order to be a good theory—consistency with our intuitions and systematicity; (2) the problem of respecting the common sense; and (3) the maxim of honesty.

In *Counterfactuals* (Lewis 1973) Lewis introduces the thesis of modal realism: the idea according to which possible worlds are real entities in the exact same way the world we happen to inhabit is.

> I believe there are possible worlds other than the one we happen to inhabit. If an argument is wanted, it is this. It is uncontroversially true that things might have been otherwise than they are. [...] Ordinary language permits

> the paraphrase: there are many ways things could have been besides the way that they actually are. On the face of it, this sentence is an existential quantification. It says that there exist many entities of a certain description, to wit, 'ways things could have been'. I believe things could have been different in countless ways. I believe permissible paraphrases of what I believe; taking the paraphrase at its face value, I therefore believe in the existence of entities which might be called 'ways things could have been'. I prefer to call them 'possible worlds'. (Lewis 1973 p. 84)

Afterwards, Lewis gives an overview of alternative conceptions to modal realism and rejects them. In brief, his argument in favour of modal realism could be sketched as follows: the fact that 'things might have been otherwise than they are', and in particular that they 'could have been different in countless ways', is one of our common convictions; modal realism is a satisfactory exposition of this conviction; moreover, unlike non-realist conceptions, modal realism does not have problematic consequences; therefore, modal realism is the expression of the correct view on the nature of possible worlds.

As already mentioned, Lewis outlines and highlights the limits of alternative conceptions. I will not delve here into the arguments Lewis advances for proving this point. Neither, I will assess the plausibility of modal realism itself. Rather, the question I am going to deal with is: why should the fact that we express certain convictions concerning what things are, or could have been, be decisive in order to ascertain the correctness of a specific theoretical proposal on the nature of possible worlds? Or, more generally, why should our convictions (what we say or think about any X)be relevant in order to establish the correctness of a theory on X? My concern is a methodological one.

In *Counterfactuals*, Lewis advances some considerations in response to this question:

> One comes to philosophy already endowed with a stock of opinions. It is not the business of philosophy either to undermine or to justify these preexisting opinions, to any great extent, but only to try to discover ways of expanding them into an orderly system. It succeeds to the extent that (1) it is systematic, and (2) it respects those of our pre-philosophical opinions to which we are firmly attached. In so far as it does both better than any alternative we have

> thought of, we give it credence. There is some give-and-take, but not too much […] So it is throughout metaphysics; and so it is with my doctrine of realism about possible worlds […] Realism about possible worlds is an attempt, the only successful attempt I know of, to systematize preexisting modal opinions. (Lewis 1973, p. 88)

Our pre-philosophical ('preexisting') opinions—Lewis continues—not only favour realism about possible worlds, but also inform us about which version of realism we should prefer.

In sum: the metaphysician's work consists in looking for the theoretical framework that better systematizes our pre-theoretical opinions. In particular, the process leading from the initial question—What is X?—to a well-established theory on X is conceived in terms of a procedure, in which our pre-existing convictions are balanced with our theoretical ones. Pre-theoretical opinions play a central role not only in the constructing phase, but also while evaluating a proposed theory: each theoretical hypothesis is assessed on the basis of its coherence with intuitions (condition 2), as well as on the basis of its systematicity (condition 1).

In 'The incredulous stare' (*On the Plurality of Worlds* 1986), Lewis seems to be less confident about the intuitiveness of modal realism. Since Lewis's defences of modal realism have been typically met by incredulous stares, we may think that the very thesis he stood by was largely counter-intuitive. If so, he needed to make explicit the consistency between his theory and the alleged pre-theoretical intuitions. So, Lewis's first step is to acknowledge the striking counterintuitiveness of his proposal. The second step is to try to explicitly state which is the (alleged) counterintuitive consequence of modal realism.

> Modal realism *does* disagree, to an extreme extent, with firm common sense opinion about what there is. (Or, in the case of some among the incredulous, it disagrees rather with firmly held agnosticism about what there is). When modal realism tells you—as it does—that there are uncountable infinite donkeys and protons and puddles and stars, and of planets very like Earth, and of cities very like Melbourne, and of people very like yourself, … small wonder if you are reluctant to believe it. And if entry into philosophers' paradise requires that you believe it, small wonder if you find the price too high. (Lewis 1986, p. 133)

The third (and crucial) step is to ask: what do non-philosophers mean when they say that they cannot believe that 'uncountable infinite donkeys and protons and puddles and stars, and of planets very like Earth, and of cities very like Melbourne, and of people very like yourself' exist?

Lewis distinguishes two opinions: the first one he agrees with, the second one he does not agree with. The point, Lewis argues, is that it is not clear at all whether it is the first, or the second, opinion that the layman has when he or she expresses his or her incredulity. In particular, it seems that non-philosophers are not able to say whether, by claiming that they cannot believe uncountable infinite donkeys exist, they mean 'that there do not actually exist an uncountable infinity of donkeys' (Lewis 1986, p. 133) [first opinion], rather than 'that there do not exist an uncountable infinity of donkeys—with the quantifier wide open, entirely unrestricted, and no "actually" either explicit or tacit in the sentence' (Lewis 1986, p. 133) [second opinion]. Hence, if this is true, it is then difficult to conclude that the thesis of modal realism effectively contradicts a common-sense intuition: which thought is contradicted? The first? The second? Both?

In general, philosophers make distinctions where laypeople do not; that is the reason why their opinions and theses seem to diverge from common sense. They appear to diverge, but it is not obvious they do. Philosophers, concludes Lewis, are entitled to ignore judgements like 'I cannot believe uncountable infinite donkeys exist'.

Lewis asks if ignoring layman's judgements might be a trouble for those who believe that common sense should be respected. This is his answer: common sense, Lewis explains, has a certain but not absolute authority in philosophy. As regards the centrality (authority) of common sense, he highlights two aspects: on the one hand, our pre-theoretical convictions are the only possible starting points of our inquiry (uniqueness of the starting point); on the other hand, intuitions are important for the acceptance of the theory: namely, a theory cannot be credible if patently contradicts our pre-theoretical opinions (credibility). However, the plausibility of a (real) theory is also given by its systematicity and simplicity. So, for a gain on the side of systematicity, opinions must, at times, be abandoned. Apropos, Lewis introduces a less binding version of the maxim 'respect common sense': no matter if a philosophical theory accounts for anything the layman thinks or says; what really matters is that the theory does not contradict what the philosopher (who embraces it)thinks or says in his or her 'least philosophical and most commonsensical moments' (Lewis 1986, p. 135). This is, Lewis specifies, a simple maxim of honesty.

In conclusion: the agreement of the theory with intuitions does not so much consist in the agreement between the philosopher's theses and the layman's convictions, as in the agreement between the theses a person, as a philosopher, proposes and the convictions this same person has, regardless of his or her engagement in the philosophical field.

In the introduction to the *Philosophical Papers*, Lewis largely reaffirms what has been said up to here. The metaphysician's work consists in looking for the theoretical picture which best systematizes our pre-theoretical opinions. The procedure leading from the initial question—what is X?—to a 'mature' theory on X is described as balancing our pre-theoretical and theoretical convictions. In this place, Lewis speaks of opinions: both intuitions and theories are described in terms of opinions. The idea seems to be the following: given that intuitions are opinions, theories (that are based on intuitions) cannot be other than opinions in turn. Certainly, theories are much more articulated and present a higher degree of generality than intuitions; however, their nature must be similar to that of the data composing the basis they rest on.

At every stage of the procedure of bringing (intuitive and theoretical) opinions into equilibrium the metaphysician measures the price: the cost of a loss, in face of a gain. In no (or almost no) case, is it a matter of finding *the* argument (the crucial, the decisive argument) in favour of, or against, a certain theoretical option. Speaking of the theories exposed in the essays collected in the volume, Lewis writes:

> The reader in search of knock-down arguments in favour of my theories will go away disappointed. […] Philosophical theories are never refuted conclusively. (Or hardly ever. Gödel and Gettier may have done it.) The theory survives its refutation—at a price. Perhaps that is something we can settle more or less conclusively. But when all is said and done, and all tricky arguments and distinctions and counterexamples have been discovered, presumably we will still face the question which prices are worth paying,

> which theories are on balance credible, which are the unacceptably counterintuitive consequences and which are the acceptably counterintuitive ones. On this question we may still differ. (Lewis 1983a, p. x)

Lewis concludes: 'once the menu of well-worked-out theories is before us, philosophy is a matter of opinion' (Lewis 1983a, p. xi).

3.3.2 'What Good Are Counterexamples?'

In the last part of this chapter, I will analyse an article by Weatherson, 'What good are counterexamples?' (2003). The author moves from the methodological theses described by Lewis and analyses the consequences of the application of Lewis's conception to a specific problem: the problem of whether it is legitimate to reject the JTB theory of knowledge on the basis of Gettier's intuition. Contrary to the opinion of the large majority of the epistemologists (and of Lewis himself), Weatherson claims we should reject Gettier's intuition and keep the JTB theory of knowledge.

Weatherson's article offers an example of how we can understand Lewis's view on the compromise between a theory's credibility and its systematicity. Furthermore, it will serve as starting-point for our discussion of one of the main criticisms of the appeal to intuitions in philosophy: the objection that using intuitions as evidence is incompatible with the aims and results of philosophers' inquiries.

First, Weatherson asks: how do we assess the correctness of a theory? In particular, how do we come to establish that a certain theory of knowledge is true? One knows if a theory is correct by evaluating whether this theory meets the following conditions (a) and (b).

> The true theory of knowledge is the one that does best at (a) accounting for as many as possible of our intuitions about knowledge while (b) remaining systematic. A 'theory' that simply lists our intuitions is no theory at all, so condition (b) is vital. And it is condition (b), when fully expressed, that will do most of the work in justifying the preservation of JTB theory in face of the counterexamples. (Weatherson 2003, p. 7)

According to Weatherson, the idea of the priority of (b) over (a) is common to different disciplines. Let us take, for instance, logic:

> If we just listed intuitions about entailment, we could have a theory on which disjunctive syllogism (*A* and *~A* ˅ *B* entail *B*) is valid, while *ex falso quadlibet* (*A* and *~A* entail *B*) is not. Such a theory is unsystematic because no concept of entailment that satisfies these two intuitions will satisfy a generalized transitivity requirement: that if *C* and *D* entail *E*, and *F* entails *D* then *C* and *F* entail *E*. (This last step assumes that *~A* entails *~A* ˅ *B*, but that is rarely denied). Now one can claim that a theory of entailment that gives up this kind of transitivity can still be systematic enough, and Neil Tennant (1992) does exactly this, but it is clear that we have a serious cost of the theory here, and many people think avoiding this cost is more important than preserving all intuitions. (Weatherson 2003, pp. 7–8)

Weatherson goes on specifying conditions (a) and (b) in two further sub-conditions each: the agreement with a significant part of our intuitions (test 1), a limited presence of counterintuitive consequences (test 2),theoretical importance (test 3) and simplicity (test 4). Since (b) has a priority over (a), in the following paragraphs, I shall focus on the two sub-conditions that specify it: theoretical relevance (test 3) and simplicity (test 4). First, I will explain the role that, according to Lewis, these two sub-conditions play in the process of determining the meaning of terms. Secondly, I will present the consequences of this view for the problem Weatherson is interested in.

First, it is interesting to notice that Weatherson's references are neither *Counterfactuals*, nor the introduction to the *Philosophical Papers*—texts in which Lewis presents his methodological theses, but rather 'New work for a theory of Universals' (1983b) and 'Putnam's Paradox' (1984)—articles where Lewis introduces the notion of naturalness and explains its role in the process of determining the meaning of terms. This is not accidental. In fact, Weatherson conceives the process of construction of a theory on X as the process of determining the meaning of the term 'X'. Let us take the case under inquiry—the problem of defining what knowledge is and establishing whether a proposed theory of knowledge is correct or not: what does the epistemologist do when he or she asks what knowledge is? And how does he or she assess the correctness of a certain theoretical hypothesis about its nature? According to Weatherson, the epistemologist asks what the meaning of the term 'knowledge' is, that is to say, which is the property that, once its uses have been considered, turns out to be the most natural. Apropos, in 'New work for a theory of Universals', Lewis writes:

> Reference consists in part of what we do in language or thought when we refer, but in part it consists in eligibility of the referent. And this eligibility to be referred to is a matter of natural properties. (Lewis 1983b, p. 371)

And here is how Weatherson presents Lewis's thesis:

> The meaning of a predicate is a property in the sense described by Lewis (1983b): a set, or class, or plurality of possibilia [...] The interesting question is determining which property it is. In assigning a property to a predicate, there are two criteria we would like to follow. The first is that it validates as many as possible of our pre-theoretic beliefs. The second is that it is, in some sense, simple and theoretically important. [...] Lewis canvasses the idea that there is a primitive 'naturalness' of properties which measures simplicity and theoretical significance, and I will adopt this idea. (Weatherson 2003, p. 11)

So, in view of the determination of the meaning of a term 'X', Lewis identifies two criteria: (1) the agreement with our intuitions on X (or our uses of 'X'); and (2) naturalness. According to some interpretations of his theory (Weatherson 2013; Brown 2012), naturalness—which is primitive—can be identified on the basis of two characteristics: theoretical significance and simplicity, that is to say, the two conditions that, according to Weatherson, specify (b). So, let us suppose that a certain property, picked out on the basis of the uses of the term 'X', turns out to be theoretically useful and simple. If so, we are allowed to conclude that that property is the most natural one, and therefore the meaning of 'X'.

I will return to naturalness later. Let us now see the consequences of this conception—Lewis's methodological theses plus the idea of naturalness—to the question Weatherson is interested in: should we consider the JTB theory of knowledge refuted by GC or could we 'still keep an open mind to the question of whether it is true' (Weatherson 2003, p. 10)? This is the way Weatherson argues.

> My main claim is that even once we have accepted that the JTB theory seems to say the wrong thing about Gettier cases, we should still keep an open mind to the question of whether it is true. The right theory of knowledge, the one that attributes the correct meaning to the word 'knows', will do best on balance at these four tests. Granted that the JTB theory does badly on test one, it seems to do better than its rivals on tests two, three and four, and this may be enough to make it correct. (Weatherson 2003, p. 10)

The idea emerging from this passage and broadening in the rest of the article is this: if it is true that condition three and four count, then one cannot take for granted that the JTB theory of knowledge should be modified or abandoned in the light of Gettier's intuition. On the contrary, the theory *could* prevail on the intuition. However, given that condition three and four have priority, the theory *should* prevail over the intuition. Hence, in the passage we have just quoted, Weatherson seems to be open to both possibilities: (1) rejecting Gettier's intuition in the name of theory relevance and systematicity; (2) adjusting or abandoning the classical theory in the name of the consistency with our intuitions. However, as the argument proceeds, he is much more inclined to embrace (1): as the JTB theory analyses knowledge 'in terms of a short list of simple and significant features' (Weatherson 2003, p. 11), and accounts for many others of our intuitions, then the classical theory of knowledge is the correct theory about knowledge. Gettier's intuition can be set aside.

In order to further support (1), Weatherson also deals with the question of 'rival' theories: as far as simplicity and relevance are concerned, no alternative theory can compete with the JTB theory of knowledge. Moreover, each alternative is, in turn, subject to further counterexamples.

Weatherson makes this point with some emphasis. However, this is not decisive for the concerns of his discussion: even if it is true that the JTB theory of knowledge does better than all alternative theories with respect to the compromise between the two factors (or among the four tests), the problem is not deciding which among the alternative theories proposed is the best, but whether the best among those alternatives can be legitimately adopted in the face of a contrasting intuition.

Weatherson's idea is that the JTB theory should be considered correct in spite of Gettier's intuitions. This is the motivation:

> Let's say I have convinced you that it would be better to use 'knows' in such a way that we all now assent to 'She knows' whenever the subject of that pronoun truly, justifiably, believes. You may have been convinced that only by doing this will our term pick out a natural relation, and there is evident utility in having our words pick out relations that carve nature at something like its joints [...] I have implicitly claimed above that if you concede this you should agree that I will have thereby corrected a *mistake* in your usage. (Weatherson 2003, p. 10)

Naturalness is the key: namely, why should a theory describing the appropriate natural property (or relation) that the term 'X' denotes be challenged by a judgement that is the mere description of what we would say / be inclined to say in a specific situation?

Weatherson insists on the 'evident utility in having our words pick out relations that carve nature at something like its joints'. The idea of naturalness looks promising, indeed. It is true that, in order to explain what X is, philosophers start from what we say or think that X is. It is also true that, in order to assess whether a specific theoretical hypothesis on X is acceptable, philosophers have to ascertain that this hypothesis coheres with a significant part of our intuitions. However, in a perspective in which our research on X is an inquiry on what X is for real (and not just on what we think or say that X is) there are further criteria (simplicity and theoretical importance) that constrain the correctness of a theory with respect not only to what we say or think, but to how things stand in the world.

Weatherson's conclusions, however, are questionable. It is obvious that a philosophical theory cannot be equated to a chaotic set of intuitions: beliefs constituting a theory are consistent with each other, well-articulated; by contrast, our pre-theoretic beliefs are inaccurate and incoherent with each other. Therefore, it is true that philosophical theories appear to be extremely distant from a layman's beliefs (and also from the beliefs the philosopher had before starting his or her inquiry). However, the mere fact that theoretical beliefs on a matter are well-articulated and consistent with each other does not seem sufficient to make these beliefs substantially different from the beliefs they are grounded on. After all, in order to be *a theory*, any theory should be systematic (no matter the nature of its object); and in order to be a *good* theory, any theory would better be simple and fruitful.

Secondly, as the *real* story of the Gettier case and the JTB theory of knowledge shows, Gettier cases counted as counterexamples. Epistemologists thought that the theory was refuted by Gettier's intuition. Apparently, they didn't think that theoretical significance and simplicity were reasons enough to reject the intuition and keep the theory. Let us think about RE: the method allows us to reject the intuition in favour of the theory. On which occasions? In all cases when a theory is strong enough to 'cause' in people embracing it a sort of *Gestalt switch*—to make them 'see' (judge) a specific situation accordingly to the theory, and differently from the way in which they would have judged the situation before embracing it. One could point out that this is not even the case with Weatherson: namely, he criticises the decision to revise the JTB theory of knowledge and nevertheless recognizes the force of Gettier's intuition, that is, he embraces it. According to the supporters of reflective equilibrium, Weatherson's position is untenable. Also Lewis claims something along this line when he states the 'maxim of honesty' and when he writes:

> If our official theories disagree with what we cannot help thinking outside the philosophy room, then no real equilibrium has been reached. Unless we are doubleplusgood doublethinkers, it will not last. And it should not last, for it is safe to say that in such a case we will believe a great deal that is false. (Lewis 1983a, p. x)

Last, if one acknowledges, as Weatherson does, that intuitions are the basis we build our theory on and we test theoretical hypotheses against, then it is hard to explain why a specific theoretical hypothesis—just because it's simple, fruitful and at an advanced stage of the reflective process—shouldn't be put into discussion when its inconsistency with an intuition was shown. In order to prove that the theory is 'immune' from the attack of intuitions, one should demonstrate that that theory is correct regardless of intuitions. It is hard to imagine how one could in the case at hand.

In general this last argument could be used against any philosopher aiming to answer Socratic questions who replies to other philosophers showing a mismatch between his or her theory and an intuition with: 'it's just an intuition!', or 'I am not committed to look for reasons for rejecting that specific judgement because that judgement is a mere intuition (a judgement on what we think or say that X is), whereas the theory is the correct theory on X (a theory that tells what X really is)'.

A philosopher claiming that is not only refusing to take the disagreeing intuition seriously, but seems to be asking the philosopher who is presenting the counterexample to produce an argument in favour of the fact that a simple intuition could falsify a theory: an argument proving the robustness of that judgement, or, in general, of the epistemic status of intuitions.

According to the perspective I defend, the philosopher exhibiting the counterexample is not committed to do anything like that. Rather, it is the person refusing to consider the counterexample who has to show us that his theoretical picture is correct irrespective of intuitions. It is unlikely that the last one will find a way out. By trying to demonstrate that his or her theory on X is correct, one could not do anything other than appeal to the fact that his or her theory agrees with many other intuitions on X and to the fact that it presents the characteristics that we all believe a good theory should present. Hence, unless one tries some sort of *ad hoc* move, he or she will be obliged to face the unwelcome intuition and to reopen the negotiation that has led to the theory he or she embraces.

In this section, I argued against Weatherson and in favour of the standard praxis. However, as I anticipated at the beginning of the section, Weatherson's view is interesting also because it serves to introduce one of the most important criticisms of the appeal to intuitions in philosophy: the objection that casts doubt on the possibility for judgements of *intuitive* nature to support or attack philosophical theories.

In the next chapter ('Introduction to Part II and Part III'), I will explain why Weatherson's thesis on the priority of the JTB theory over Gettier's intuition can be considered an example of this general objection to the use of intuitions as evidence in philosophy.

References

Brown, J 2012. Words, concepts, and epistemology. In *Knowledge Ascriptions*, eds. J. Brown and M. Gerken. Oxford: Oxford University Press.

Carnap, R. 1950. *Logical Foundations of Probability*. Chicago: University of Chicago Press.

Daniels, N. 1979a. Moral Theory and the Plasticity of the Person. *The Monist* 62(3): 265–287.

Daniels, N 1979b. Wide Reflective Equilibrium and Theory Acceptance in Ethics. *The Journal of Philosophy* LXXVI(5): 256–282.

———1980a. Reflective Equilibrium and Archimedean Points. *Canadian Journal of Philosophy* X(1): 83–103.

———1980b. On Some Methods of Ethics and Linguistics. *Philosophical Studies* 37: 21–36.

———1985. Two Approaches to Theory Acceptance in Ethics. In *Morality, Reason and Truth: New Essays on the Foundations of Ethics*, eds. David Copp and Zimmerman David. Totowa, NJ: Rowman and Littlefield.

Goodman, N. 1955. *Fact, Fiction, and Forecast*. MA: Harvard University Press.

Gutting, G. 2009. *What Philosophers Know: Case Studies in Recent Analytic Philosophy*. Cambridge: Cambridge University Press.

Lewis, D. 1973. *Counterfactuals*. Oxford/Cambridge: Blackwell Publishers/ Harvard University Press (reprinted with revisions, 1986).

Lewis, D 1983a. *Philosophical Papers*, vol I. Oxford: Oxford University Press.

———1983b. New work for a theory of universals. *Australasian Journal of Philosophy* 61(4): 343–377.

———1984. Putnam's paradox. *Australasian Journal of Philosophy* 62(3): 221–236.

———1986. *On the plurality of worlds*. Oxford: Blackwell Publishers.

Nagel, J. 2007. Epistemic Intuitions. *Philosophy Compass* 2(6): 792–819.

Rawls, J. 1951. Outline of a Decision Procedure for Ethics. *Philosophical Review*, 60(2): 177–97; reprinted in Rawls, J. 1999. *Collected Papers*, 1–19. Cambridge, MA: Harvard University Press.

Rawls, J 1971. *A Theory of Justice*. Cambridge, MA: Harvard University Press; edition used: Rawls, J. 1972. *A Theory of Justice*. Oxford: Clarendon Press.

Weatherson, B. 2003. What good are counterexamples? *Philosophical Studies* 115: 1–31.

Part II

What Is Philosophy?

4

Introduction to Part II and Part III

4.1 The Legitimacy of the Appeal to Intuitions: Two Big Questions

In the last section, I analysed 'What good are counterexamples?' In that article Weatherson criticizes the decision to revise the JTB theory of knowledge in the face of Gettier's counterexamples: as opposed to the opinion of the majority of epistemologists, Weatherson thinks one should reject the intuition and keep the theory. This choice is due to the role Weatherson assigns to characteristics such as simplicity and theoretical significance: a theory that is built from our uses of 'X', and that is simple and theoretically useful, is a theory describing the appropriate natural property (or relation) that the term 'X' denotes. The JTB theory of knowledge is a simple and theoretically significant theory, or better yet, it is the only simple and theoretically significant theory of knowledge we have. So, the question Weatherson raises is the following: why should a theory describing a natural property (or relation) be questioned by a judgement that is the mere description of what we would say/be inclined to say in a single specific situation?

© The Author(s) 2016

S.M. Nicoli, *The Role of Intuitions in Philosophical Methodology*,
DOI 10.1057/978-1-137-56715-4_4

It is not difficult to see that this question goes much further than the issue of what impact Gettier cases should have on the JTB theory of knowledge. First, the issue raised appears to question the legitimacy of the practice of attacking simple and fruitful theories on the basis of imaginary or real cases. The problem is: is it correct to reconsider the validity of a well-established theory in the light of an intuition elicited by a case? Would it not be better to dismiss the intuition and keep the theory unchanged—at least in some cases? Dismissing the intuition would not amount to rejecting it 'in force' of the theory (that is, because the theory allows us to 'see' things differently), but rather it would amount to denying the effectiveness of the judgement elicited by the case as that judgement is 'just an intuition'.

Problems do not stop here. The qualms about the role of intuitions elicited by thought experiments (TEs) could be extended to the appeal to intuitions in philosophy *tout court*, that is to say, they could be used to call into question the legitimacy of the use of intuitions both in theory-building and in theory-justification. Namely, if we reject intuitions in some specific case, why should we endorse them in others? Here is the question which describes, in general terms, the problem: why should intuitions (what we would say/be inclined to say about X in specific circumstances), be relevant in order to construct, justify or attack theories aiming to establish what X really is (as different from or exceeding what we say or think that X is)? The problem is, ultimately, a problem of compatibility between the nature of the aims and results of the philosophical inquiry, and the nature of the evidence philosophers appeal to.

As we will see in Part II (Chaps. 5 and 6), this is a question for all of those who advocate the legitimacy of the methodology philosophers do in fact adopt, that is, the possibility of achieving the expected results moving from judgements of the sort of those expressed at the end of TEs.

In Chapter. 5, I will take into account Williamson's position. Like Weatherson, Williamson faces the question of the compatibility between the nature of the evidence and the nature of the aims and results of philosophical inquiries from the perspective of TEs. He asks: how can the judgement we express on X at the end of a TE be used to attack a theory that is supposed to state necessary truths about the entity X?

Williamson thinks that the practice of attacking theories on the basis of judgements of this kind is perfectly legitimate. For instance, the judgement epistemologists express at the end of Gettier examples falsifies the JTB theory of knowledge. His solution? Denying that when dealing with this kind of judgement, one has to do with intuitions; rather, one has to do with facts. Williamson argues that the knowledge we acquire by thinking about real or imaginary cases should not be described in terms of what it seems to us that X is, that is, in terms of knowledge of our mental states, but rather as knowledge of X, the non-psychological object of our inquiry. For example, by thinking about Gettier cases, one has access to the fact that the subject does not know; not to the fact that it seems to one as the subject does not know (*see* Williamson 2007, p. 211). This thesis does not concern Gettier's verdict exclusively nor, in general, the verdicts expressed at the end of TEs, but all judgements philosophers characterize as *intuitive* and treat as evidential basis for their inquiries.

Williamson presents the discontinuity between the nature of intuitions and that of a theory as a conflict between judgements of a psychological kind and results of a non-psychological kind. However, his objection is directed towards any position on the matter which refers to uses or concepts, and therefore against the most obvious and acknowledged view on the nature of intuitions, that is, the idea that judgements are the expression of what we would say or think about X in a series of cases. In short, Williamson's thesis is this: if we want the actual method, the armchair method of philosophy, to provide us with evidence pro or contra philosophical theses of a non-psychological/non-linguistic/non-conceptual subject matter, then we have to recognize the non-psychological/non-linguistic/non-conceptual nature of the evidence we obtain by thinking about the cases.

I criticize Williamson's decision to re-describe the so-called *intuitive* judgements in factual terms and propose to maintain the classical view on judgements about cases: the idea that intuitions are the expression of what we would say about X, in the light of our linguistic (or conceptual) competence, or, depending on the object under inquiry, in light of our inferential competence, our moral capacity and so on. In addition, I propose to conceive the request of a theory on X as a demand for explicit norms for the use of the term 'X'/for the application of the concept *X*.

Sections 5.1.2 and 5.1.3 expound the criticisms Williamson would probably make of this idea, in particular, the reasons he advances against the renegotiation of the nature of the aims and results of philosophical inquiry in conceptual terms. Williamson believes that conceiving the object of philosophical inquiry as conceptual amounts to conceiving it as psychological, and that, therefore, research on concepts cannot satisfy the expectations that philosophers have when they ask for a theory on X (see Williamson 2007, p. 211).

Hence, Sect. 5.2 is a defence of my proposal against these two criticisms. Specifically, Williamson's second criticism will be re-phrased and discussed along the following broad lines: judgements philosophers move from in order to build and justify their theories are the expression of what we would say (or think) about X in the light of our linguistic (conceptual) competence (or, depending on the object under inquiry, our inferential competence, moral capacity, and so on). However, philosophical results cannot be described as mere systematizations of what we say (or think)/would say (or think) about X: in fact, most (perhaps, all) philosophical theories are corrective of our uses and judgements. Moreover, given the way philosophers present their goals and results, theories should be the expression of what X really is—as different from or exceeding what we say or think that X is. But, when so, how can the appeal to intuitions match the expectations philosophers have on a theory on X?

I will argue that, in order to answer this question, one needs to consider: (1) the exact nature of the non-psychological object of philosophy; (2) the philosophical method of inquiry; and (3) the evidence philosophers appeal to. I will claim that the appeal to intuitions is legitimate if one satisfies three conditions: (a) one conceives the request of a theory on X as a demand of coherent and precise norms for the use of 'X'/for the application of *X*; (b) one takes reflective equilibrium (RE), a descriptive and revisionary (corrective and enhancing) method, as the method for theory construction and justification; and (c) understands intuitions as the expression of what *should* be said, that is, of what philosophers think is correct to say in light of their competence and reflection.

The arguments concerning points (a) and (b) are presented at the end of Chapter. 5. The arguments supporting the idea that intuitive judgements have a normative aspect (point (c)) are developed in Chapter. 6. To defend (c), I appeal to the common assumption that the linguistic (inferential) practice is normative and consider the reflection which is needed to assess (at least *prima facie*) the possibility that a certain use (a certain inference) has to be explained by a norm, that is, by a consistent and accurate framework. To explain why, in spite of their apparently descriptive aspect, intuitive judgements have a normative nature, I introduce a parallel with Millikan's Pushmi-Pullyu Representations.

In Part III, I will claim that (c), the normative approach to intuitions, and (b), the idea of RE as a descriptive and revisionary method, are also effective against the second big criticism of the appeal to intuitions in philosophy, that is to say, experimental philosophers' objection to intuitions' variability and disagreement.

Although experimental philosophers sometimes hint at the first problem (that is, at whether what we would say about the cases could be legitimately used in order to build, justify or attack theories that are not descriptive of our way of judging), their main interest lies 'at the roots' of the appeal to intuitions. On closer inspection, the judgements philosophers express at the end of TEs are hypotheses on the way in which people would respond to the cases. Are these empirical hypotheses well-grounded? Experimental philosophers maintain that philosophers can legitimately appeal to intuitions only if intuitions describe people's actual responses to cases. However, they continue, that is not something that can be assessed from the armchair. They conclude that philosophers need to verify empirically the correctness of so-called intuitions.

Experimentalists' claims are discussed in Chapter. 7. My arguments against experimental philosophers' results and main tenets are presented in Chapter. 8. In detail, Sect. 8.2 is devoted to the discussion of some general issues concerning the methodology of experimental psychology and to the analysis of experimental philosophers' studies under this aspect. Apropos, it is argued that experimental philosophers' inquiries should not be considered conclusive evidence for any theory about how Westerners and East Asians, women and men, and so on, tend to judge philosophically relevant cases. Two conclusions in particular cannot be drawn: (i) that the way people judge depends on/is shaped by their gender, culture, and so on; (ii) that philosophy is a prerogative of groups of persons having specific characteristics ('the Philosophy club'). Moreover, I cast some doubts about the possibility of qualifying these studies as really informative surveys. Finally, I introduce an alternative explanation for the (apparent) tendency of a large proportion of people belonging to under-represented categories in the philosophical community to give non-standard answers to TEs. This explanation is based on the notion of stereotype threat and on the phenomenon of the confirmation of the negative stereotype (Gendler 2011).

In Section. 8.3, I argue that empirical investigations such as those carried out by experimental philosophers are not philosophical and, more interestingly, not necessary to philosophy. In particular, studies showing a mismatch between non-philosophers' and philosophers' judgements would not be reason enough to question the reliability of philosophers' judgements. Namely, philosophers' verdicts are neither hypotheses on how laymen would judge the cases, nor hypotheses on how people would use a term (would reason, behave, and so on) in those circumstances. Philosophers' judgements are, rather, verdicts of correctness concerning a certain use (reasoning, action), developed on the basis of philosopher's competence and reflection.

References

Gendler, T.S 2011. On the Epistemic Costs of Implicit Bias. *Philosophical Studies* 156: 33–63.

Williamson, T 2007. *The Philosophy of Philosophy*. Malden, MA: Blackwell.

5

The Nature of the Philosophical Enterprise

5.1 The Philosophy of Philosophy

5.1.1 Philosophy Versus Psychology

What is the nature of the evidence provided by thinking about hypothetical cases, such as those presented in thought experiments (TEs)? Is it psychological, as those who speak about intuitions seem to think, or not? This problem is closely related to that of the nature of the subject matter of philosophy, which most philosophers tend to conceive as non-psychological.

In the seventh chapter of *The Philosophy of Philosophy* (2007), Williamson argues against a position that he dubs 'the psychological view'. He takes the psychological view to be the mainstream stance on the nature of the verdicts expressed at the end of the TEs and, in general, of all the verdicts that philosophers treat as the evidential basis for their

This chapter is the extended version of an article, 'Williamson on the psychological view', forthcoming in *Argumenta* (Nicoli 2016).

© The Author(s) 2016

S.M. Nicoli, *The Role of Intuitions in Philosophical Methodology*,
DOI 10.1057/978-1-137-56715-4_5

inquiries. His presentation of the view and argument against it can be briefly reconstructed as follows.

Many contemporary analytic philosophers 'think that, in philosophy, ultimately our evidence consists only of intuitions (to use their term for the sake of argument). Under pressure, they take that to mean not that our evidence consists of the mainly non-psychological putative facts which are the contents of those intuitions, but that it consists of the psychological facts to the effect that we have intuitions with those contents, true or false. On such a view, our evidence in philosophy amounts only to psychological facts about ourselves' (Williamson 2007, p. 235). What philosophers are then supposed to do is 'to infer to the philosophical theory that best explains the evidence. But since it is allowed that the philosophical questions are typically not psychological questions, the link between the philosophical theory of a non-psychological subject matter and the psychological evidence that is supposed to explain becomes problematic' (Williamson 2007, p. 5). In particular, as Williamson argues, the psychological view ends up encouraging scepticism, since 'psychological evidence has no obvious bearing on many philosophical issues' (Williamson 2007, p. 234), which mainly concern non-psychological matters. This view, and intuitions-talk altogether, should therefore be abandoned. Indeed, philosophers should recognize the non-psychological nature of the evidence they have access to: 'our evidence in philosophy consists of facts, most of them non-psychological, to which we have appropriate epistemic access' (Williamson 2007, p. 241).

Let us take, for instance, the Gettier cases and the theory these cases are supposed to provide evidence against—the justified true belief (JTB) theory of knowledge. A genuine counterexample to the JTB theory of knowledge would be a case of a justified true belief without knowledge, not, as tradition presents it, the fact that it seems to one that this is the case. This would, in fact, raise 'the challenge of arguing from a psychological premise, that I believe or we are inclined to believe the Gettier proposition [the proposition that the Gettier subject has a non-knowledge justified true belief] to the epistemological conclusion, the Gettier proposition itself. The gap is not easily bridged' (Williamson 2007, p. 211). So the psychological proposition that it seems to one as if the subject in the Gettier case has a non-knowledge justified true belief is not and cannot be a counterexample to the JTB theory of knowledge (call this 'the gap problem'). What is rather needed, for the argument to work, is the fact itself: the fact that the subject does not know.

Therefore, Williamson concludes, if we want to keep the idea that TEs can in fact provide us with evidence pro or contra certain generalizations, we have then to acknowledge that, thinking about the scenarios described in the TEs, we have access to the relevant facts themselves. More specifically, in the Gettier case we have access to the fact that the subject does not know, which is expressed through the (non-psychological) proposition that the subject lacks knowledge, or better, given that the scenario presented is a counterfactual, through the counterfactual judgement that if someone was in a Gettier case, than she would have a justified true belief without knowledge.

As Brown (2011) points out, a similar way to argue is set out by Deutsch (2009). Discussing Kripke's Gödel case (see Section 7.1, p. 000), Deutsch argues that a genuine counterexample to the descriptive theory of reference would involve a claim about a term and its referent, rather than claims about anyone's intuitions. He writes:

> The predictions of the theory of reference concern terms and their referents, not competent speakers and their intuitions. For example, *D* [the descriptivist theory of reference] predicts that, in Kripke's fiction, the relevant speakers' uses of 'Gödel' refer to Schmidt, not Gödel. If the prediction is false, so is the theory, but the theory makes no predictions at all concerning who will intuit what. Hence, in presenting the Gödel case, Kripke does not, and need not, make any claims about competent speakers' intuitions. He need only say, as he does, that the speakers' uses of Gödel, in the cases he describes, do not refer to Schmidt, contrary to the prediction about the case implied by *D*. (Deutsch 2009, p. 445)

Briefly, this is how Williamson's and Deutsch's argument can be summarized: given the fact that the subject matter of philosophy is non-psychological and that the way to investigate it is widely based on judgements expressed at the end of TEs, how should we conceive the nature of those judgements? Not as psychological—we would in fact expose ourselves to the sceptical challenge—but as directly concerning the object of our inquiry.

Williamson's and Deutsch's argument is straightforward, but nevertheless suspicious. One could in fact object that, by arguing that the evidence must be conceived as non-psychological in order to make a certain kind of project plausible, they are addressing the wrong problem: the problem does not consist in deciding how the evidence provided by armchair methodology should be understood in order to avoid the sceptical challenge. On the contrary, what we should ascertain is whether the data collected through thinking about actual and hypothetical cases could in fact provide evidence for the relevant object of inquiry. Namely, what we are trying to understand is whether the actual methodology is suitable for a certain kind of project, or not. We could end up answering yes, that given the nature of the evidence provided by thinking about what we would say in actual and counterfactual situations, one could in fact engage in that kind of project; or no, that one could not. If the latter was the case, we would have two choices: we could conclude—as the sceptics do—that the actual methodology is inappropriate, and that philosophy, as it is pursued, is a hopeless enterprise; or we could conclude that the methodology is appropriate, but we have characterized the object of philosophical inquiry in the wrong way.

An attempt to avoid the move Williamson makes is proposed by Brown (2011). Brown agrees with Williamson on the two points we have just considered: the nature of the subject matter of philosophy and the characterization of its method. However, Brown denies that arguing in favour of the possibility of this method providing evidence for the object, so conceived, forces us to review the classical position about the nature of the evidence provided by TEs (and by armchair methodology in general): if the only problem with the classical view on intuitions were its vulnerability to the sceptical challenge, then, in order to go on supporting it, it would be sufficient to find a good argument against the sceptics.

Let us see how Brown argues. First, she proposes an analogy with the question of the nature of perceptual evidence. Then, she examines the different strategies supporters of the psychological view for perception adopt against scepticism and consider their effectiveness for the matter at issue. I will not take into account the detailed report of the internalist solutions that Brown makes and the reasons she gives for rejecting them, but I will move directly to the approach she favours: reliabilism.

Pursuing the parallel between perception and intuition, she argues in this way: suppose that the method of forming beliefs about the external world—about the non-psychological facts of philosophy—on the basis of perceptual experiences—on the basis of psychological/intuited propositions—is reliable, then beliefs formed in that way—perceiving or intuiting—are correct. But how can we establish if a certain belief-forming method is generally reliable? Usually, she answers, by evaluating whether the appropriate external relations hold. Here is an example: 'suppose that, when one has the experience as of a large barking dog in front of one, one forms the belief that there is a large barking dog in front of one. On the externalist approach to justification, such as reliabilism, as long as the appropriate external relations hold, the beliefs so formed are justified' (Brown 2011, p. 513). Applying this remark to the case we are interested in—the Gettier case—gives the following answer: suppose that, when one has the intuition that the subject in the Gettier case does not know, one forms the belief that subject does not know. According to reliabilism, as long as the appropriate external relations hold, the beliefs so formed are justified. Therefore, in order to affirm (not just to suppose) that a belief-forming method is reliable and the particular belief so formed is correct, the challenge is to check whether the external relations, that is, the relations in the world corresponding to those expressed by the majority of the beliefs formed by that beliefs-forming method, generally hold.

Unfortunately, Brown does not go further: she does not explain either how these relations are conceived, nor how one is supposed to check if they hold.

5.1.2 Substantial Inquiry Versus Conceptual Analysis

In the following section, I introduce a third solution to 'the gap problem', which is alternative to Williamson's and Brown's. At the beginning, however, I present an attempt to implement Brown's suggestion.

Let us go back to Gettier cases. The argument is traditionally presented in the following way: if the JTB theory of knowledge were true, it would follow that the subject in the Gettier case would know that *p*, but it is clear that we would not say that he does. Here, the idea is that, given the way we use 'knowledge' (or given our concept of knowledge), this belief would not be called 'knowledge' (would not be categorized as *knowledge*).

Let us then face the question at issue: what does entitle us to say that Gettier's conclusion is not wrong? How do we know that the judgement Gettier or, in general, epistemologists express at the end of the case is correct? One can answer that we can legitimately believe that the judgement they express is correct—can in fact provide evidence for the object under inquiry, where the object under inquiry is the norm governing the use of the word 'knowledge' (or the norm governing the application of the concept *knowledge*) in our community—on the basis of their semantic (conceptual) competence, that is to say their ability to use the word 'knowledge' (to apply the concept *knowledge*) in a manner that tends to reflect the way that word is used (that concept is applied) in the community.

So, let us generalize this proposal and see how it can be applied to Brown's suggestion: the particular judgement on X the philosopher expresses at the end of the TE is correct—it gives the inquirer legitimate evidence about X, where X (the non-psychological object of our inquiry) is the norm that governs the use of the term 'X' in our community—under the proviso that the way the inquirer herself forms judgements on X is generally reliable. But how do we know that the way the inquirer forms her judgements on X is generally reliable? Usually, by considering whether she is semantically (conceptually) competent—that is to say, capable of using 'X' in inferences that connect 'X' with other words (that connect the concept *X* with other concepts) and able to apply the word 'X' (the concept *X*) in correspondence to real and imaginary situations, in a way that tends to reflect the way that same word is used in the community. Hence, if the way she usually uses 'X' actually matches the way the same word is used in the community—if these external relations hold—then we could say that she is competent. Furthermore, being the particular judgement she expresses at the end of the TE yielded by her competence, we can conclude that this judgement is correct. Namely, competence, the ability that guides one in everyday life, enabling one to be successful, is the same ability that guides one in the imaginary case.

At this point, two remarks are needed: the former concerns competence and the idea that the capacities involved in the evaluation of the cases described in the TEs are ordinary capacities. The latter concerns the way in which the nature of the non-psychological object, the psychological data at our disposal are supposed to provide evidence for, can be conceived.

First, saying that the ability we use to evaluate a case like Gettier's is an ordinary ability means recognizing—as Williamson does—that no special faculty is involved in the evaluation of TEs. In this respect, my position is close to his. Williamson, however, would refuse to describe the ordinary capacity (or better, one of the ordinary capacities) involved in the evaluation of TEs in terms of linguistic or conceptual competence: notably, all the first part of *The Philosophy of Philosophy* is devoted to argue against the idea that philosophy is committed with something peculiarly conceptual or linguistic. Later on, I will illustrate Williamson's criticisms of the idea of philosophy as conceptual analysis. In the next paragraphs, I shall merely illustrate the differences between my position and his, insofar as the evaluation of the imaginary scenarios described in TEs is concerned.

> Williamson argues: imagining counterfactual scenarios and inferring justified and informative conclusions from them are not the results of a mysterious *sui generis* faculty, but rather of an activity that is based on the application of ordinary cognitive capacities. I, too, reject the idea of a special faculty of intuition. Nevertheless, there is an important difference between my description of the judgement-formation process and Williamson's. In particular I think that the abilities involved in the evaluation of Gettier case-like scenarios, even if mundane (that is, not peculiarly philosophical), are in fact different from those involved in the evaluation of other kind of scenarios, that is, scenarios that nobody takes to be of any philosophical interest. To explain this point, it will be useful to look Glock (2010). As Glock points out in 'Discussion: from armchair to reality?' Williamson's proposal is based on a three-step argument: first, philosophers can attain knowledge of metaphysical necessity and possibilities through knowledge of counterfactuals. Thus, we can achieve the insight that it is possible for a subject *S* to have a justified true belief that *p* without knowing that *p* by appreciating that if *S* were in a Gettier-type situation, *S* would have true justified belief yet lack knowledge. Secondly, the ability to attain knowledge of such counterfactuals is a straightforward extension of our ability to know empirical counterfactuals such as 'If the bush had not been

> there, the tumbling rock would have ended up in the lake.' Thirdly, in both cases we rely on the use of imagination. (see Glock 2010, p. 347)

So, according to Williamson, we evaluate a philosophical counterfactual like (1) 'If *S* were in a Gettier type situation, *S* would have true justified belief yet lack knowledge', in the same way as we evaluate an empirical counterfactual like (2) 'If the bush had not been there, the tumbling rock would have ended up in the lake', that is to say by 'using a mixture of imaginative simulation, background information and logic' (Williamson 2007, p. 2). Now, Glock argues, 'this is ingenious but not entirely convincing'. One of the several reasons he pits against Williamson's proposal is the following:

> even if empirical and philosophical counterfactuals have the same form, distinct cognitive capacities could be required for establishing them. Even if a verdict like (2) is based on estimate rather than measurement and calculation, that estimate relies on empirical beliefs, e.g., the weight of the rock, the gradient and constitution of the slope and the sturdiness of the bush; and it is honed through experience of similar events. By contrast, verdicts on Gettier cases do not depend on empirical knowledge. (Glock 2010, p. 347)

So, we could say: while the imaginative development of the consequent of (2) is based on (and constrained by) our knowledge of empirical facts, or, as Williamson would say, is based on (and constrained by) folk physic, that of the consequent of a counterfactual as (1) is based on our knowledge of the facts concerning the use of the word 'knowledge' in our community. In this process there is nothing peculiarly philosophical at stake: our semantic (or conceptual) competence is all we need. This ability is, in fact, an ordinary ability; however, it cannot be easily likened, as Williamson does, to those involved in the evaluation of (2).

Let us now pass to the question of the nature of the non-psychological object, the psychological data at our disposal are supposed to provide evidence for.

I see my solution as a development of Brown's idea but in no way do I mean to suggest that she would endorse it. I said that the judgement on

X stated at the end of the TE is correct as long as a person is semantically/conceptually competent, that is, capable of using 'X' in a manner that tends to reflect the way the word is used in the community she and we belong to. But at this point, one could point out the following: this solution works under the assumption that what we are in fact looking for are the norms for the use of word 'X' (norms for the application of concept *X*). Yet, this is in tension with Williamson's assumption: that what we are looking for are the necessary true propositions about the entity X. In other words, it is pretty clear (although, as I will point out in Section 5.2, not trivial) how competence can provide mostly reliable data about the object under investigation when the query of it is understood in terms of a query of rules for the usage of the term in question. By contrast, it is not clear how competence could provide any evidence if the object of our inquiry were substantial, were the entity itself, as Williamson puts it.

Later on, I say more about how an inquiry starting from our competence is supposed to lead us to the norms for the use of 'X'. For the time being, it is important to remark that Brown would probably disagree with the way in which I tried to substantiate her proposal. She in fact believes, as Williamson does, that, when we ask 'What is X?', we are asking about the necessary truths about the entity X. Brown does not mention the possibility of conceiving the aims and results of philosophical inquiry differently. However Williamson does.

In the seventh chapter of *The Philosophy of Philosophy*, Williamson asks: could a reinterpretation of the nature of the aims and results of the philosophical inquiry in conceptual terms lead to an alternative resolution of the gap problem? Here is his answer:

> Attempts have been made to close the gap by psychologizing the subject matter of philosophy. If we are investigating our own concepts, our application of them must be relevant evidence. But this proposal makes large sacrifice for small gains. As seen in early chapters, the subject matter of philosophy is not distinctive in any sense. Many epistemologists study knowledge, not just the ordinary concept of knowledge. Metaphysicians studying the nature of identity over time ask how things persist, not how we think or say they persist. (Williamson 2007, p. 211)

This consideration is made in the wake of the criticisms Williamson made in the first part of *The Philosophy of Philosophy*. There he argued against the idea of philosophy as a conceptual inquiry on the basis of two points: (1) the incapacity to discriminate between allegedly conceptual questions and obviously non-conceptual ones; and (2) the non-viability of the notion of analytic truth, that he sees as the best explication of the notion of a conceptual truth.

Let us briefly illustrate these points and start with (1): most philosophers would consider such a question as 'Was Mars always either dry or not dry?' to be conceptual. They would say this question is about the predicate 'dry', or the concept *dryness*, and that its explicit content is 'Is the concept of dryness vague?'. However, Williamson objects, there are no criteria on whose grounds we can prove this question is different from 'Was Mars always either uninhabited or not dry?'—a question that is obviously factual. It is important to remark that Williamson does not deny that, in order to answer the first question, linguistic and conceptual considerations do in fact play a central role. In other words: he agrees with the common denominator of the traditional approaches to vagueness, that is, addressing the features of the logic adopted. However, Williamson disagrees with the traditional accounts of vagueness on what the problem is about: although linguistic and conceptual considerations play a central role in the solution of the problem, we do not need to regard the question to be about language or thought. On the contrary, the problem, 'Was Mars always either dry or not dry?' concerns what in fact it seems to concern: the physical object (Mars), and the property it is literally about (dryness).

The necessity to take philosophical questions and claims literally is a central concern for Williamson: philosophical questions, he thinks, are not about language or thought, unless they are so explicitly. The refusal of the thesis of the intrinsic linguistic or conceptual nature of philosophical questions is connected to the refusal of the thesis of isolationism about philosophy, that is then specified in the positive thesis of the continuity between philosophy and natural science. Philosophy, like science, investigates reality, the one and same reality, or, otherwise expressed, philosophy, like science, aims to provide substantive knowledge of the world.

But, if this is the idea, it is then clear that Williamson does not take reality to coincide only with the physical world. Reality, for Williamson, 'would include numbers and other logical and mathematical entities, as well as knowledge, truth, time, and more' (Marconi 2011, p. 97). Once again, it is important to stress that, despite defending the thesis of the continuity between philosophy and science regarding to the subject matters of their inquiries, Williamson recognizes the peculiarity of the philosophical method with respect to the scientific one: philosophical investigations are usually performed from the armchair, whereas experimental investigations of science are generally not. It is exactly from the relation of this point with the thesis of the substantiality of the subject matter of philosophy that the problem here at issue stems: how can the method actually applied by philosophers provide philosophers with knowledge about the objects of their inquiries?

5.1.3 Armchair Reflections: Is Philosophy Like Mathematics?

Williamson's main argument in favour of armchair methodology is built on a parallel with logic and mathematics: mathematicians and logicians acquire substantive knowledge of the world by theorizing from the armchair, and philosophers do likewise. This argument, however, is questionable: namely, that logical and mathematical reflection in fact provides substantive knowledge of the world is not an issue that we can regard as settled. As Glock rightly points out 'it is a moot point whether they [logicians and mathematicians] provide knowledge about reality, rather than explicating mathematical concepts' (Glock 2010, p. 341). In defence of Williamson one could argue that, after all, this is not really important: the nature of the results is not an issue that needs to be decided preliminarily.

> We can say that epistemology is about knowledge (not about the concept of knowledge or the word 'knowledge') or that ethics is about the good even without having preliminarily established that knowledge and the good are physically respectable entities. It is not obvious that there are only physically respectable entities (Marconi 2011, p. 98–99)

Nevertheless, this consideration only partially supports what Williamson says: what Williamson was in fact trying to do, by appealing to mathematics and logic, was exactly to defend the possibility of the philosophical method providing substantial knowledge of the world. Hence, he should rather justify or, at least give some reasons in favour of, the thesis that logical and mathematical knowledge are in fact substantial. But, as Frascolla (2011) notices, no attempt of that kind is made:

> Williamson makes use of some fairly questionable arguments in that they assume as unproblematic claims that are, in effect at least as problematic as those for which they should furnish support. For instance, in favour of the claim that, for all that it is restricted to the methodology of armchair thought, philosophy is able to furnish 'substantive knowledge of the world', Williamson appeals to the overwhelming evidence that would be given by the way mathematics and logic, two typical disciplines based on the same methodology, are able to provide very rich information about how things stand in the world. Now, even if Williamson is right when he notes that no-one (not even Wittgenstein or Carnap) has ever managed to make convincing defence of the claim that mathematics and logic are 'trivial' or 'non-substantial', it is still the case that the idea that 'thinking just as much as perceiving is a way of learning how things are' (Williamson 2007, p. 47) looks to me like a claim that cannot be simply taken for granted. Indeed, it looks to me like a claim in urgent need of defence, at least as much as does the claim of the non-substantialness. (Frascolla 2011, p. 89)

Apart from these considerations, there is at least one other possible argument against the parallel between the philosophical and the mathematical methods. Mathematicians reach conclusions as to what mathematical objects are; however they do it on the basis of axioms: they do not say what a set is, but rather what a set is in the Zermelo–Fraenkel system, or in Von Neumann's, and so on. Philosophers are in a different position: what are the axioms of philosophy? Could they be something different from registrations of linguistic uses? Let us take, for instance, a sentence like 'Propositions that are known are true': could it mean something different from 'we do not call a propositional content "knowledge" if we do not take it to be true?' (I shall return to this in Section 5.2.1.) Furthermore, mathematicians can create the objects they argue about, philosophers cannot. So, whereas a mathematician can study objects such as groups or circles by examining particular structures, that person knows such structures are groups or circles (they meet the definition provided of them), a philosopher cannot study the good by studying specific behaviours: in fact, one does not know these behaviours to be good (there is no set of axioms characterizing good things), one only knows they are called 'good', considered good, and so on.

What about (2), that is, Williamson's idea that the notion of analytic truth is the best explication of the notion of a conceptual truth? As Marconi (2011) argues, the idea that a conceptual investigation amounts to the search for conceptual truths and that conceptual truths are analytic truths is, from a historical point of view, 'somewhat out of focus': 'the idea that philosophy is an attempt at establishing analytic truths—truths such as are expressed by "Rectangles have four sides" or "Groundhogs are woodchucks"—is not easily reconciled with our picture of what Wittgenstein, Ryle, or Strawson were up to' (Marconi 2011, p. 91). For instance, according to Wittgenstein, the results of conceptual analysis are the so-called grammatical 'propositions', which 'are not to be identified with the analytic propositions (e.g. Baker and Hacker 1985; Andronico 2007), even though it has been argued that the two sets have a non-empty intersection (Schroeder 2009, pp. 102–105)' (Marconi 2011, p. 92). I will return to this in the next section.

For time being, it is important to remark that, despite being in the wake of the reasons against conceptual analysis Williamson provides in the *pars destruens* of his book, the objection presented in his seventh chapter appears to be slightly different from (1) and (2). Here the idea is that conceptual analysis is not viable because what philosophers are looking for, when they ask what knowledge, identity, reference, causation are, is something non-psychological. Conceptual analysists look rather for concepts, so, Williamson says, something mental and therefore uninteresting for the concerns of philosophy, when they are properly understood.

In what follows, I am going to argue against the problems he raises in the seventh chapter. So, more specifically, against the idea that (3) conceiving the object of the philosophical inquiry as conceptual necessarily amounts to conceiving it as psychological, and the idea that (4) a research on concepts cannot satisfy the expectations that philosophers have when they ask for a theory on X.

5.2 What Kind of Conceptual Analysis?

5.2.1 Philosophers' Aims

Let us start with (3). First of all, claiming that the aims and results of the philosophical inquiry are conceptual does not necessarily amount to saying that they are psychological: indeed, some rules are not. In particular, rules which are outlined and adjusted on the basis of Gettier case-like judgements (that is, judgements stating what we would say in specific circumstances) are not.

At this point a clarification of what I mean exactly by saying that philosophical theories are (non-psychological) norms is in order. To explain it, I will follow the arguments proposed by Marconi in 'Wittgenstein and Williamson on conceptual analysis' (2011). There, Wittgenstein's view of philosophy as conceptual analysis and Williamson's view of philosophy as substantial inquiry are compared.

From an historical point of view, the conception of philosophy as an activity devoted to the discovering of norms can be brought back to Wittgenstein and can be presented in the following way. Let us take an uncontroversial result of epistemology as that expressed by (a1) 'Knowledge entails truth', or by the equivalent: (2) 'Propositions that are known are true'. Wittgenstein (1974, p. 415) would refuse to describe these expressions as propositions stating necessary connections between entities (or properties), that is, the entity/property Knowledge and the entity/property Truth, as Williamson does; but he would neither present them in terms of propositions about the concept under inquiry, as Williamson thinks any friend of conceptual analysis would do. So, if (1) and (2) are not propositions stating necessary connections between entities (or properties), and neither propositions on concepts (conceptual truths), what are they?

In the perspective Wittgenstein defends, statements like (1) and (2) express rules/explicit instructions for the use of the words 'knowledge' and ‘truth’ (or, as he would also say, rules for the application of the concepts *knowledge* and *truth*); or better, rules that set connections among ‘X’ and other words (among *X* and other concepts). Among, for instance, the notion of knowledge and the notion of truth; or, as in the case of the complete formulation of the JTB theory of knowledge, the notion of knowledge, on the one side, and that of belief, truth and justification, on the other. So, being norms, (1) and (2) could be formulated as ‘Call a proposition “a piece of knowledge” only if you are prepared to call it a truth’, or ‘Only apply the concept of knowledge to contents to which you are prepared to apply the concept of truth’. Such connections ‘are called “grammatical” because the rules that establish them are similar to rules of grammar, such as “Medial verbs do not allow manner adverbials” or “The passivization marker is SV-internal”’ (Marconi 2011, p. 92).

It is important to remark that what epistemologists discover is a fact: they discover that, in the practices in which ‘know’ and ‘truth’ are used, speakers use these words in a way that leads them to conclude that they do not generally call a propositional content ‘knowledge’ if they do not take it to be true. Like any other fact, this kind of fact can be discovered, Wittgenstein says, by observing (or by describing) how in fact things go, and specifically how, as a matter of fact, we use the words ‘knowledge’ and ‘truth’. So, discovering how in fact we use the words, we also gain knowledge of how we ought to use them.

Hence, let us go back to Williamson’s argument. Is it true that, in order to defend the armchair methodology of philosophy, we must reconsider the nature of the evidence obtained by thinking about the cases? The answer is no. Namely, while the evidence is psychological, it can lead us to discover something that is not (at least not entirely) psychological: facts about how we use concepts and words. In other terms: if the object of the philosophical inquiry is norms governing the use of ‘X’ in the community, and if intuitive judgements are the expression of the competence, that is, the capacity to use ‘X’ in a manner that tends to reflect how the term is used in the community, then the appeal to them is justified.

Of course, that is not to say that the position I have just defended is without its own problems. Indeed, this view has several well-known problems, especially in the ‘crude’ version I provided above—a version in which it seems that only by observing and describing uses can something like a norm can be achieved.

For instance, a classic objection is the following: everyday linguistic behaviour includes errors and idiosyncratic uses; it is then difficult to say how, by observing and describing uses, or just by considering judgements that are descriptive of the way people (would) use a certain term 'X', one could define a theory on X, that is, a norm for the use of 'X'. What one obtains will be, eventually, a list of all the idiosyncrasies within the speakers' community. Furthermore, uses are often incoherent or inaccurate. So one can ask: how could they then be subsumed by a theory that, in order to be a theory, has to be consistent and accurate?

Moreover, the idea that philosophical theses and theories are just description of uses, or of judgements on uses, or, in a broader sense, the registration of the received opinion on X, seems to be plainly false. First, describing our uses or our intuitions is not what philosophers say they are doing when they investigate X: 'many epistemologists study knowledge, not just the ordinary concept of knowledge. Metaphysicians studying the nature of identity over time ask how things persist, not how we think or say they persist' (Williamson 2007, p. 211). Moreover, describing what we say or think, or what we suppose is correct to say or think, is not what philosophers have done up to now: many (perhaps all) philosophical theories are corrective of our uses and competence. In particular, they are seen as—and in fact are—means to discriminate between what is really correct to say and what we just say or we just think is correct to say.

So, in order to make the third solution plausible, one has to answer the question of how an inquiry starting from judgements that are the product of one's competence can lead to a theory that: (1) has the characteristics a theory has to have in order to be a theory (consistency, accuracy); and (2) satisfies the expectations philosophers have for a theory on X, that is, those expectations that are plausibly subsumed by claims such as 'philosophers want a theory on X, not just on what we think or say that X is'.

In the remaining part of this chapter, I will delve into the questions I have just mentioned and start outlining a strategy for supporting the view just described. In my arguments, two aspects are going to play a central role: the nature of intuitions and the structure of the method leading from intuitions to theory.

In short: against the first battery of objections, I am going to insist on the fact that the evidence philosophers appeal to in order to build, support or attack theories is not the actual use of the relevant words/application of relevant concepts by ordinary speakers/thinkers, but judgements of competent speakers/thinkers. Moreover, these judgements do not consist in mere descriptions of usage, nor in predictions on how people would use a term (or respond to a question asking whether it is appropriate or not to use such a term) in a given situation. An intuitive judgement is rather the description of what a *very* competent speaker would take to be the *correct* usage of that term. In other words: intuitions are hypotheses of correctness that philosophers develop in the light of their competence and reflection. Against the second battery of objections, I am going to point out that norms are not the results of a plain generalization from an initial set of intuitions, but rather the results of the descriptive and revisionist (corrective and enhancing) dynamics of reflective equilibrium (RE).

5.2.2 Describing, Correcting and Enhancing

Let us start with the first battery of objections: how could one come to a theory on X just by observing and describing uses, or just by considering judgements that are descriptive of the way people (would) use 'X' [question 1]? Furthermore, in the light of the errors and idiosyncrasies, does it make any sense to speak about the practice as a norm-governed practice [question 2]?

Let us tackle [question 2]. First, it is worth noticing that it does not make any sense to face the objection by denying the fact itself: in fact, the practice is not wholly uniform, that is, usage includes errors and idiosyncrasies. A reasonable reply to [question 2] consists in pointing out that the idea of a norm-governed practice does not presuppose a complete uniformity of behaviour: it just requires the existence of a shared assumption of normativity about the practice (Marconi 1997, Chapter V).But how can the existence of a shared assumption of normativity support the thesis of the linguistic practice as a norm-governed practice? The main point appears to be the following: any speaker generally aims to conform his or her speech to the same constraint as the others. So, for instance, by using ‘X’, any of us generally assumes that he or she is using the word ‘X’ more or less in the same way the others use it, that is to say with respect to the same cases (not necessarily all cases: agreeing on the most ‘central’ ones is generally enough), and that he or she shares with the others a certain number of beliefs which contain the word ‘X’. It is then plausible to think that the fact that everyone seeks to conform their speech (and thought) to the same constraint as others actually shapes the way in which people speak (and think). In other terms, the fact that any single speaker is inclined to speak as others in the community do makes one speak as others do. An obvious consequence is that uses end up converging.

In the next chapter, I am going to further assess this. For now, the central point is the following: in spite of errors and idiosyncrasies, it is possible to conceive linguistic practice as a norm-governed practice. Moreover, in light of the effect of this common assumption (that is, convergence), it seems that the standpoint of this objection (the risk of deviant behaviours), is exaggerated. In conclusion, [question 2] turns out to be less incisive than expected.

Let us now tackle [question 1]. I have just said that every speaker generally aims to conform his or her speech to that of others. But, obviously, this does not mean that, at times, it would not be preferable to willingly decide not to do so; nor that, despite aiming to do so, one actually succeeds; nor that any of us is able to do so as successfully as some others. Is this a problem for those who aim to define a norm (and not just a list of idiosyncrasies)? I argue it is not: supporters of conceptual analysis do not present their activity as research on some particular person’s use of language—or, more generally, people’s use of language—but rather as a reflection starting from the judgements of correctness of *very* competent speakers. The emphasis on judgements and competence ought to help contain the problem of error and idiosyncrasies. Furthermore, intuitive judgements should not be seen as mere descriptions of usage—nor as empirical hypotheses about usage—but as descriptions of what competent speakers would take to be the *correct* usage. Therefore, judgements philosophers express at the end of TEs—judgements of the form ‘given the way we usually use “X”, this *Y* would (or wouldn’t) be categorized as

X'—should be read as: 'given the way we usually use "X", this *Y* should (or shouldn't) be categorized as *X*'.

In Chap. 6, I will examine this thesis in depth and argue that the normative aspect of judgements is connected, on one hand, to the assumption of normativity about usage, and on the other hand, to the idea of the theory (of the norm) as a consistent and accurate structure.

After this brief anticipation of some of the points connected to the normativity of judgements, let us go back to the point at issue. It is worth noticing that the appeal to the intuitions of a competent speaker is useful not only because it offers a solution to the objection of the error and idiosyncrasies, but also because it enables us to 'enlarge' our reflection on X. Let us imagine a philosopher asking himself or herself 'What is X?' A possible way to proceed is by describing the explicit and implicit connections speakers set between 'X' and other terms. An epistemologist who tries to answer the question 'What is knowledge?' will look, for instance, at the way the term 'knowledge' is used, especially in connection with other terms, such as 'belief', 'truth' and 'justification'. On the basis of his or her observations and reflections on usage, the epistemologist could conclude that knowledge is justified true belief. Nevertheless, as is clear from the Gettier cases, this conclusion is not correct: being justified and true is necessary for a belief to be knowledge, but is not sufficient. Apart from the significance these cases have for epistemology, what Gettier examples teach us is that reflection on uses of 'knowledge' in situations we generally come across is not enough. That is why philosophers, besides reflecting on cases close to home, try to imagine other cases—unusual or imaginary cases—and ask themselves: what would we say in this situation? Would we call this belief 'knowledge' or not? Clearly, the answer to this question will not consist in the description of some particular person's use of 'knowledge', but in a verdict of who is asking the question (the epistemologist) about the applicability of the term 'knowledge' to that situation.

It is interesting to notice how the reasons for the indispensability of the appeal to intuitions in epistemology are identical to those in favour of the appeal to intuitions (of acceptability/unacceptability) in linguistics. As Schütze (2010) argues, in linguistics the observation of the use is not enough for two reasons: it is possible that we never come across certain uses that are considered acceptable, and especially it is possible that we never come across uses that are considered unacceptable. Furthermore, spontaneous use (even our own use) includes errors that we, by ourselves, would judge to be errors. That is why acceptability judgements are taken to be superior to usage (Schütze 2010, p. 210).

Let us now consider the second kind of objection presented earlier: those criticisms according to which, given the expectations philosophers have of a theory on X and the kind of results they do in fact reach, a methodology based on intuitions cannot be adequate. What philosophers are looking for—critics say—is a theory on X, not about what we say or think that X is (Williamson's objection (4)). Moreover, by looking at the theories philosophers build or have built in the past, it is clear that they are not descriptions of what we say or simply think is correct to say: in fact, theories are correctives of our uses and intuitions. In sum, theories are seen as—and are—means to discriminate between what is really correct and what we just say or we just think is correct.

The problem can be presented also in this way: when we speak about norms what we seem to presuppose is a clear distinction between what seems right and what is right. So normative claims about correct use cannot be derived from facts about actual usage or from simple intuitions of correctness. The very notion of normative constraint opens a gap between how 'X' is used (or how we think it is correct to use it) and how 'X' should be used (or how it is correct to use it). A gap that, in a perspective founded only on the exercise of the competence only, cannot be bridged.

What kind of strategy could be adopted against objections of this kind? My answer consists in an appeal to the method the reflective equilibrium (RE). In Chap. 3, I pointed out that philosophical theories which are the product of RE cannot be seen as the results of a plain generalization from an initial set of intuitions. They are rather the result of an articulated process, in which different theoretical hypotheses—obtained by inference to the best explanation (IBE) from the initial set of intuitions—are confronted with other intuitions, that have not been taken into account before and which emerge from the reflection on factual or counterfactual cases. Usually, when a theory disagrees with these intuitions, then the theory is modified or eventually abandoned. The same thing, however, can happen to intuitions. Namely, we may happen to find a certain theory especially persuasive and the way we normally would have judged the case changes: we start seeing things the way the theory predicted.

Generally speaking, the process taking the analyst from his or her intuitions to the norm can be described as a procedure which, through IBE and through corrections and improvements, aims to set an equilibrium (a reflective equilibrium) between the norm and the intuitions. Eventually, also other beliefs (for instance scientific ones), we think are relevant to the problem at issue, can be used in this process.

So, to conclude and to answer to objection (4): it is clear that the product of a process of this kind cannot be seen as the simple description of what we think or say about X, and neither as the simple description of what we believe is correct to say about X. The norm that we obtain is corrective and enhancing of what we think and say. In particular, it can be used to discriminate between what is correct and what we just thought or said was correct. And this—it can be argued—is enough to satisfy the expectations philosophers have when they ask for a theory on X, as opposed to a theory merely describing what we believe X to be.

It is worth noticing that the process described above, that is, a process which is corrective and enhancing of the products of our competence, is different from the one presented in the section devoted to Wittgenstein, that is, a merely descriptive procedure. Still, it could be argued that the way a conceptual philosopher like Wittgenstein in fact works is much more similar to the way lined up by the supporters of RE:

> The kind of reflection that Wittgenstein has in mind, which is to a large extent reflection about, or from, one's own semantic competence rather than about some particular person's 'use of language', does not just amount to exercising one's semantic competence. As we know from Wittgenstein's writings, it involves reflection of considerable 'length and depth': it includes recourse to thought experiments and the creation and comparison of many examples; it does not rule out appeal to scientific results and to logic; and it exercises ordinary reasoning and, to a vast extent, imagination. (Marconi 2011, p. 95)

After this brief digression, let us go back to the main argument of this section: I replied to objections that lay claim to the necessity of distinguishing correct uses from only seemingly correct uses, by explaining how the results of RE procedure do in fact allow us to make this distinction. In the light of the characterization of the analyst's method that I have provided, it would then equally be possible to conclude that those who have undertaken a procedure of that kind really do have a complete mastery of the norm. In fact, they are not only able to use 'X' in a way that approximately matches how others use it, but know—are able to tell—what is the rule for a correct use of 'X'. In particular, they know which, among our effective uses and uses that we tend to judge as correct, are really correct, which are not, and why. I will call these persons 'experts'.

Hence, the correctness of a particular use, or of a particular intuition, will be determined on the basis of the rules resulting from a process of RE, or, if we prefer, on the basis of the convergence with the uses experts would sanction to be correct. These uses match, to a significant extent, the uses of the speakers of the community the expert is in, but they also partially differ from them for reasons that the experts, having conducted the process of negotiation typical of the reflective dynamics, know and are able to provide.

However, objection (4) might be suggesting something stronger than that which we answered to by saying that particular uses or only seemingly correct uses can be corrected in light of authoritative uses, or on the basis of the results of the reflective process. The objection might be suggesting that to make sense of the notion of normative constraint, norms for the use of 'X' should be conceived as norms that determine whether 'X' should be applied (or not) to a specific case, independently of what any of us—experts included—actually judge or will judge. In other words, it may be that one should, on the basis of these rules, judge that X (or judge that non-X), even if everybody in fact judges (or would judge) otherwise. As a reply to the objection, it can be asked where, if not from our use and intuitive application of 'X', we could start from to define the rules for its use. In addition, if the alleged norms for the application of 'X' (let us say 'knowledge') can be unrelated to how we in fact apply 'knowledge', then it is natural to ask: why should anyone adopt these rules? As the question underlines, these rules would be essentially impossible to comply with, and, after all, useless, with respect to what we, by asking ourselves 'what is X?', were plausibly looking for, that is, precise and explicit rules that enable us to distinguish which of our uses are really correct and which are not.

References

Andronico, M. 2007. Analitico/sintetico vs. grammaticale/fattuale: l'analisi concettuale ai tempi della naturalizzazione. *Rivista di Estetica* 47: 41–59.

Brown, J. 2011. Thought Experiments, Intuitions and Philosophical Evidence. *Dialectica* 65(4): 493–516.

Deutsch, M. 2009. Experimental Philosophy and the Theory of Reference. *Mind and Language* 24(4): 445–466.

P. Frascolla. 2011. On the face value of the 'Original Question' and its weight in *Annuario e Bollettino della Società Italiana di Filosofia Analitica* (SIFA).

Glock, H.J. 2010. Discussion: from armchair to reality? *Ratio* 23: 339–348.

Gordon P. Baker & P. M. S. Hacker (1985), Wittgenstein: Rules, Grammar and Necessity, Blackwell

Marconi, D. 1997. *Lexical Competence*. Cambridge, MA: MIT press.

Marconi, D. 2008. *Filosofia e scienza cognitiva*. Bari-Roma: Laterza.

D. Marconi (2011) 'Wittgenstein and Williamson on conceptual analysis', in Annuario e Bollettino della Società Italiana di Filosofia Analitica (SIFA).

Nicoli S. M. 2016. Williamson on the psychological view. *Argumenta.*

Schroeder, S. 2009. Analytic truths and grammatical propositions. In *Wittgenstein and Analytic Philosophy*, eds. H.-J. Glock and J. Hyman. Oxford: Oxford University Press.

Schütze, C.T. 2010. Linguistic evidence and grammatical theory. *Wiley Interdisciplinary Reviews: Cognitive Science* 2(2): 206–221.

Williamson, T 2007. *The Philosophy of Philosophy*. Malden, MA: Blackwell.

———2011. Philosophical Expertise and the Burden of Proof. *Metaphilosophy* 42: 215–229.

Wittgenstein, L. 1974. In *On Certainty*, eds. G.E.M. Anscombe and G.H. Von Wright. Oxford: Blackwell.

6

The Nature of Intuitive Judgements

6.1 Intuitions: Empirical Hypotheses or Judgements of Correctness?

In Chap. 5, I claimed that when one asks 'What is X?', one is asking about explicit norms for the usage of 'X' (for the application of the concept *X*): norms that can be used to discriminate between correct uses of 'X' and only seemingly correct uses of 'X'. This view was presented in juxtaposition to Williamson's—a position according to which, what philosophers are looking for when they ask what knowledge, reference, justice are, are necessary truths about entities (the entity Knowledge, the entity Reference and so on). In advocating my position, I pointed out that, by describing the object of the philosophical inquiry in terms of norms that are corrective and enhancing of our intuitions, one can satisfy the expectations implicit in such requests as: 'we want a theory on X, not a theory on what we say or think that X is'. Furthermore, I showed how this conception allows us to avoid the problem Williamson has to face: explaining in which way judgements like those stated at the end of thought experiments (TEs) can legitimately provide evidence in favour or against theories of a non-psychological subject matter. As regards

© The Author(s) 2016

S.M. Nicoli, *The Role of Intuitions in Philosophical Methodology*,
DOI 10.1057/978-1-137-56715-4_6

this issue, I argued that, in the attempt to keep hold the idea that the object of the philosophical inquiry is substantial (in the sense of non-psychological, but also non-conceptual/non-linguistic) and the idea that the method philosophers in fact use is legitimate, Williamson ends up adopting a perspective on judgements conflicting with the most obvious and acknowledged view on their nature. But what is this view exactly?

By explaining the value of the Gettier examples for epistemology, I presented Gettier's argument in the following way: if the JTB theory of knowledge were true, it would follow that the subject in the GC would know that *p*, but it is clear that we would not say he does. I also argued that, though formulations such as (1) 'we wouldn't say that the subject knows that *p*' or (2) 'given the way we use "knowledge" (or given our concept of knowledge), this belief wouldn't be called "knowledge" (wouldn't be categorized as *knowledge*)' may sound like empirical hypotheses (1) on how people would answer the question 'does the subject know that *p*?', or (2) on how people would use the term 'knowledge' (or apply the concept *knowledge*) if they were experiencing a real-life Gettier case, they are not. Judgements of this kind are the expression of what philosophers think is correct to say in that situation. For instance, a Gettier judgement is a hypothesis of correctness about the use of the term 'knowledge' (application of the concept *knowledge*) that epistemologists elaborate in light of both their acquaintance with the practice (their own semantic/conceptual competence) and their capacity to reflect on it. In general, expression like 'given the way we usually use "X", this Y would (or wouldn't) be called "X"' should be read as: 'given the way we usually use "X", this Y should (or shouldn't) be called "X"'.

What motivated the decision to speak of intuitions in those terms was the objection of errors and idiosyncrasies in usage: how could one come to a norm for the use of 'X' just by observing and describing uses, or just by considering judgements which are descriptive of the way people (would) use 'X'? Very briefly, my strategy consisted in specifying that doing analysis does not amount to describing some particular person's use of the language, or the whole of speaker's uses of the language, but to a descriptive and revisionary practice that, moving from the uses competent speakers sanction as correct, using logics and, perhaps, scientific knowledge and common sense, leads to the formulation of the norm. In other words: the process leading to the definition of the norm is based on judgements of correctness of experts; not on judgements aiming to describe the use of ordinary agents.

While in Chap. 5 I focussed on the method, in the next sub-sections I shall deepen the question relative to the normative aspect of intuitive judgements. In order to argue in favour of the normative view on intuitions, I will appeal, on the one hand, to the common assumption that the practice is normative. This assumption explains why descriptions of the practice tend to be conceived as prescriptions. In particular, it explains why judgements like 'given the way we use "X", this Y would (or wouldn't) be called "X" count as hypotheses on the correct use of the term "X"', rather than as predictions on speakers' usage. On the other hand, I will consider the reflection which is needed to assess (at least *prima facie*) the possibility that a certain use (a certain inference, if we consider logic) has to be explained by a theory, that is, by a consistent and accurate framework.

Before doing this, I shall propose a parallel between intuitions and Millikan's Pushmi-Pullyu Representations (PPR)—representations that, as well as intuitions, have a prescriptive function, in spite of a descriptive content. This analogy will help me introduce some topics that are important for my discussion.

6.2 The Twofold Nature of Intuitive Judgements

I have suggested that an intuitive judgement should be understood as the expression of what the philosopher thinks one should say in the relevant circumstances, rather than as the expression of what the philosopher supposes people would say. In other words, the judgement aims to convey a prescription, not (just) a description of some state of affairs.

Because of their twofold nature, intuitions resemble Pushmi-Pullyu Representations (PPR). PPR, as Millikan describes them; they are those representations—human, but also animal—simultaneously fulfilling two purposes: describing and prescribing. Or better, they are the kind of representation that, while describing a state of affairs, aims to prescribe or inform someone about a certain norm of behaviour. Here is how Millikan introduces them:

> PPRs have both a descriptive and a directive function, yet they are not equivalent to the mere conjunction of a pure descriptive representation and a pure directive one but are more primitive than either. […] Perhaps the most obvious PPRs are simple signals to conspecifics used by various animals—bird songs, rabbit thumps, and bee dances, for example. But PPRs also appear in human language and probably in human thought. Illustrations in language are 'No Johnny, we don't eat peas with our fingers' and 'The meeting is adjourned' as said by the chair of the meeting […] Our inner representations by which we understand the social roles that we play as we play them are probably also PPRs. The natural way that we fall into doing 'what one does', 'what women do', what teachers do', and so forth, suggests this. I suspect that these primitive ways of thinking are an essential glue helping to hold human societies together. (Millikan 1995, p. 186)

In which sense is the content of a statement like 'we don't eat peas with our fingers' similar to the content of an intuition? Here is my answer: just as with a statement describing a social behaviour that is more or less widespread ('we don't eat peas with our fingers'), we do not wish to inform our interlocutor about some state of affairs, but rather about a norm of behaviour ('it is not allowed to eat peas with fingers'), with a statement describing a disposition speakers are supposed to have, we do not intend to make a prediction on the behaviour speakers would actually have, but to point out the behaviour speakers should have in that case.

First and foremost, it is clear that, in the case of most PP statements, the descriptive aspect, though syntactically salient, is secondary with respect to the prescriptive aspect. Furthermore, although it is true that we think it is inappropriate to eat peas with fingers, it is not true that everybody does in fact avoid eating peas with their fingers. On the contrary, the behaviour of some members of our community may not conform to the behaviour the judgement describes. So the question is: does the fact that a PP statement is not entirely adequate as a description jeopardise its truth? The PP statement truth would be challenged, if the statement was in fact descriptive (if by saying 'we don't eat peas with our fingers', a mother wanted to give her child an empirical information). But, since the statement has a prescriptive/normative value (a mother intends to teach her child good manners), then the *empirical* untruth of the PP statement is beside the point—at least, within certain limits.

It needs to be acknowledged that while I have offered a broad outline of the relationship between normative and descriptive aspects of PP statements, there can be large variations in various cases. For example, there can be instances in which the descriptive interpretation is roughly true and has to be true for the norm to be sound (for example, 'in Italian, "un'amica" is written with the apostrophe'), and cases in which the descriptive interpretation is glaringly false and the normative reading of the statement prevails (for example, 'taxes are paid', 'sentences are not to be discussed, but to be executed').

But let us focus on the general aspect of the thesis. I said that a PP statement's truth needs not be challenged by forms of behaviour not conforming to the one described in the judgement. The same thought seems to apply to intuitions. As a PP statement such as 'we do not eat peas with our fingers' needs not to be challenged by the presence of people in our community that eat peas with their fingers, likewise an intuition like 'this belief wouldn't be called "knowledge"' needs not to be challenged by people in our community that use 'knowledge' to describe the belief of the subject in the Gettier case.

In the previous sections, I proposed to rephrase the judgement a philosopher expresses, when confronted with a particular case, in the following way: 'given the way we use "X", this Y would/should (wouldn't/shouldn't) be called "X"'. It is interesting to remark that a similar discourse can be done apropos of the judgements logicians employ to justify inferential rules, and can therefore be used to complete the discussion of Goodman's proposal, that is, the suggestion of describing the process of construction and legitimation of inferential rules as 'the delicate one of making mutual adjustments between the rules and accepted inferences' (Goodman 1955, p. 67).

I am not going to recall here the workings of the process leading from accepted inferences to a possible rule, and then back from the rule to accepted inferences to check whether the rule is satisfactory or not; but I will just focus on the nature of the evidence provided by the so-called 'accepted inferences' or, as Goodman also says, by the 'judgements rejecting or accepting particular inferences'.

It is worth noticing that Goodman speaks about judgement of acceptance/non-acceptance and that when he speaks about the practice, that is about the inferences actually made, he specifies they are *accepted* inferences. So, it is the fact of sanctioning a particular inference as legitimate that is central in his understanding of the nature of the evidence, not the practice directly.

Hence, suppose that a logician engaged in the endeavour of constructing and justifying rules is confronted with a singular case and expresses a judgement like: 'given the ordinary inferential practice, this particular inference would (or wouldn't) be deemed valid'. What does the logician aim to express by a verdict of this kind? Not a prediction on the way in which, either in everyday life or in an experimental context, people would in fact judge, but rather a hypothesis of validity/invalidity about a certain reasoning: given the way we usually infer, this inference should (or shouldn't) be judged as valid. Indeed, what the Goodmanian logician does in these situations is similar to what a philosopher does when judging the applicability of 'X' in a given context: the logician supposes a certain practice—in this case, the inferential practice—is mostly a rule-governed practice and, on the basis of his or her competence and reflection, concludes that is reasonable to conceive that particular inference as conforming (or not) to that rule.

Hence, the philosopher/the logician seems to be guided by the following idea: there are norms for the use of 'X'/norms of reasoning and, aside from errors and idiosyncrasies, it makes sense to think that these norms govern a significant part of our uses/of the inferences we make.

In conclusion, concerning the analogy between philosophers' intuitions and PP statements (that is, statements having a prescriptive role but a descriptive content), I pointed out that an intuitive judgement such as 'this Y would (or wouldn't) be called "X"' does not aim to describe the manner in which the term 'X' would actually be used in relation to a specific case—even if the phrasing of the judgement undoubtedly suggests so. Rather, claiming that this Y would (or wouldn't) be called 'X' amounts to claiming that, in such a case, 'X' should (or shouldn't) be used to refer to Y. So, according to this perspective, the verdict is not the description of a mere regularity, or a prediction on a not-yet-observed use of the term in light of already-observed uses of the term. The verdict is rather the description of a norm/rule, or better, a hypothesis on the correctness of a certain use. The same seems to apply to logicians' judgements rejecting or accepting particular inferences.

In the next sections, I am going to illustrate the difference between a regularity and a rule (between regular and rule-guided behaviour) and the distinction between failing to comply with a practice and violating a norm. I will then explain the reason why some practices—like the linguistic one—are conceived as rule-guided practices and some others are not. Finally, I shall deal with the assumption of normativity about the linguistic practice and its consequences. Regarding this last point, I shall explain how this assumption is put into effect in the conception philosophers have of their goals, in the methodology they adopt and in the characterization of the evidence they appeal to.

6.3 Normative Assumptions, Normative Aims and the Normative Aspect of Intuitions

So, what is the difference between a regular and a rule-guided behaviour? First, 'regular' and 'ruled' are not co-extensive. There are practices that are perfectly regular but not ruled (the motion of planets) and practices that are completely ruled but not regular (drivers' behaviour). Linguistic practice is a roughly regular practice that we tend to conceive as rule-guided. What makes us judge some regular (to some extent) practices as rule governed, and others as not rule governed? Or, coming back to the problem of the ambiguity between the descriptive and the normative aspect in judgements: what is it that makes a description of some praxis the description of a rule, and a description of some other praxis the mere description of a praxis? For example: what does distinguish the praxis of using 'spoon' to denote spoons, from the local Turinese praxis of having dinner between 8 p.m. and 9 p.m.? More precisely, why do we consider the use of 'spoon' to denote cups as violating a norm, whereas we regard having dinner in Turin at 10 p.m. as simply disregarding a custom?

To answer these questions let's consider the following scenario. Let's suppose that we conduct a survey among the community of the chess players and that we notice that, before making a move, every chess player (or the majority of the chess players) scratches their nose. Would we conclude that 'one scratches one's nose before making a move' is the description of a norm, in the same way as a statement like 'the bishop moves diagonally' counts as the description of a norm? Obviously, we would not. However, what does distinguish the praxis of moving the bishop in diagonal from that of scratching one's nose before moving? More specifically: how is it possible that the statement describing the former is seen as a prescription and the statement referring to the latter as a simple description? The consideration that is usually made is this: while we would say that those who do not move the bishop diagonally are not playing chess (our game of chess), we would not say that those who do not scratch their nose before making a move are not playing chess. In other terms: while we conceive moving bishops in certain manners as constitutive of playing chess, we do not conceive scratching one's nose as such.

Let us then return to the example I introduced at the beginning of this section: why are we committed to using 'spoon' the way English speakers use it, whereas, when in Turin, we are not committed to dine between 8 and 9 p.m.? As in the case of the chess players' community, the idea is this: while we would say that those who do not use 'spoon' in the same way English speakers do are not speaking English, we would not say that those who are having dinner in Turin between 8 p.m. and 9 p.m. are not having dinner in Turin. Namely, using an English term in a certain way (in the way English people use it) is constitutive of speaking English. On the contrary, having dinner in Turin between 8 p.m. and 9 p.m. is not constitutive of having dinner in Turin.

The idea just stated could also be expressed by conditionals like 'if you want to play chess, then move the bishop so and so'; 'if you want to speak English, then use "spoon" to denote spoons (or, in general, as English people do)', and so on. This is appropriate, provided that conditionals of this kind are not assimilated to conditionals like 'if you want to get from Turin to Milan in less than one hour, then take the *Freccia Rossa* train', that is, conditionals that convey technical or instrumental norms, which only count if one has certain desires or ends. Differently from 'if you want to get from Turin to Milan in less than one hour, then take the *Freccia Rossa* train', a conditional like 'if you want to speak English, then use "spoon" to denote spoons' does not convey a technical or instrumental norm, but a constitutive condition: using the term 'spoon' in the same manner English speakers do is not just a means to speak English (it is not a means to an end); it is constitutive of it.[1]

[1] To deepen the discussion on the normativity of meaning see Glüer and Wikforss (2015).

Apropos, the parallel of Hattiangadi (2006) between conditionals expressing semantic norms and conditionals expressing moral norms (she calls them categorical prescriptions) is especially fitting. Let us consider the conditional: 'if you are a moral agent, you ought to maximise happiness' (Hattiangadi 2006, p. 228). Aiming to maximize happiness is not a means to being a moral agent (in contrast with taking a *Freccia Rossa* train, which is a means to get from Turin to Milan in less than one hour), but a constitutive condition of being one. Equally, using 'spoon' to denote spoons is not a means to being an English speaker, but a constitutive condition of being one.

In conclusion: presenting semantic norms as conditional and constitutive norms is a way to describe the normativity of linguistic use and, in general, a means to clarify the difference between (more or less) regular praxes, whose descriptions count just as descriptions of facts, and (more or less) regular praxes whose descriptions (also) count as prescriptions.

Hitherto, I have observed that linguistic usage is a roughly regular practice that we tend to conceive as rule-guided and I clarified in which sense some practices are seen as rule-guided and others are not. In the next section, I will deepen the issue about the relation between regular and rule-governed in regards to language.

In Chap. 5, I dealt with the normativity of usage. There, I addressed the question of whether it makes sense to think of the linguistic practice as a norm-governed practice in the light of errors and idiosyncrasies. The response to this question consisted in pointing out that the idea of a norm-governed practice does not presuppose a complete uniformity of behaviour, but merely requires the existence of a shared assumption of normativity about the practice: of speakers thinking of the linguistic practice as a norm-governed practice (Marconi 1997, Chapter V). This assumption is revealed by a fact: by speaking (or thinking), any of us assumes to conform one's speech (or thought) to the same constraints that others conform to when they intend to speak (or to think) correctly. One of the immediate consequences of this attitude is the convergence of uses: by assuming to speak in the same way as others do, speakers end up using terms in a similar way. According to this view, the relation between regular and rule-governed should be thought of in the opposite direction to that which is usually imagined. It is not because of the (relative) uniformity of our uses (or, in general, behaviour) that we believe our language is a norm-governed structure; it is rather because we conceive, from the beginning, the practice of our language being used in a relatively uniform way to be norm-governed. In other words, the convictions that there are correct criteria for the use of 'X', that one's own behaviour conforms to these criteria, and that these criteria guide, at least broadly, others' behaviour, directs everybody's behaviour towards uniformity.

The normativity of usage can be used to explain the normative aspect of intuitive judgements as well. Judgements of the form: 'given the way we use "X", this Y would (or wouldn't) be called "X"' should not be conceived as mere empirical hypotheses, but rather as hypotheses of correctness about a certain use of 'X'. The reason lies in the normative assumption on usage: given that a philosopher's judgements are the products of his or her capacity to use the term in a manner that tends to reflect the way the term is used in the community, and given that usage is the source of the norm, such verdicts will count as prescriptions (as well as descriptions).

However, in order to account for the conception philosophers have of their goals, for the way they see themselves as proceeding in the attempt to give an explicit formulation of the rule and for their conception of the evidence they appeal to, the assumption of normativity about usage is not the sole aspect that should be considered.

I said that philosophers' goal is the outlining of norms, that is, consistent and accurate frameworks. I also pointed out that, even though the practice is conceived as rule-governed (and, therefore, as the source of the norm), there are irregularities in it: errors and idiosyncrasies, on the one hand, incoherencies and inaccuracies, on the other. Let us concentrate on incoherencies and inaccuracies. In the case in which ordinary usage would be the only source of the norm, using 'X' inconsistently or inaccurately would not necessarily amount to making a mistake (in the sense of making an idiosyncratic or deviant use of the word with respect to other speakers' uses). Namely, an incoherent or inaccurate use of 'X' made by an individual speaker might reflect trends that are actually present in the community. Anyway, let us put ourselves in the philosopher's shoes, that is, from the perspective of one who works to make norms for the use of 'X' explicit. It is then clear that, in this case, inconsistent or inaccurate uses of ‘X’ will be problematic: one looks for norms, that is, consistent and accurate frameworks.

Hence, for a philosopher who wants to use ‘X’ correctly, it will not be enough to abide by the meta-norm ‘comply with usage’. The philosopher should also determine whether uses of the term can be brought back to an accurate and coherent framework. The corrective and enhancing (as well as descriptive) method of reflective equilibrium (RE) offers a strategy for determining that (*see* Chap. 3 and Sect. 5.2.2).

Finally, the idea of the norm as non-reducible to usage is also reflected in the conception philosophers have of the nature of the judgements they appeal to.

Intuitions, as I have claimed, are hypotheses of correctness, not mere hypotheses on usage. My first attempt to explain this was based on a parallel with the PPR and on the thesis of the normativity of usage: judgements appearing to be mere descriptions of a state of affairs should be understood as normative because of the assumption of normativity about usage. However, in the light of the arguments raised in the preceding paragraphs, my claims should be put in a slightly different way.

Let us imagine a philosopher considering a case, asking himself or herself whether it would be legitimate to use a certain term (apply a certain concept) in a certain situation. In the framework I am proposing, the question would be: would it be correct to use the term (apply the concept) in this situation? In a perspective that recognizes both the centrality of the practice and the revisionary work of philosophers, the question would be expressed as: could this use of the term (application of the concept) be explicated by a norm, that is, by a consistent framework accounting for a significant part of our uses (applications)?

It is clear that, to answer this question, the exercise of one’s competence is not enough; a lot of reflection is needed—on the practice in general, and on the case in particular.

Reflecting upon the practice means considering whether a particular use is convenient (or not) in light of the circumstances in which this use is made. Moreover, it requires comparing different cases. According to some philosophers engaged in the debate on the method, such as, for instance, Kauppinen (2007), Jackman (2009), and Casati (2011), a philosopher struggling with a case usually imagines similar or only slightly dissimilar cases to the one he or she is concerned with and tries to make his or her intuitions about all those cases consistent. As Casati (2011) puts it: philosophers reason in a parametric way.

Rather than starting from the 'updated' idea of normativity, this same idea could be stated by referring to philosophers' goals. In the perspective of research aiming to define a consistent and accurate theory on X, the judgements this theory is going to be built on cannot be the results of competence alone. They will also come from reasoning: from some sort of negotiation 'prior' to RE's negotiation phase. In short, my suggestion is this: even before passing to the phase of the RE process in which one compares and negotiates among competing instances, the philosopher expresses his or her judgement on the applicability of the term (of the concept) on the basis of his or her competence, of the comparison of the case with other case and, also, of general criteria (as those relative to the conditions under which a proper exercise of the competence can be done).

Hence, there are two aspects that one should take into account: the level of acquaintance with the practice of the person who expresses the judgements and one's capacity to reflect upon such practice. Now, the skill that enables the philosopher to become acquainted with the practice could be described as 'an anthropologist's skill': it is a matter of observing the practice, of having the broadest possible overview of usage. By contrast, the other skill (that of reflecting on the practice) appears to be 'typically philosophical', that is, acquired by training or, more generally, by exposure to the *modus operandi* that is typical of the discipline.

The point raised in the preceding paragraphs will be central in my discussion of experimental philosophers' theses (Sect. 8.3). Namely, though it is plausible that non-philosophers (for instance, a sizeable proportion of the subjects interviewed by the experimental philosophers) are sufficiently competent in respect to the usage of several terms (such as 'knowledge'), it is unlikely that non-philosophers will undertake a reflection of the sort we have just presented if asked to evaluate a philosophical TE. The reason is simply that non-philosophers are generally not engaged in the kind of project philosophers are engaged in— the enterprise of outlining norms.

As Jackman points out in 'Semantic intuitions, conceptual analysis, and cross-cultural variation':

> The preference for philosophers' intuitions is not because philosophers are especially good at detecting what is already there, but rather because they are the only ones actively concerned with the sort of consistency-driven construction that conceptual analysis involves (Jackman 2009, p.13)

The idea is that philosophers' constant search for accurate frameworks shapes their judgements. Asking themselves whether a use could be explicated by a norm, philosophers know they must reflect on the practice, consider the situation they are evaluating in the light of other cases, negotiate their intuitions to make them consistent, consider the possible consequences of their responses, and so on. This is something the subjects interviewed by experimental philosophers plausibly don't, and can't, do. Therefore, non-philosophers' evaluations of TEs will register, in the best-case scenario, their disposition to use the relevant term (apply the concept) in the light of the way in which, in everyday contexts and with no particular concern about coherence and precision they use the term (apply the concept).

References

Casati, R. 2011. *Prima lezione di filosofia*. Roma-Bari: Laterza.

Glüer, K., and Å. Wikforss. 2015. The Normativity of Meaning and Content. *The Stanford Encyclopedia of Philosophy* (Summer 2015 Edition), ed. Edward N. Zalta, URL = http://plato.stanford.edu/archives/sum2015/entries/meaning-normativity/. Date accessed 28 Dec 2015.

Goodman, N. 1955. *Fact, Fiction, and Forecast*. MA: Harvard University Press.

Hattiangadi, A. 2006. Is Meaning Normative? *Mind & Language* 21(2): 220–240.

Jackman, H. 2009. Semantic intuitions, conceptual analysis, and cross-cultural variation. *Philosophical Studies* 146(2): 159–177.

Kauppinen, A. 2007. The Rise and Fall of Experimental Philosophy. *Philosophical Explorations* 10: 95–118.

Marconi, D. 1997. *Lexical Competence*. Cambridge, MA: MIT press.

Millikan, R. 1995. Pushmi-pullyu representations. *Philosophical Perspectives* 9: 185–200.

Part III

Experimental Philosophy

7

Empirical Inquiries Versus Armchair Investigations

In the previous two chapters I dealt with the question of whether—and how—a method largely based on intuitions could lead to the kind of results philosophers wish. The problem is, ultimately, a problem of compatibility between the nature of the aims and results of the philosophical inquiry, and the nature of the evidence philosophers appeal to. Here is the question which describes, in a more detailed manner, the dilemma: why—and how—should intuitions (what we would say, or be inclined to say about X in specific circumstances), be relevant in order to construct, justify or attack theories aiming to establish what X really is (as different from or exceeding what we say or think that X is)?

In Chap. 5, I wondered about the way in which it would have been more appropriate to conceive this goal: what does looking for a theory on X amount to? Does it amount to the search of necessary truths about the entity X, as Williamson claims? Or to the search of coherent, accurate and intersubjectively binding norms, which can be used to discriminate between correct uses of 'X' and only seemingly correct uses of 'X'? I claimed that, if one wants to hold the most obvious and acknowledged view on the nature of intuitive judgements, then the latter way is preferable.

© The Author(s) 2016
S.M. Nicoli, *The Role of Intuitions in Philosophical Methodology*,
DOI 10.1057/978-1-137-56715-4_7

At the end of Chap. 5 and at the beginning of Chap. 6, I explained in which sense this alternative is viable, that is, I showed how an inquiry starting from our intuitions can indeed lead to norms. Apropos, I claimed that two aspects should be taken into account: (1) the structure of the method; and (2) the normative aspect of the intuitive judgements.

Concerning the second topic, I argued that judgements philosophers express are neither mere registrations of what philosophers feel inclined to say in specific cases, nor empirical hypotheses on what speakers say or would say in those cases. Rather, intuitive judgements are the expression of what philosophers think is correct to say in the light of their competence and of their capacity to reflect. The point is that intuitive judgements have a normative aspect that is not immediately evident which allows them to legitimately perform their role in the process of theory-building and justification. I also claimed that this argument would have been useful to counter experimental philosophers' criticisms of the traditional methodology of philosophy.

Hence, in this chapter, I shall address experimental philosophers' theses. In Chap. 8, I shall explain how, through our characterization of the nature and the role intuitive judgements play in reflective dynamics, one can take on their challenges to the armchair methodology of philosophy.

7.1 Experimental Philosophy: Who, How, When and Why

Over the last fifteen years, experimental philosophers have been among the most influential participants in the debate on philosophical methodology: they have questioned the traditional armchair methodology of philosophy and presented their project as an alternative to—or, at least, as an essential integration of—the work of armchair philosophers.

'Normativity and epistemic intuitions' (2001), a paper by Weinberg, Nichols, and Stich (WNS), is generally taken as the manifesto of the movement in its contemporary form. Although experimental philosophy is quite a recent creation, the sceptical questions it is founded on were elaborated two decades earlier by one of the fathers of the movement:

Steven Stich. In the 1980s, Stich wrote two articles (Stich and Nisbett 1980; Stich 1988) in which he criticized the methodological proposal of reflective equilibrium (RE), as described by Goodman (1955) and Rawls (1951 and 1971). His focus was already on the risk of intuitions' (transcultural) variability. However, I will not take into account Stich's articles from that period, rather, I will discuss and criticize experimental philosophers' theses as they have been presented in the last fifteen years, that is, since the manifesto of 2001.

First, it is worth underlining that, although experimental philosophers sometimes hint at the question debated in the previous two chapters (whether what we would say could be legitimately used in order to build, justify or attack theories that are not descriptive of our way of judging), their main interest does not lie here, but 'at the roots' of the appeal to intuitions. On closer inspection, X-Philers maintain, judgements philosophers express at the end of thought experiments (TEs) are empirical hypotheses on the way in which people would respond to cases (apply a concept or use a term, judge a moral scenario, and so on). Intuitions are, therefore, empirically testable. But, though being testable, they have never been tested: no philosopher has ever taken upon himself or herself the task of checking whether matters effectively stand in the way they claim, that is, whether their judgements actually are the expression of what people would actually say about such and such cases.

Experimental philosophers focus on the responses to TEs. However, the problem they raise does not concern exclusively the appeal to the verdicts philosophers express when assessing TE scenarios, but intuitions of the folk in general. Are the intuitive judgements philosophers appeal to the expression of what we—all of us, philosophers and non—would respond to cases (would apply a concept or use a term, judge a moral scenario, and so on)?

In short, experimental philosophers' attack to armchair philosophy is the following: the appeal to intuitions could be legitimate only if intuitions did in fact describe people's responses to cases. But, this cannot be assessed 'from the armchair'. What one should do is check, empirically, whether people's responses to cases match what philosophers predict they would be.

> For 2500 years, philosophers have been relying on appeals to intuition. *But the plausibility of this entire tradition rests on an unsubstantiated, and until recently unacknowledged, empirical hypothesis*—the hypothesis that the philosophical intuitions of people in different groups do not disagree. Those philosophers who rely on intuition are betting that the hypothesis is true. If they lose their bet […], then a great deal of what goes on in contemporary philosophy, and a great deal of what has gone on in the past, belongs in the rubbish bin. (Stich 2009, p. 232)

A parenthetical remark: I said that the experimentalists' criticism is 'at the roots' of the question whether judgements of intuitive nature could legitimately be used to build and justify philosophical theories. This fact is strictly connected to another. Experimental philosophers neither take any particular stance about the problem of compatibility between the nature of intuitions and that of the aims of the philosophical inquiry, nor declare a leaning toward a certain description of philosophical goals rather than to any other: given their main concerns, they do not need to do so.

Furthermore, in the last decades, anyone who has been interested in methodological questions, and resolved to argue in favour of the legitimacy of philosophers' work, felt 'forced' to take a stand against experimental philosophers' theses. Everyone did it, regardless of his or her position on the nature of the philosophical aims and results.

Two things can be gathered from these facts: first, in spite of disagreeing on what practising philosophy amounts to (do philosophers make metaphysic in a classical sense? Revisionary metaphysics à la Strawson? Analysis of language?), the majority of philosophers seem to agree on the way in which philosophy is conducted. Second, the characterization that X-Philers give of intuitive judgements seems to be agreed upon.

A couple of remarks about these last two claims. As a matter of fact, there is no actual agreement on the way in which philosophy is performed. General methodological questions are little discussed by experimental philosophers: the debate is focused mainly on the use of intuitions in TEs, and when the broader dynamics in which intuitions are involved must be explained, the parties involved limit themselves to extremely concise remarks. These remarks are similar to those made by WNS in their manifesto: ‘an IDR [Intuition Driven Romanticism] strategy can be viewed as a “black box” which takes intuitions (and perhaps other data) as input and produces implicitly or explicitly normative claims as output’ (Weinberg et al. 2001, p. 8); or similar to those made by Buckwalter and Stich (BS): ‘Philosophical theories that are compatible with the content of the intuition are supported, and philosophical theories that are incompatible with the content of the intuition are challenged’ (2014, p. 311). On this very general idea—on a version of the method in which no mention of the inner mechanisms of the ‘black box’ is made—there seems to be an agreement.

A similar question can be raised apropos of the second aspect. By saying that experimentalists’ characterization of the judgements is more or less agreed upon, I am not claiming that characterizations of intuitions that critics of experimental philosophy offer are in fact identical to those offered by X-Philers. What I want to say is that many supporters of armchair philosophy take the idea at the basis of the experimentalists’ project as true: they agree that the reliability of intuitions depends on the fact that they are widely shared or, at least, shareable.

In the following sections, I shall explain in detail why experimental philosophers believe that empirical research is needed. In order to do that, I will start by pointing out that X-Philers’ general remarks could be countered by appealing to an argument that was advanced by Jackson in *From Metaphysics to Ethics*: *A Defence of Conceptual Analysis*.

My responses—Jackson claimed—are reliable (that is, they reveal the folk conception) ‘in as much as I am reasonably entitled, as I usually am, to regard myself as typical’ (Jackson 1998, p. 32). According to this thesis, verifying the effective diffusion of philosophers’ intuition is useless.

Jackson’s argument recalls one of the considerations I advanced in Chap. 5, when, discussing the reliability of the judgements on X that philosophers express at the end of TEs, I argued that those judgements are reliable in so far as philosophers are competent, that is, able to use ‘X’ (to apply concept *X*) in a way that tends to reflect the way it is used (applied) in the community. In other terms: the fact that a philosopher usually uses ‘X’ in a way that actually matches the way it is used in the community guarantees his or her assessment about the single case.

Experimental philosophers could reply in the following way: Jackson has many good reasons for regarding himself as typical; in general, any philosopher, being an adult and educated member of his or her community, is justified in judging himself or herself competent. Hence, it is perfectly plausible that their judgements actually are grounded hypotheses on how people will respond to the case. However, why not check whether it is really so, that is, whether people actually judge in the way philosophers suppose they would? After all, empirical research could substantiate philosophers' hypotheses. In particular, adding the work of the experimental philosopher to that of the traditional philosopher could help philosophy, could offer philosophers a valuable chance to further test their theories.

It is worth pointing out that the remark I have just made—a 'prudential' remark—is not the kind of remark experimentalists usually advance when arguing for the suitability of their project. On the contrary, they refuse arguments *à la* Jackson and claim that integrating (or even replacing) the armchair with experimental work is necessary. In other terms: according to the experimental philosophers, one does not check what people say just for doubt's sake, but because there are reasons to suspect that philosophers' judgements do not match those of non-philosophers at all.

This idea could be stated in two ways: the first is the way in which Pust and Goldman present it; the second is that of the experimentalists.

Let us start with Pust and Goldman. In 'Philosophical theory and intuitional evidence', they introduce the problem of theory-ladenness as regards philosophers' intuitions: 'a second possible source of error is theory contamination. If the person experiencing the intuition is a philosophical analyst who holds an explicit theory about the nature of F, this theory might warp her intuitions about specific cases' (Pust and Goldman 1998, p. 183). This is the reason why one should prefer 'informants who can provide pre-theoretical intuitions about the targets of philosophical analysis, rather than informants who have a theoretical "stake" or "axe to grind"' (Pust and Goldman 1998, p. 183).

Unlike X-Philers, Pust and Goldman support the use of intuitions in philosophy. Furthermore, they present the philosophical project as a descriptive enterprise: philosophers describe concepts and concepts are psychological entities. In this respect, their position is distant from mine and from that of X-Philers. Nevertheless, Goldman and Pust's objection can be examined also in relation to our problem: it is possible that judgements philosophers present as neutral (as descriptions of what speakers would say when presented with the cases) are the upshot of some theory they embrace. Anyone who engages in a process leading him or her to embrace certain theses could come to conceive as neutral judgements which, as a matter of fact, are the upshot of their theses. According to this perspective, philosophical training is not an added value, but rather an obstacle to the acquisition of the relevant data.

Let us now consider the reasons experimental philosophers adduce to motivate their scepticism towards philosophers' judgements, and to argue, thereby, for the necessity of empirical surveys. Experimental philosophers advance the following hypotheses.

First (Hypothesis A) it is possible that intuitions philosophers present as 'correct', that is, as obvious and widely shared, are the expression of the way of thinking (or speaking) of a very limited portion of the population (and in particular of that portion of the population philosophers belong to).

Moreover (Hypothesis B) it is possible that intuitions vary in connection with a series of philosophically irrelevant factors, as, for instance, culture, gender, socio-economic status, level of education of the person who assesses the case. In addition to the characteristics of the subject (the so-called demographic factors), intuitions could vary accordingly to the characteristics of the experiment, that is, according to the order in which the cases are presented, to the negative or positive framing of the single cases, to the environmental conditions in which the subjects are while assessing the case (order effects, framing effects, environmental effects).

Hence, what does the empirical inquiry X-Philers have in mind consist in? Although the literature on the surveys they have designed is pretty large, their way of proceeding could be summarized as follows. Experimental philosophers take into account a series of well-known TEs and present a version of them to groups of non-experts. After having presented the story, experimental philosophers do not report the expected outcome (let us name it '*H*'), but rather a question which can be answered with '*H*' or 'non-*H*'. X-Philers generally don't give further options and don't ask subjects to motivate their choice. In some cases, they pose another question in order to verify that subjects have really understood the case. Secondly, experimental philosophers select the sample of subjects they interview in a way that should help them check whether the differences eventually occurring in people's responses are somehow connected to their degree of education, cultural background, gender, and so on. In order to do so, they take two groups of people that differ in the aspect that they want to investigate (gender, culture, and so on), and check whether the reactions of the subjects belonging to one group differ from those of the people belonging to the other. In other cases, experimental philosophers do not work on the characteristics of the subjects, but rather on those of the experiment. In this circumstance, two different versions of the same case are designed (for example, one version in which the case is positively framed and another in which is negatively framed). One version is presented to half the subjects of the sample and the other one to the other half. Finally, it is checked whether there are significant differences between the answers of the subjects of the first group and those of the second group.

The presentation I have given is extremely general. I will not enter here into the merits of all the different tests experimental philosophers have designed, but I will illustrate one of them as an example. Here is the description of the Gödel–Schmidt case (Kripke 1980) given by Mallon, Machery, Nichols, and Stich (MMNS) in 'Semantics, cross-cultural style'.

> Suppose that John has learned in college that Gödel is the man who proved an important mathematical theorem, called the incompleteness of arithmetic. John is quite good at mathematics and he can give an accurate statement of the incompleteness theorem, which he attributes to Gödel as the discoverer. But this is the only thing that he has heard about Gödel. Now suppose that Gödel was not the author of this theorem. A man called 'Schmidt,' whose body was found in Vienna under mysterious circumstances many years ago, actually did the work in question. His friend Gödel somehow got hold of the manuscript and claimed credit for the work, which was thereafter attributed to Gödel. Thus he has been known as the man who proved the incompleteness of arithmetic. Most people who have heard the name 'Gödel' are like John; the claim that Gödel discovered the incompleteness theorem is the only thing they have ever heard about Gödel. When John uses the name 'Gödel', is he talking about:

(a) the person who really discovered the incompleteness of arithmetic? or
(b) the person who got hold of the manuscript and claimed credit for the work? (Macherie et al. 2004, pp. 6–7)

As in the TEs in general, a mismatch is shown between the way in which we judge and the way in which a theory would require us to judge. According to the descriptivist theory of meaning for proper names, a name, *N*, as used by a certain speaker, *S*, refers to the object/person which is the denotation of all, most, some of the definite descriptions *S* associates to *N*. Hence, if descriptivism were true, in the Gödel–Schmidt case, 'Gödel' would refer to the denotation of the definite descriptions (in this case of the unique definitive description: 'the man who discovered the incompleteness of arithmetic') most of the people associate to the name 'Gödel', that is, Schmidt. However, Kripke said, it isn't so, or better, it doesn't seem correct to us to say it is so. What we would say is that 'Gödel' refers to Gödel. Descriptivism is challenged.

MMNS presented the case to a group of undergraduate students from Rutgers University (31 people) and to another group of undergraduates from the Hong Kong University (42 people). As the choice of the sample suggests, there were two hypotheses X-Philers wanted to test: whether non-experts' intuitions diverge from experts' and whether Westerners' and East Asians' intuitions about reference differ. In other terms, the question MMNS addressed was: is Kripke's intuition (b) typical of our cultural context solely, or not?

What emerged from their survey? First, a significant number of the subjects did not answer in the manner Kripke predicted. More specifically, whereas a great proportion of the subjects with a Western cultural background (56.5 per cent) answered in the way in which the great majority of Kripke's colleagues answered (b), the majority of the subjects with an Asian cultural background turned out to be inclined towards the opposite option (only 31.5 per cent of them chose (b)).

Beyond this specific survey, analogous results were obtained in all other experimentalists' studies: a complete convergence on the verdicts philosophers express at the end of TEs never occurred. In particular, in each case, a significant percentage of the subjects did not answer in the way philosophers do. Furthermore, it turned out that there are factors (characteristics of the subjects or of the experimental setting) affecting the trend of the answers. For instance, cultural background seemed also to influence people's intuitions relative to the Gettier cases (Weinberg et al. 2001). Other studies showed that intuitions are sensitive to people's socio-economic status (Haidt et al. 1993) or to their emotional state (Wheatley and Haidt 2005), or to the order the cases are presented (Swain et al. 2008).

What conclusions do X-Philers draw from these data? In general, X-Philers take their data to prove that their work is needed: traditional philosophers make improper generalizations from their perspective. Their activity cannot do without the work of experimental philosophers.

Some experimental philosophers (Stich 2009) also take their data as a demonstration of the bankruptcy of the armchair methodology: if it is true that philosophers build, support, and attack theories on the basis of intuitions, then variability of intuitive judgements (a pervading and wide-ranging fact) shows how an intuition-based activity can only give rise to partial and relative theories. Take the example above, that is, a TE which elicits different intuitions about reference in different subjects: some people (mainly Westerners and few East Asians) have an intuition which is consistent with the causal-historical theory of reference, other people (mainly East Asians and a few Westerners) have the opposite intuition, that is, the intuition agreeing with the descriptivist view of reference. Now, if one holds the idea according to which a theory is built, supported or challenged on the basis of intuitions, then, in the light of the variability of intuitions, one should recognize the futility of the appeal to them. In this case, for instance, one cannot conclude what Kripke claims to have established: that descriptivism is false and the causal-historical alternative should be preferred. This is true for all intuitions and theories: in the light of the disagreement on intuitions, no choice among competing theories can be made.

Finally, data do not show a generic kind of variability. There are factors in correspondence to which intuitions vary systematically (in our example, culture is the determinant factor). But, if this is true, then either one abandons the idea according to which the agreement with intuitions determines the correctness or the incorrectness of theoretical options, or else, if one maintains it, then one is committed to accept that theories are group-specific (in our example, culture-specific).

Yet, (cultural) relativism is not the worst consequence. The fact that philosophical intuitions (about reference, meaning, causation, moral right and wrong, and so on) change in connection with the change of philosophically irrelevant factors (culture, gender, and so on) reveals the dependence of these intuitions on these factors: having an intuition that *H*, rejecting a theory which contradicts *H* and embracing one which is coherent with *H*, seems to be determined by the mere fact of belonging to a certain group of people having certain characteristics, or by some characteristics of the experimental setting. But, if this is true—if intuitions are the result of these accidents—then it would seem that any decision in favour of a theory, rather than another, would be arbitrary.

I have said that the variations experimental philosophers register are wide-ranging: they investigate different fields of philosophical inquiry and depend on different factors. In particular, as Buckwalter and Stich (2011 and 2014) have recently pointed out, philosophical theses, and the intuitions supporting them, seem to be the result of the mindset of a community ('the philosophy club') which is formed by people having specific and clearly recognizable characteristics: Westerners, males and highly educated. In fact, whereas Westerners', males' and highly educated people's intuitions generally match philosophers' intuitions, non-Westerners', females' and poorly educated people's intuitions tend to diverge from them. The most general consequence has already been stated: theories systematizing philosophers' intuitions cannot fulfil the aspirations of generality which philosophers have about their theories. But another consequence could be that these intuitions and theories end up constituting an unintentional filter by means of which people having certain characteristics (the same characteristics as philosophers of previous generations) are selected as new philosophers, while persons whose spontaneous reactions diverge from those of established philosophers are excluded from the games. Obviously, this would happen neither patently nor purposely, but by means of a misguided methodology based on intuitions.

7.2 Experimental Philosophers: Empirical Branch of Armchair Philosophers or 'Saboteurs'?

I said that experimental philosophers regard their data as supporting the importance of carrying out empirical investigations, and that some of them also see such data as demonstration of the bankruptcy of the traditional methodology. Let us linger on this topic for a little longer.

First it is worth pointing out that experimental philosophers have not had (still do not seem to have) a completely straightforward attitude towards the armchair methodology of philosophy and, consequently, an unequivocal idea of their own role as philosophers investigating the actual agreement on intuitions. According to experimental philosophers' initial declarations of intent, the empirical methodology they proposed should have merely represented an important integration of the armchair one: experimental philosophers would have checked which, among traditional philosophers' judgements, were well-confirmed hypotheses and which were not. Traditional philosophers would then have accepted experimental philosophers' results and limited their work to the well-confirmed ones. However, experimental philosophers' attitude towards armchair philosophy appears to have turned out to be much less accommodating than the one I have just described: they seem to have ended up claiming that, in light of an ascertained and wide-ranging disagreement on intuitions, armchair philosophers' theses and methodology should be abandoned.

An interesting article analysing the different approaches experimental philosophers have to their own project is 'The past and future of experimental philosophy' (2007). Here Nadelhoffer and Nahmias distinguish three different branches: Experimental Descriptivism (ED), Experimental Analysis (EA) and Experimental Restrictionism (ER). ED is a psychological project: the focus is on philosophical intuitions, but the goal is to explore human psychology, that is, the psychological processes and cognitive mechanisms generating those judgements. EA and ER are described as philosophical projects. The goal of EA is to use empirical data to support philosophers in their work: experimental philosophers determine which intuitions are widely shared; traditional philosophers should then use these judgements to build (or privilege) theories enjoying 'squatters' rights'. On the contrary, ER uses empirical data to criticize philosophers' work. The idea is that, since intuitions are unreliable, they cannot be used to build and to justify philosophical theories.

By showing that experimental philosophy is not a monolithic project, Nadelhoffer and Nahmias aim to counter some objections to the experimental project (Kauppinen 2007). I will not analyse these objections and Nadelhoffer and Nahmias' arguments, but I will just focus on the plausibility of the distinction between EA and ER.

The distinction between EA and ER is definitely grounded in the declaration of intents of the different members of the experimental movement. However, in light of the data experimental philosophers have collected, what is to be said about the 'destiny' of EA? If one takes those data to be reliable, than it is hard to see how one could claim they could be used to support philosophers' work. Namely, even in those cases in which there is a statistically significant majority responding in one way, there is still a nontrivial minority responding in other ways. How is a traditional philosopher supposed to treat this disagreement? Why should he or she privilege someone's intuitions to someone else's? Or, why should he or she privilege the intuitions of the majority over those of the minority? As Nado points out in 'Intuition, philosophical theorizing, and the threat of skepticism', in light of an ascertained and wide-ranging disagreement on intuitions, 'the skeptical interpretation seems to be the default one' (Nado 2015, p. 206). In conclusion, experimental philosophy, allegedly the empirical branch of traditional philosophy, in fact undermines philosophical armchair investigations.

I shall return to this problem in the first section of the next chapter, when I will explain why the experimental philosophers' project can neither be regarded as integrative, nor as substitutive of that of armchair philosophers. In the last part of this section, I will consider a fourth possible way to characterize the experimental philosophers' project and to describe its relation to armchair philosophy. According to this view, experimental philosophy can be considered a naturalistic project and, more importantly, an attempt to naturalize our discipline.

This thesis is defended by Bishop in 'Reflections on cognitive and epistemic diversity: Can a Stich in time save Quine?' (2009). In his article, Bishop acknowledges that Stich in 'Naturalizing epistemology: Quine, Simon and the prospects for pragmatism' (1993) explicitly rejects the idea that a normative discipline—like epistemology or philosophy—can be reduced to a descriptive discipline. However, Bishop argues that Stich's and experimental philosophers' empirical arguments against analytic philosophy do in fact 'vindicate Quine's moribund naturalism' (Bishop 2009, p. 133).

When Bishop speaks of 'Quine's moribund naturalism', he refers to the idea defended by Quine in 'Epistemology naturalized' (but repudiated by the majority of the epistemologists and later rejected by Quine himself) according to which 'epistemology, or something like it, simply falls into place as a chapter of psychology and hence of natural science' (Quine 1969, p. 82–83).

How do experimental philosophers' empirical arguments against analytic epistemology (and philosophy) vindicate Quinean naturalism? First, by claiming that philosophers need to 'use experimental methods to figure out what people really think about hypothetical cases' (Knobe 2004, p. 37), experimental philosophers assume that philosophical investigations (should) consist, for a large part, in ethnographical (or social psychological) researches, and

> […] to identify the core of analytic epistemology with ethnography is to suggest that it is empirical. Of course, analytic epistemology is not entirely empirical. Analytic philosophers extract lots of normative, epistemological claims from their descriptive, ethnographic theories. If this is right, then the method of contemporary analytic epistemology has been broadly Quinean all along. (Bishop 2009, p. 119)

Furthermore, experimental philosophers describe their own project as philosophical. But since the main goal of experimental philosophy is to explore the way in which people would respond to cases—that is, to collect empirical data—then experimental philosophers can be considered, for all intents and purposes, Quinean naturalists.

References

Bishop, M.A. 2009. Reflections on cognitive and epistemic diversity: Can a Stich in time save Quine? In *Stich and His Critics*, eds. D. Murphy and M.A. Bishop. Wiley-Blackwell: Malden.

Buckwalter, W., and S. Stich. 2011. Gender and the Philosophy Club. *The Philosophers' Magazine* 52: 60–65.

Buckwalter, S. Stich. 2014. Gender and Philosophical Intuition. In *Experimental Philosophy*, vol 2, eds. J. Knobe and Shaun Nichols. Oxford: Oxford University Press.

A. Goldman, J.Pust (1998) 'Philosophical Theory and Intuitional Evidence' in M. DePaul and W. Ramsey (eds.), 1998

Goodman, N. 1955. *Fact, Fiction, and Forecast*. MA: Harvard University Press.

Haidt, J., S. Koller, and M. Dias. 1993. Affect, culture, and morality, or is it wrong to eat your dog? *Journal of Personality and Social Psychology* 65: 613–628.

Jackson, F. 1998. *From Metaphysics to Ethics: A Defense of Conceptual Analysis*. Oxford: Oxford University Press.

Kauppinen, A. 2007. The Rise and Fall of Experimental Philosophy. *Philosophical Explorations* 10: 95–118.

Knobe, J. 2004. What is Experimental Philosophy? *The Philosophers' Magazine* 28: 37–39.

Kripke, S. 1980. *Naming and Necessity*. Cambridge, MA: Harvard University Press.

Macherie, E., R. Mallon, S. Nichols, and S. Stich. 2004. Semantics, cross-cultural style. *Cognition* 92: B1–B12.

Nadelhoffer, T., and E. Nahmias. 2007. The Past and Future of Experimental Philosophy. *Philosophical Explorations* 10(2): 123–114.

Nado, J. 2015. Intuition, philosophical theorizing, and the threat of scepticism. In *Experimental Philosophy, Rationalism, and Naturalism: Rethinking Philosophical Method*, eds. E. Fisher and J. Collins. London: Routledge.

Quine W.V.O. 1969. Natural Kinds. In *Ontological Relativity & Other Essays.* New York: Columbia University Press.

Rawls, J. 1951. Outline of a Decision Procedure for Ethics. *Philosophical Review*, 60(2): 177–97; reprinted in Rawls, J. 1999. *Collected Papers*, 1–19. Cambridge, MA: Harvard University Press.

Rawls, J 1971. *A Theory of Justice*. Cambridge, MA: Harvard University Press; edition used: Rawls, J. 1972. *A Theory of Justice*. Oxford: Clarendon Press.

Stich, S., and R.E. Nisbett. 1980. Justification and the psychology of human reasoning. *Philosophy of Science* 47: 188–202.

Stich, S. 1988. Reflective equilibrium, analytic epistemology and the problem of cognitive diversity. *Synthese* 74: 391–413.

Stich, S 2009. Reply to Sosa. In *Stich and His Critics*, eds. D. Murphy and M.A. Bishop. Malden: Wiley-Blackwell.

Swain, S., J. Alexander, and J. Weinberg. 2008. The instability of philosophical intuitions: running hot and cold on Truetemp. *Philosophy and Phenomenological Research* 76: 138–155.

Wheatley, T., and J. Haidt. 2005. Hypnotically induced disgust makes moral judgments more severe. *Psychological Science* 16: 780–784.

Weinberg, J., S. Nichols, and S. Stich. 2001. Normativity and Epistemic Intuitions. *Philosophical Topics* 29: 429–460.

8

Against Experimental Philosophy

8.1 Experimental Philosophy: A Non-philosophical Project (Part I)

Experimental philosophers' sceptical conclusions—explicitly acknowledged (Experimental Restrictionism) or not (Experimental Analysis)—are the starting point from which I will develop the first of my criticisms of their project.

According to experimental philosophers, for the general reasons illustrated earlier: group-specificity and arbitrariness of the theories built on intuitions, and for the (political) reasons connected to the systematic exclusions of certain categories from the philosophical community (Sect. 7.1), the theories philosophers elaborated, together with the intuitions supporting them, should be abandoned. Hence, my question is: in favour of what should these theses and intuitions be abandoned? Maybe in favour of the theories which could be build starting from the intuitions of people who are generally left out of the games (women, East Asians, and so on)? Obviously not. If one takes the data that experimental philosophers have collected to be reliable, then it seems that an intersubjectively acceptable theory (a theory enjoying the universality philosophers wish) could not be constructed in any case.

© The Author(s) 2016

S.M. Nicoli, *The Role of Intuitions in Philosophical Methodology*,
DOI 10.1057/978-1-137-56715-4_8

At this point, one could claim that philosophy should be a relativistic enterprise, in which anyone, depending on one's gender, social status, and so on (depending on his or her intuitions) builds or embraces the most congenial theory for him or her. However, this does not seem to be the X-Philers' idea: what they suggest should be abandoned are not primarily the theses and intuitions supporting them, but rather traditional philosophers' activity itself. In other words: what belongs to the 'rubbish bin' (Stich 2009) are not only the results philosophers achieved, but the philosophical enterprise as it has been conceived up to now.

Hence, the question is: how should philosophy be conceived? And, more importantly: could it be conceived otherwise? Even though experimental philosophers frequently ask for a revision of a methodology having the defect of being largely based on intuitions, they do not give any precise indication of what this revision should consist of: shall we restrict our appeal to intuitions to just some of them? If so, how can we determine which are the intuitions that we can legitimately appeal to? It would seem that any decision in favour of the intuition of a certain group, rather than of that of other, would be arbitrary. Hence, should we abolish the appeal to intuitions entirely? If so, what would be the alternative? None of these questions finds a clear answer in X-Philers' texts.

The reply to the last claim may be the following. It is not true that experimentalists only make generic admonishments in favour of a revision of the actual methodology. On the contrary, they propose an alternative to armchair philosophy. They also show us, in practice, what this consists in: experimental philosophy itself, together with its empirical methodology, is the alternative that we need. This move, however, is precluded: if one acknowledges X-Philers' conclusions, then one cannot think of experimental philosophy as a possible substitute for traditional philosophy. In which way could experimental methodology replace armchair methodology, given that X-Philers collect people's intuitions, register important and systematic differences among people's reactions to cases, and conclude that theories founded on them cannot fulfil the aspirations of generality that philosophers have about their theories? If anything, experimental methodology could have represented an important integration of the armchair one. However, in order to do so, it had to have maintained the aspect of neutrality it originally had. But, as I pointed out in Sect. 7.2, data showing widespread disagreement and variability emerged from experimental philosophers' studies and the sceptical aspect of the project prevailed.

All these considerations count as objections to experimental philosophy in so far as it is presented as an alternative, a replacement, or an integration of armchair philosophy, or, more generally, in so far as it is described as a philosophical project. But, obviously, they cannot be considered a reply to the challenges X-Philers advanced to armchair methodology: they do not constitute an answer to the problem of the variability of intuitions. Furthermore, they do not disqualify experimental philosophy as a non-philosophical, that is, psychological, project.

Henceforth, I shall start defending traditional philosophy from X-Philers' criticisms. My strategy will develop along two fronts: a critique of the methodology of experimental philosophers, and a critique of the idea that motivates their project.

In Sect. 8.2, I shall introduce arguments in support of the first line of argument. The thesis I will endorse is the following: experimental philosophers' inquiries should not be considered conclusive evidence for any theory about how Westerners and East Asians, women and men, and so on, tend to judge philosophically relevant cases. Two conclusions in particular cannot be drawn: (1) that the way people judge depends on/ is shaped by their gender, culture, and so on; and (2) that philosophy is a prerogative of groups of persons having specific characteristics ('the Philosophy club'). Moreover, I cast some doubts about the possibility of qualifying these studies as really informative surveys.

In Sect. 8.3, I shall present the arguments in support of the second line of argument. In particular, I will claim that the kind of inquiries experimental philosophers intend to do are irrelevant for philosophy: finding out that non-philosophers' reactions to the cases differ from philosophers' ones neither proves the unreliability of the judgements of the latter, nor the wrongness of philosophers' theses. That is to say, the very idea that motivates the experimental project—the idea that in order to ascertain the validity of philosophers' judgements, one should check people's reactions to the cases described in TEs—is misleading.

As can be expected, my strategy will not consist in questioning the assumption that intuitions play a central role in philosophers' work—as Cappelen in *Philosophy Without Intuitions* (2012) and Deutsch in 'Experimental philosophy and the theory of reference' (2009) did—consequently understating the importance of experimentalists' surveys and the import of intuitional variability on philosophers' work. On the contrary, I will defend my suggestion (Chap. 4) that judgements we appeal to are not empirical hypotheses about the way in which people would assess the case, but rather judgements of correctness that philosophers formulate in the light of their competence and reflection.

In sum, while I will show with the first battery of objections that the results of experimental philosophers' studies do not prove intuitions' variability (neither in a specific sense—cultural variability, gender variability and so on—nor in a generic one), with the second battery of objections I will explain the sense in which the kind of variability X-Philers draw attention to would not pose a threat to the reliability of philosophers' judgements or to the theoretical conclusions they come to.

8.2 Experimental Philosophy: A Non-experimental Project

Are the results obtained by means of the kind of surveys conducted by experimental philosophers reliable? Can these data be used to support the claim that (epistemic, semantic, moral) intuitions depend on factors such as gender and culture, or on other philosophically irrelevant factors such as framing, order, and so on? In particular, can they be used to draw conclusions on the way in which non-experts and experts, Westerners and East Asians, women and men tend to judge, and, hence, the idea that Western academic philosophy, as it has been conceived up to now, is a prerogative of groups of persons having specific characteristics?

In this section, I will argue that the data experimental philosophers have collected could be used for these purposes only if the tests which produced them were real experiments, that is, studies that allowed us to conclude that there is a relation of statistically reliable causality between the independent (gender, culture, socioeconomic status of subjects) and the dependent (subjects' responses to the cases) variables. However, no relationship as such can be established on the basis of experimental philosophers' inquiries. Namely, no 'experiment' where the independent variable is a subjective variable (for instance, gender or culture) can be considered a 'real' experiment, that is, a tool for determining whether there is a relation of statistically reliable causality between the independent and the dependent variables—in the case of experimental philosophers, between gender, culture, socio-economic status of subjects and subjects' responses to the cases.

Specifically, after an opening part where I shall introduce the concepts of experiment, independent variable, dependent variable and internal validity, I will explain why X-Philers' researches cannot be considered experimental, and hence why they cannot provide adequate evidence for any theory about how Westerners and East Asians, women and men, and so on, tend to judge philosophically relevant cases.

8.2.1 Real Experiments and Quasi-experiments

Let's start by introducing an experiment from social psychology: the study conducted by Latané and Darley (LD) on the phenomenon of the *diffusion of responsibility*.

At the end of the 1960s, the two social psychologists elaborated the notion of the diffusion of responsibility after hearing of the murder of a woman that happened in a densely populated area of New York. The attack lasted 45 minutes and none of the residents gave the alarm. Many of them afterwards admitted to have heard a woman screaming and to have rushed to their window. LD wondered what could be the reasons for this behaviour. By focusing their attention on the context in which people witnessed the murder, they elaborated this hypothesis: it is possible that each of the woman's neighbours, knowing that many others were witnessing the emergency and assuming that someone else had called the police, did not felt obliged to help or to give the alarm.

Hence, the observation of a fact allowed the two psychologists to formulate an interesting general hypothesis on people's behaviour in emergency situations: the greater the number of persons witnessing an emergency is, the less likely it is that any of them intervenes.

Once the hypothesis was formulated, the psychologists proceeded to test whether it was effective: whether a causal relationship existed between the number of the bystanders and the helping behaviour. In order to do that, LD used the experimental method. I will not report in detail the preparation and execution of their experiment, but just illustrate its structure in general terms, so that we can understand how the relation of causality between different phenomena can be experimentally established.

LD kept under control the variable (the number of witnesses) that was supposed to be the cause of a certain behaviour (the helping behaviour). Some subjects in the study were exposed to a certain stimulus—that is, believing to witness an emergency (an epileptic attack) alone—and others to another stimulus—that is, believing to witness the emergency in a small group (four persons in all: the victim, the subject and two other witnesses), or in a larger group (six people in all: the victim, the subject and four other witnesses). By changing only one aspect (group size), researchers were able to assess whether this factor was the cause of the problematic behaviour, or not.

In technical jargon, the number of witnesses, that is, the factor which is thought to be the cause of (or to have an effect on) a certain behaviour, is the independent variable; whereas the helping behaviour, that is, the factor on which the independent variable is supposed to cause a change, is the dependent variable. The goal of the experiment—of this and of any other experiment of this type—was to assess whether a relationship of statistically reliable causality existed between these two factors, or better, whether (and to what extent), by changing or varying the independent variable, the dependent one changed or varied in turn.

Did LD confirm their hypothesis? Did the independent variable (the number of the bystanders) affect the dependent variable (the helping behaviour)? LD observed that in the cases in which participants believed that four other people were witnessing the emergency, only 31 % of them helped. When they believed that two other bystanders were there, 62 % of them gave the alarm. Finally, when the participants believed they were the only bystander, 85 % of them helped (*see* Aronson, Wilson, and Akert (AWA), 1997, pp. 46–47).

In light of these results, the answer to the initial question could seem to be obvious: LD confirmed their hypothesis. However, this is not so: data can confirm (or disconfirm) a certain hypothesis, only if the study which produced them was designed in a way which allows us to rule out (a) that other variables are causes of the behaviour we are interested in; (b) that results are accidental, that is, obtained by pure chance, and that in other situations (different subjects, settings, times) they would not be replicated; (c) that there are other theoretical explanations of the data which are more plausible than that we believe is corroborated by the research. In sum, one tries to rule out the presence of the most common threats to internal, external, statistical and construct validity.

Condition (a) is extremely important. Let us take LD's experiment: how can one be sure that differences in the rate of help across the different experimental settings were due to the size of the group of witnesses and not to other factors, such as the different personalities of the participants, their previous experiences as regards emergencies in general, or this particular kind of emergency, and so on? This problem is a problem of internal validity: for an experiment to be (internally) valid, one should make sure that no other factor besides the independent variable (the group size) affects the dependent variable (the modifications in the 'degree' of helping).

As regards LD's study, experimental psychologists agree that it is a valid experiment. Namely, while designing and running the experiment, LD made sure that the procedures were the same in all different conditions: all subjects received the same instructions and the description of the experiment; furthermore, the rooms in which they received the stimulus and the way to convey it were identical, and so on. The only aspect which changed was the aspect experimenters were interested in: the number of subjects taking part to the experiment. Moreover, LD proceeded by randomly assigning the participants to the conditions which they had previously defined: witnessing to the emergency alone, with two other possible helpers and with four other possible helpers.

Let us dwell on this aspect. It is pretty obvious why the control of the conditions in which subjects perform their task is functional to this goal: putting participants in the same environmental conditions and presenting the stimulus to them in the same way enables one to rule out that environmental factors, or factors relative to the stimulus, determine (or significantly contribute to determining) the differences in participants' behaviour. Furthermore, there is random assignment. Random assignment is one of the most efficient techniques experimental psychologists have to neutralize (or at least to drastically reduce) the impact of personal and background differences among the participants on the experimental outcome. In this specific case, it allows them to exclude (or at least to minimize) the possibility that the responses of the participants are due to a greater or lesser inclination of the subjects to help, to their previous experiences as regards emergencies, or to any other aspect of their personality, education and past experience (*see* AWA 1997, pp. 49–50).

In other terms, random assignment is advantageous as it guarantees that the confusion of subject-related variables with the experimental variable occurs just by chance: 'only chance can cause the groups to be unequal with respect to any and all potential confounding variables associated with the group members' (McBurney and White (MW) 2004, p. 196). Obviously, the possibility that these characteristics were not distributed in a perfectly even manner across conditions cannot be completely ruled out. However, it is drastically reduced. Furthermore, thanks to the casual assignment, this possibility becomes statistically measurable. As we will see, the problem with the random assignment is not in the technique itself, but rather in the fact that this is a strategy that is impossible to adopt in specific circumstances. I will return to it later.

Beyond random assignment of the subjects to experimental conditions, there are other strategies for keeping the action of 'unwelcome' factors under control. One of these is building nuisance (or confounding) variables into the experiment: the experimenter designs the experiment so that factors which could interfere with the independent variable become independent variables in the study as well. Hence, one will have an experimental design with more than one variable.

In conclusion (1) random assignment of the subjects to the conditions that the experimenter has previously defined (the two or more values of the independent variable), and (2) the possibility for the experimenter to control confounding variables, are the two aspects characterizing the experimental method as such and distinguishing it from observational and correlational methods (*see* AWA 1997, p. 50).

This last consideration allows me to introduce an important distinction for experimental psychology and for the arguments that will follow: the distinction between experimental and non-experimental (or quasi-experimental, or correlational) research. This is the way in which McBurney and White describe it:

> The distinction between experimental research and nonexperimental research is based on the degree of control that the researcher has over the subjects and the conditions of the research. The key words are manipulation and assignment versus observation. An experiment is a kind of investigation in which some variable is manipulated. The researcher has enough control over the situation to decide which participants receive which conditions at which times. [...] Nonexperimental research is often called correlational research. [...] What makes research correlational in the common usage is the inability to manipulate some variable independently. (MW 2004, pp. 214–215)

Still apropos of the distinction between quasi experiments and true experiments, they write:

> Whereas it is possible to *assign* subjects to conditions in a true experiment, in a quasi experiment it is necessary to *select* subjects for the different conditions from previously existing groups. [...] Another way to look at the difference between true experiments and quasi experiments is to note that in true experiments we manipulate variables, whereas in quasi experiments we observe categories of subjects. When we take two pre-existing groups and consider a difference between them to be the independent variable, we are not manipulating a variable but simply labelling groups according to what we think is the important difference between them. (MW 2004, pp. 330–331)

Let us suppose ourselves to be interested in establishing which one of two teaching methods (A and B) is the best, that is, which is the method causing better performances in terms of scores among students.

'Robust' scientific research would require the experimenter to set up the conditions and to assign subjects to them (that is, to constitute, by means of random allocation, two groups *ex novo* and to treat the members of the first group with method A and those of the second group with method B), in order to nullify or at least minimize the effect of subject-variables on the outcome of the study, and to justify the conclusion that it is exactly a difference in methods that has caused a difference in the performances of the students belonging to the two groups. However, let us suppose that the social scientist does not have this possibility and that the two groups are already formed (in technical terms, are *ex post* or *post factum*). In particular, let us suppose that the researcher has to deal with existing classes. In that case, he or she will be allowed to state that there are differences in the quality of students' performances in the two classes (if there are any differences), however, he or she will not be licensed to draw conclusion of the following kind: students treated with the method A show better performances than students treated with method B, therefore method A is more effective than method B. In other terms: the researcher could highlight a correlation between method A and the performances of the students treated with A, but no cause–effect relation between them. Namely, the cause of a difference in learning between the students exposed to method A and those exposed to method B could have been any pre-existing difference between the two classes. For instance, students of the class in which the method A is used could have a higher average IQ than students exposed to method B.

In regards to this specific inquiry, the impossibility of manipulating the variable and randomly assigning subjects to the conditions is due to contingency: the social scientist does not have the resources and/or the time necessary to form two new classes and to treat them with the two different teaching methods. In principle, however, this could be done: one could proceed by constructing a real experiment.

The point I have just highlighted is important: namely, there are studies that cannot be 'transformed' into real experiments. These are all the inquiries that take as an independent variable demographic factors (gender, age, socioeconomic status, and so on): inquiries in which the independent variable is not accidentally, but rather intrinsically *post factum*.

Let us imagine we are interested in gender differences in detecting hidden figures (embedded figures task).[1] No statistic cause–effect relationship can be established between a subjective factor, such as gender, and a specific attitude or behaviour, such as the way people detect hidden figures, as one can (1) neither manipulate the independent variable nor (2) randomly allocate the subjects in the two groups: 'the experiment takes place long after the subjects become males and females' (MW 2004, p. 331). Namely, without random assignment one cannot rule out (or, at least, drastically reduce) the possibility that other factors besides the independent variable (gender) affect the dependent variable (determine a difference in the responses to the cases). Hence, one cannot conclude that the distinguishing factor one has singled out as salient between the (two) groups has caused a modification in the dependent variable. In principle, any other difference between the subjects of the two groups could have done so.

Another problematic aspect of the researches in which the independent variable is a subjective variable is the problem of so-called confounded variables.[2] Let's imagine we are interested in studying gender differences in colour preference. We observe that colour preferences do in fact vary with gender. Does it mean that females prefer certain colours, since they are females, and males other colours, since they are males? This is (at least) dubious. A difference in colour preference between males and females could be the result of past experience and not of gender (*see* MW 2004, p. 171). Any inquiry into the differences in psychological processes between men and women is equally debatable: in our society (and in many others), males and females are subject to greatly differing influences from the time they are babies and 'all the innumerable experiences and the resulting learned attitudes and skills are confounded in the simple term gender differences. To the extent to which these confounded

[1] In the embedded figures test, the research participant is shown a complex background figure and asked to describe it. After this, the participant is shown a target (such as the outline of a triangle) and asked to locate the target amid the background figure).

[2] Confounded variables should not be confused with nuisance (or confounding) variables: whereas the nuisance variable is a factor which varies independently from the independent variable and can be controlled by projecting, for instance, a factorial design, the confounded variable varies together (co-varies) with the independent variable in a way that their separate effects cannot be distinguished.

variables contribute to the sex differences found in research is the crux of the sex difference controversy' (MW 2004, p. 171).

The point made about gender is valid for any other subjective factor, and hence for the majority of experimental philosophers' studies: as in the case of gender, in the case of studies involving any subjective factor whatsoever as independent variable, one cannot talk of *real* experiments. The problem I am raising is not merely verbal. I am not claiming merely that experimental philosophers' studies should not be called *experiments*, or that experimental philosophers should stop using the adjective *experimental.* What I am claiming is that no conclusion on the way in which Westerners and East Asians, women and men tend to judge philosophically relevant cases can be drawn on the basis of experimental philosophers' studies.

Secondly, by saying that experimental philosophers' studies cannot be considered real experiments (tools for determining whether there is a relation of statistically reliable causality between gender, culture, socio-economic status of subjects and subjects' responses to the cases), it might seem that I claim that X-Philers *explicitly* state that their studies seek to confirm or disconfirm causal hypotheses. They do not explicitly state this and I do not intend to claim they do. However, if one considers the conclusions experimental philosophers draw on the basis of their surveys (that is, group-specificity of philosophical theories, arbitrariness of any position adopted on intuitions basis, philosophy as a prerogative of a group of persons having specific characteristics, and so on), then it seems that they assume that these causal connections exist and that their studies prove them. Again, by showing that experimental philosophers' studies are not *real* experiments, my goal is to argue that these studies should not be considered appropriate evidence for any theory about how Westerners and East Asians, women and men tend to judge philosophically relevant cases. Two conclusions in particular cannot be drawn: (1) that the way people judge depends on/is shaped by their gender, culture; and (2) that philosophy is a prerogative of groups of persons having specific characteristics ('the Philosophy club').

What I have claimed up to this point applies to studies of demographic factors in general. In addition to these observations, one could raise more specific concerns about particular studies.

Let us look again at the Mallon, Machery, Nichols, and Stich (MMNS) study. The study cannot be used to draw conclusions on cultural relativism for the reasons I adduced previously: MMNS take as independent variable a subjective factor (the cultural background of the interviewed subjects). Consider two groups which are different in respect to the desired characteristic (a group of undergraduates from Rutgers and a group of undergraduates from Hong Kong University); present them with a written story (the Gödel–Schmidt case) followed by a multiple-choice question (with an option describing the intuition shared by the philosophical community and another option describing the opposite response); register the answers of the two groups; observe there is a certain discontinuity between the preferences of the members of the first group and those of the second group (more than half of American university undergraduates agreed with the prevailing philosophers' evaluation and little more than a half of Hong Kong University students with the opposite one); and conclude that culture is determinant in respect to the way persons intuit: culture shapes their semantic intuitions and determines the semantic theory they will embrace.

We know that studies taking a subject variable as independent variable do not enable the experimenter to control the independent variable, nor to randomly assign the subjects to the conditions describing it. Moreover, these studies are characterized by the dilemma of confounded variables. But then, MMNS (and all experimental philosophers carrying out studies of demographic factors) are wrong in characterizing their studies as experiments and in claiming that cultural background (or any distinguishing factor they single out as salient between the two groups) shapes the way people judge.

There are also specific reasons to doubt the conclusions MMNS draw on the basis of this study. First, MMNS tell us that the subjects recruited in order to assess the influence of culture on semantic intuitions are, respectively, 31 undergraduates from Rutgers and 41 from the University of Hong Kong. Aside from the perplexities about the size and representativeness of the sample,[3] one could point out that MMNS's choice of the sample reminds us of the experimenter who, in order to assess the

[3] I will discuss this point when dealing with the problem of sampling.

effectiveness of different teaching methods, studies two classes that are already formed. As in that case, MMNS focus their attention on pre-existing groups, take a difference between the two groups (the cultural background) as salient and connect the differences in the ways the members of the two groups answer their question to this factor. This conclusion is illegitimate: between the students attending the University of Hong Kong and those attending Rutgers, there could be many other differences (not specifically cultural) affecting the trend of the answers—for instance, a difference of the average IQ of the students of the two groups or a difference in their capacity to read and understand written texts.

Secondly, experimental philosophers claim that subjects' intuitions can vary according to a series of demographic factors: culture, gender, socio-economic status and subjects' level of education. MMNS focus on the culture factor and conclude that culture shapes people's intuitions on reference. But if one has reasons to think that also other factors might shape our inclination to judge, then this conclusion cannot be drawn. Consider that participants in MMNS's survey were 18 females and 13 males from Rutgers, 25 females and 16 males from Hong Kong University. In proportion, the Hong Kong group contains more females than Rutgers. The sample is unbalanced. Let us suppose that the idea that intuitions vary according to gender is confirmed. Then, it could be the case that the greater proportion of women in the Hong Kong group determine a wider presence of descriptivist intuition in the group, rather than (just) subjects' cultural background.

The point is general: if one has reasons to think that more than one factor could have an impact on a certain behaviour, then, when designing an inquiry meant to check the effect of just one of these factors, one should make sure that the two groups forming the sample are balanced in all the other aspects.

Finally, as Sytsma and Livengood (2011) point out, there might be an ambiguity in the question MMNS ask. MMNS do not indicate whether the choices should be read from the narrator's epistemic perspective or from John's perspective:

> From the narrator's point of view, 'the person who really discovered the incompleteness of arithmetic' denotes Schmidt and 'the person who got

> hold of the manuscript and claimed credit for the work' denotes Gödel; but, from John's perspective 'the person who really discovered the incompleteness of arithmetic' denotes Gödel. (Sytsma and Livengood 2011, pp. 320–321)

Furthermore, the test might be ambiguous with regard to speaker's reference and semantic reference:

> Assuming that 'Gödel' actually refers to Gödel (semantic reference), John might nonetheless be taken to intend to be talking about Schmidt (speaker's reference). A participant might therefore answer (A) despite sharing Kripke's intuitions about the semantic reference of the name 'Gödel'; she does so because she thinks that John intends to be talking about Schmidt. (Systma and Livengood 2011, p. 321)

In sum, for general reasons concerning studies having a demographic factor as independent variable, and for specific concerns about the sample and the ambiguities of the question MMNS ask, no conclusion about the cultural relativity of our semantic intuitions can be drawn on the basis of their study.

8.2.2 Inquiries

In the last section, I argued that because of the impossibility of randomly assigning subjects to conditions, given the risk of confounded variables, and because of the absence of alternative strategies aimed at controlling possible 'disturbing' factors, experimental philosophers' studies cannot be qualified as real experiments. This fact, however, does not rule out the possibility that their studies could highlight a correlation between subjects' gender, culture, and the way they evaluate TE scenarios. And, in fact, experimental philosophers' experiments resemble surveys.

Surveys—together with observational research, archival research and case studies—belong to the field of non-experimental (or correlational) research. The primary goal of a survey is that of collecting people opinions. MW suggest that 'a major function of surveys is to dispel myths' (MW 2004, p. 238). Indeed, one of the declared goals of the experimentalist is to dispel the myth of a general agreement on philosophers' judgements. This characterization may suggest that X-Philers' studies belong to the field of correlational research. However, correlational research has its own methodological standards as well. Do experimental philosophers' studies meet them? Are they real surveys?

One of the most important aspect of surveys is sampling: 'surveys differ greatly in value according to how respondents are sampled' (MW 2004, p. 247). Given that answers to a survey are not useful if they only reflect the opinions of the subjects who are actually tested, the selection of subjects constituting the sample has to be made so as to enable the generalization of results to the population. Let us imagine we are interested in generalizing the results of a certain survey Y to a population X: what we should do is randomly select the elements of our sample from the population X, that is, proceeding in a way such as any person in that population has an equal and independent chance of being selected for the survey Y. Only in the case in which a sample has been randomly selected it is possible to assume that subjects' responses are reasonably similar to those of the entire population. Random sampling is also called probabilistic sampling as it allows the researcher to apply various statistical techniques and to calculate the probability that any person of a certain population has of being in the sample (*see* MW 2004, p. 244). Among the different strategies of sampling only the probabilistic one is fully satisfactory.

The problem with the studies of experimental philosophers is that they do not use any form of probabilistic sampling. Their studies use convenience samples. Usually, a convenience sample consists in a selection from students enrolled in introductory psychology courses—in the case of experimental philosophers, philosophy courses (*see* MMNS' study sample)—or in a selection from the population of researchers' home city (*see* Weinberg, Nichols, and Stich's survey sample). Indeed, this is a form of sampling which is widely used in psychology, for merely practical reasons: convenience sampling is economically convenient. Adopting a convenience sampling isn't a wrong choice in itself. However, in the case in which one decides to proceed in this manner, it is necessary (1) to have solid reasons for claiming that, as regards to the question(s) one is interested in, the restricted population from which the sample is extracted (student population) is uniform with the population to which one wants to generalize the results (for instance, the entire adult population of North America); (2) to sample randomly at least from within the restricted population (*see* MW 2004, p. 207).

What about experimental philosophers' surveys—for example, MMNS's survey? Do we have good reasons for believing that, as regards the question X-Philers are interested in, the student population is uniform in respect to the population they want to generalize the results to (a population with a Western cultural background on one side, a population with an East Asian cultural background on the other)? Moreover, could the sample X-Philers used (Rutgers undergraduates and Hong Kong University undergraduates) be considered a randomly selected sample from the student population? The answer to the second question is no: the study takes into account only a group of undergraduates from Rutgers and a group of undergraduates from Hong Kong University. This consideration alone could be enough to doubt the statistical validity of MMNS's survey.

However, the first question is also interesting as it allows us to deal with another serious shortcoming of this and other X-Philers' studies. According to MMNS, the trend registered among Rutgers students should be extended to the population with a Western cultural background, and that registered among Hong Kong University students to the population with an East Asian cultural background. However, they do not specify who belongs to the population with a Western cultural background and who to the population with an East Asian cultural background.

Apart from sampling, there is another important aspect one must consider when running a written survey: one must be able to exclude the possibility that a significant proportion of the interviewed subjects have difficulties in understanding prose writings (in general), or prose writings of a certain degree of complexity. X-Philers' data reveal that a large proportion of interviewed subjects make a judgement opposed to the one philosophers believe is correct. Experimental philosophers conclude that their epistemological, semantic, etc. opinions diverge from those of professional philosophers. Are they justified in concluding this? Could it not simply be the case that some participants fail to understand the text of the experiment?

When I presented the way experimental philosophers transform a philosophical TE into a test for non-philosophers (Sect. 7.1.), I specified that (in some cases) they pose another question before the philosophically relevant one. According to them, this question alone should confirm that subjects have really understood the case. They call it a *comprehension check*. One might argue that this is not enough, that experimental philosophers need to run *specific preliminary* studies aimed at (1) determining the degree of comprehensibility of the texts used in the survey, and (2) testing participants' comprehension skills of written texts.[4]

Moreover, there is reason to think that in the case of at least some experimental philosophers' surveys, non-experts' non-standard responses to the cases are due to insufficient comprehension of texts, rather than to different conceptions of knowledge, reference, and so on. Contrary to 'classic' surveys, experimental philosophers' studies do not consist of lists of simple questions by means of which people's opinions about specific concerns are registered (for instance, simple items, such as 'women should be permitted to decide for themselves whether to continue a pregnancy', followed by a rating 7-point scale going from point 1, 'agree', to point 7, 'disagree'), but questions presupposing the reading and the comprehension of pretty complex stories. Hence, in their case, the chief risk in a comprehension problem does not involve the question solely (or primarily), but rather the written story which the subjects are asked to assess. Furthermore, the cases X-Philers propose to non-philosophers describe scenarios we very rarely come across (or hardly think about). These cases ask for reflection on aspects such as the epistemic status of a subject, the reference of a proper name and so on; that is, issues which non-philosophers are not likely to think about.

Apart from the risk of misunderstanding the text, another hypothesis can be advanced to explain the discontinuity between philosopher's judgements and those of the interviewed subjects: laymen's diverging answers

[4] Beyond conjecturing that diverging judgements are, at least in part, the result of a failed comprehension of the texts, one could argue there are quite strong reason to doubt that things really go in this way. Namely, there are studies (for instance, ALL – The Adult Literacy and Life Skills Survey) attesting to an important fact: the difficulties that a significant part of the adult and educated of Western (and non-Western) countries have in understanding prose writings, including very easy ones.

could be due to their natural tendency to load texts with implicatures. In other terms: subjects might experience no particular difficulties in understanding written texts, but the TE's text, as it is formulated, might suggest unwanted implicatures that subjects who have not been trained to 'stick to the letter' (that is, non-philosophers) may very well be inclined to pick up. That is, some experimental philosopher's tests may be analogous to the well-known Linda case (Tversky and Kahneman 1983): some of the interviewed subjects load the text with implicatures, ending up responding to the case in the wrong way.

Let us briefly illustrate the Linda case and the theory of error I have just mentioned. The case Tversky and Kahneman (TK) elaborated is the following: Linda is 31 years old, single, outspoken, and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in antinuclear demonstrations.

After having read this short story, the subjects were asked to guess which kind of life Linda conducts, or, better, to list a series of hypotheses from the most probable to the less probable one. Among the options TK gave, the interesting ones (the options which are useful in identifying the mistake subjects make) were: 'Linda is a bank teller' (*T*), 'Linda is active in the feminist movement' (*F*) and 'Linda is a bank teller and is active in the feminist movement' (*T* & *F*). In the original experiment TK conducted (and in all subsequent re-proposals of the same case) the majority of the interviewees (more or less 80–90 %) order these options in this way: *F*, *T* & *F*, *T*. In other terms, after having judged that Linda is more likely to be a feminist, rather than a bank teller (or only a bank teller), the majority of the subjects claims that that is more probable that Linda is a bank teller and feminist, rather than a bank teller. This is a patent violation of the conjunction rule: that law of probability holding 'that the mathematical probability of a conjoint hypotheses (A & B) cannot exceed that of either of its constituents, that is, $p(A \& B) \leq p(A), p(B)$' (Hertwig and Gigerenzer 1999, p. 275). On the basis of these (and other) results, TK concluded that the majority of humans are not very good at reasoning in accord with the rules of probability.

In the following years, different theories of error were produced. Among them was the theory we hinted at previously. According to this theory (*see* Marconi 2008, p. 108), subjects judge *T* & *F* to be more probable than *T* because they read *T* as *T* [& ~*F*]. Literally, *T* only says that Linda is a bank teller, that is, it does not rule out that Linda is anything else, for instance, a feminist. However, the structure of the case suggests that *T* says that Linda is only a bank teller (hence, not a feminist). Therefore, some subjects end up judging *T* & *F* to be more probable than *T*. Against this solution, one could object that the probability of *T* & *F* is, in any case, lower than that of *T* & ~*F* (namely, there are more non-feminist than feminist women) and, therefore, conclude that this theory of error is unsatisfactory. The obvious reply to this objection is that *T* & ~*F* is more probable than *T* & *F* only *a priori*. If one considers the story, *T* & *F* must be more probable than *T* & ~*F*: namely, the story describes events of Linda's life suggesting she could be a feminist, and, on the other hand, does not give any information which lets us suppose that Linda could be a bank teller. The hypothesis *T* & *F* is therefore much more easily associable to the story than *T* & ~*F* (or even than *T* alone).

In conclusion, TK's experiment shows what every teacher knows well: people have difficulty sticking to the literal meaning, as comprehension is usually enriched with implicatures (very often undesired by the author of the text). This is the reason why, while teaching, one often says: 'By saying this, I do not mean…'—that is to say, one makes an effort to cancel implicatures.

Readers' tendency to enrich texts with implicatures is related to the issue of the discontinuity of laymen's responses from philosopher's for the following reason: exposure to a certain *modus operandi* and commitment to high standards of rigour predispose professional philosophers to take the content of philosophical texts and discourses literally (or even to 'deactivate' certain mechanisms that are typical of comprehension, such as, for instance, loading texts with implicatures).

Still in fairness to the methodological validity of experimental philosophers' studies, one could raise other questions. The problem of construct validity is one of those. An experiment has construct validity only if it is possible to show—within reasonable limits—that none of the alternative theoretical explanations (which we know of) are more plausible than the theory we believe is confirmed by the data. One of the threat to construct validity is the ambiguous effect of independent variables.

> An experimenter may carefully design an experiment in which all reasonable confounding variables seem to be well controlled, only to have the results compromised because the participants perceive the situation differently than the experimenter does. (MW 2004, p. 176)

In many cases, for instance, subjects end up conducting 'their own experiment', invalidating the experiment researchers thought they were conducting. A typical prejudice pushing participants to conduct their own experiment is 'the concern that the experimental procedure in some way measures the participant's competence. Some participants are convinced that the experiment is a carefully disguised measure of intelligence and emotional adjustment. This expectancy give rise to evaluation apprehension' (MW 2004, p. 177).

One might conjecture that it is the evaluation apprehension that causes differences in the way a great proportion of the interviewed women (but also East Asians, or people with low socio-economic status) answer the cases, rather than gender (culture, social status) itself. Such conjecture may be supported by the apparent tendency of people belonging to categories that are prejudicially associated with limited intellectual capacities (or poor capacities in a specific field) to show poor performance in tasks which are presented, or merely perceived, as diagnostic of their intellectual capacities (or their capacities in that area). In the case of experimental philosophers' studies, it is possible that women (and other persons belonging to under-represented categories in philosophy), believing their abilities in the field are being tested, fear to confirm the negative stereotype they are victim of (a stereotype according to which they are less inclined to pure speculation), or to be judged on the basis of that stereotype. It is possible that this fear negatively affects performance and that many of them end up answering incorrectly, 'confirming', in spite of themselves, the stereotype.

In order to better explain the phenomenon of the confirmation of the stereotype, Gendler (2011) introduces the concept of stereotype threat.

> Stereotype threat is a well-documented phenomenon whereby activating an individual's thoughts about her membership in a group that is associated with impaired performance in a particular domain increases her tendency

> to perform in a stereotype-confirming manner. So, for example, as Claude Steele and Joshua Aronson hypothesized in their original 1995 paper, 'whenever African American students perform an explicitly scholastic or intellectual task, they face the threat of confirming or being judged by a negative social stereotype—a suspicion—about their group's intellectual ability and competence [...] The self-threat [...] may interfere with the intellectual functioning of these students, particularly during standardized tests' (Steele and Aronson 1995, p. 797). (Gendler 2011, p. 16)

In their study, Steele and Aronson proceeded by giving college students—not only African Americans—a quiz consisting in a selection of items from the Graduate Record Exam (GRE). Some of them were in a stereotype-threat condition, other in a non-stereotype-threat condition. In the first case, the test was presented as diagnostic of intellectual ability; in the second one, as a mere problem-solving task, which was not diagnostic of any ability. The results? While in the first condition the performances of African American participants were poorer than those of non-African Americans, in the second condition the performances of the former dramatically improved, matching the performance of the latter. The same kind of phenomenon has been registered in a series of tests:

> When a task is described as diagnostic of intelligence, Latinos and particularly Latinas perform more poorly than do Whites (Gonzales et al. 2002), children with low socioeconomic status perform more poorly than do those with high socioeconomic status (Croizet and Claire 1998), and psychology students perform more poorly than do science students (Croizet et al. 2004). Even groups who typically enjoy advantaged social status can be made to experience stereotype threat. White men perform more poorly on a math test when they are told that their performance will be compared with that of Asian men (Aronson et al. 1999), and Whites perform more poorly than Blacks on a motor task when it is described to them as measuring their natural athletic ability (Stone 2002; Stone et al. 1999). (Gendler 2011, p. 17)

Hence, in the case of experimental philosophers' studies on gender and culture, one could ask the following: would it not be more appropriate to suppose that the tendency registered by women and other subjects belonging to under-represented categories in the philosophical community to answer in a non-standard manner to TEs is ascribable to the widely scrutinized and ascertained phenomenon of negative stereotype confirmation, rather than to profound, systematic and important differences in the way people belonging to these groups are inclined to judge?

In this section, I have raised few questions regarding experimental philosophers' methodology: general issues on the correctness of drawing conclusions about the (cultural, gender, socio-economical) relativity of intuitions, the absence of studies on ruling out the possibility that non-experts' non-standard responses to cases are due to insufficient comprehension of the texts or to readers' natural tendency to enrich texts with implicatures, problems with sampling, and a question of construct validity (for studies on demographic factors) based on the notion of stereotype threat. This section is not meant to be exhaustive of all the weaknesses and limitations of experimental philosophers' studies, merely to give an idea of how much their methodology needs to be refined.

8.3 Experimental Philosophy: A Non-philosophical Project (Part II)

In the previous sections, I argued that the experimental philosophers' project could hardly be regarded as philosophical or as experimental. One might observe that experimental philosophers may be able, in the future, to design tests that overcome the methodological difficulties highlighted above, and that the results of surveys based on such new tests could confirm the theses currently defended by X-Philers. If this were the case, would the reliability of philosophers' judgements and methods be challenged?

In this section, I will argue that the answer to this question depends (1) on the image that one has of philosophy, that is, on whether the aims and results of philosophical inquiry are conceived as descriptive of what we would say, or normative in respect to our intuitions; (2) on the description one gives of philosophical methodology; and, more importantly, (3) on whether intuitions are conceived as empirical hypotheses or not.

Let us start from (1) and (2). As I argued in Chap. 5, philosophers generally refuse to see their theories on X as the mere registration of the received opinion on X. Moreover, describing what we say or think is not what the majority of philosophers have done up to now: many (perhaps all) philosophical theories are corrective of our intuitions; they are seen as means to discriminate between what is really correct to say and what we just say or think is correct to say. In regard to this question, RE theorists seem to have found a good compromise between the necessity of appealing to intuitions and the fact that philosophical theories are conceived as (and in fact are) normative in respect to our intuitions. Namely, according to them, the process leading philosophers from intuitions to norms is a procedure which, through IBE and corrections and improvements (both to theory and intuitions), aims to set an equilibrium between the theory and intuitive judgements. Moreover, other beliefs (for example, scientific knowledge) that we think are relevant to the problem at issue can also be used in this process (*see* Chap. 3 and Sect. 5.2.2).

Hence, in the light of these considerations, one could counter experimental philosophers' theses by arguing that philosophical theories are norms, not mere descriptions of what we would say. Furthermore, in the constructive and justificatory process leading to such norms, there is much more than the appeal to intuitions generated by means of TEs.

The last consideration is important because the impression one gets from experimental philosophers' texts is that philosophy consists, mostly, in work aimed at designing TEs. In particular, the debate among supporters of competing theories would amount to a game in which who wins and who loses can be determined on the basis of theories' consistency or inconsistency with the responses generated by TEs. This is not the way things stand.

Against this simplistic reconstruction, one can first point out that not all intuitions philosophers appeal to are elicited by TEs. Quite the contrary: a large part of the intuitions philosophical theories are built on are not introduced by means of these devices. For example, the intuition according to which it is wrong to torture innocents for mere fun is a judgement whose content is generally conveyed directly. TEs come into play when some sort of theory has already been advanced: they attack theories that are obtained by inference to the best explanation from an initial set of intuitions, the majority of which comes from a reflection on real cases. Furthermore, designing a TE is not an easy task: in several cases, a lot of time is needed to identify or imagine a case working as a counterexample against a theoretical framework (see Chap. 3).

Hence, one could ask: why do experimental philosophers test the diffusion of the intuitions generated by TEs, rather than caring about what people say or think about cases they usually come across and are used to assess? Wouldn't it be better to test people's judgements concerning these simple and common cases?

Secondly, experimental philosophers seem to forget that, in philosophy, TEs are part of an ongoing inquiry, whose partial results are well known to the philosopher who is proposing the case and to his colleagues, but not to people they interview. This is important because it would seem that, in order to assess some cases correctly, it is necessary to be familiar with the systematizations that have been done up to that moment; familiar with the theory that is attacked; and, possibly, familiar with the process that produced it. Let us take Gettier examples. When confronted with these cases, epistemologists start from a clear thesis concerning knowledge: knowledge is justified true belief. Experimental philosophers offer no definition, or, in general, no clarification of the polemical target of the experiment to their subjects when conducting their inquiries. One could conjecture that this omission is problematic, or more; that it makes subjects' error somehow predictable. The point is not merely that the controversial thesis is not explicitly stated before the case, but that a correct evaluation of the case requires those who assess it to know the systematizations of the use which have been developed up to that moment, that is, to know that, when asked whether the person in the case knows that *p*, one is not asked whether the person just believes, firmly believes, or truly believes that *p*—although, in everyday life, 'to know' is used in these senses, too.[5]

[5] As I argued in Sect. 6.3, in respect to everyday needs, inaccurate uses of the expression of a term or expression 'X', for instance of 'to know', aren't necessarily problematic. Problems arise when one starts asking what X (for instance, knowledge) is; that is, when normative goals come into play. In the perspective of an epistemologist, using 'to know' just to mean 'to truly believe', 'to firmly believe' or 'to believe' would amount to disregarding an important distinction that any of us would recognize when confronted with cases in which the differences between a simple belief and a true belief, or between a true belief and a true and justified belief, are manifest.

Lastly, experimental philosophers seem to misunderstand the role of TEs in another sense. Let us take the case described in Chap. 7: the Gödel–Schmidt case. In 'Semantics cross-cultural style', MMNS argue that giving the non-standard answer rather than the standard answer makes people descriptivists (or more inclined to embrace descriptivism) rather than Kripkian (or more inclined to embrace the Kripkian alternative); and in particular, makes East Asians descriptivists and Westerners Kripkians. Let us suppose that, as X-Philers believe, subjects' responses to the cases are the expression of their conception of the reference of the proper name 'Gödel': would this mean that subjects answering (a), rather than (b) could be qualified as descriptivists, rather than as non-descriptivist? Not really. Even if it is true that Kripke's theoretical proposal has had a certain success, an important proportion of the supporters of descriptivism, whilst acknowledging Kripke's TE and agreeing with his intuitions, remained descriptivists. In their article, MMNS themselves allude to sophisticated descriptivists, that is, supporters of forms of 'mature' descriptivism that do not clash with Kripkian intuitions. It is then surprising to see that MMNS claim there is an obvious connection among having a certain intuition, giving up a certain theoretical option (descriptivism) and embracing an alternative such as the causal-historical picture.

The point is general. Contrary to the simplistic description of the philosophical method given by experimental philosophers, an intuition does not work as a clear-cut divide between rival theories, and is not necessarily crucial in respect to the theoretical framework it contradicts. It is a burden for those who support the theory mismatching with it: they must explain, or explain away, the intuition. However, it does not commit them to abandon their theory, and, least of all, to embrace the rival theory.

In order to better illustrate this last point, it is useful to remind ourselves what possibilities are open to a philosopher when his or her theory is attacked by a counterexample. The theory could be adjusted in order to account for the intuition. Or one could find a theory of error that explains the intuition away. It might also be the case that, after having reflected, the philosopher has to acknowledge that the intuition is irresistible and has to abandon the theory. If this is the case and another theory accounting for that intuition is available, should the philosopher embrace it? It depends. The alternative could be attractive as it accounts for the intuition in question. However, this second theory has to be evaluated under other aspects as well: the capacity to account for other intuitions, and to satisfy the characteristics a good theory should have (*see* Chap. 3).

In the previous sections, I claimed that experimental philosophers have a partial view of the process leading to the construction and justification of philosophical theories: they tend to focus exclusively on TEs, underestimating the 'big picture' encompassing them. By doing that, they also seem to misunderstand the proper role of TEs. Apropos, I highlighted that TEs are a small parts of a broader process in which a theoretical picture is gradually outlined. Several TEs are introduced in developed stages of the research on a certain topic, to show the incompleteness or unacceptability of a theory which is, in turn, the result of a process where other problematic cases have been taken into account and important distinctions have been made. Generally, those who assess TEs—philosophers—are familiar with this process, or, at least, with its results. On the other hand, the subjects that experimentalists interview are not. However, being familiar with the results and the kind of procedures producing them seems to be necessary to assess TEs correctly.

Against experimental philosophers' theses, however, one could raise yet another more radical criticism. This criticism has to do with (3), that is, the question of whether intuitions are empirical hypotheses or not. The assumption experimental philosophers move from in order to assert the necessity of their project is that the verdicts philosophers appeal to are untested empirical hypotheses. I argued (Chap. 6) that they are not: intuitive judgements are not empirical hypotheses of any kind—they are neither hypotheses on how laymen would judge the cases (would answer a question such as, for instance, 'does the subject in the GC know that *p*?'), nor predictions on how people would apply a concept/use a term (reason, behave, and so on). But, what do I mean by saying that judgements such as 'we wouldn't say that the subject knows that *p*' or 'given the way we use "knowledge"/apply *knowledge*, this belief wouldn't be called "knowledge"/categorized as *knowledge*' are not empirical hypotheses?

Briefly, this is the way I answered this question in Chap. 6. Let us imagine a philosopher in front of Gettier cases, asking himself or herself whether it would be appropriate to apply *knowledge*/use the term 'knowledge' in such situations. His or her question wouldn't be: would someone eventually apply the concept/use the term in such circumstances? But rather: would it be correct to apply the concept/use the term in such circumstances? In a perspective which recognizes both the centrality of the practice and the revisionary work of philosophers, the question would sound like: could this application of the concept/this use of the term be eventually explicated by a norm, that is, by a consistent and coherent framework accounting for a significant part of our applications/uses? It is then clear that judgements in response to this question will not count as empirical predictions, but rather as judgements of correctness—hypotheses on the correct application of X. A judgement like 'given the way we usually use "X"/ apply *X*, this Y would (or wouldn't) be called "X"/characterized as *X*' could then better be reformulated as 'given the way we usually use "X"/ apply *X*, this Y should (or shouldn't) be called "X"/characterized as *X*'.

Several advocates of armchair philosophy have tried to explain the sense in which philosophers' judgements cannot be compared to those of non-philosophers (Kauppinen 2007; Ludwig 2007; Williamson 2007 and 2011; Jackman 2009; Martì 2009). The idea sketched by Jackman in 'Semantic intuitions, conceptual analysis, and cross-cultural variation' (2009), combined with RE strategy and the normative view on intuitions, seems to be promising. According to Jackman, philosophers' judgements cannot be compared to those of laymen as philosophers are those who are typically engaged in the enterprise of outlining norms, or, as he says, 'in the consistency-driven construction that conceptual analysis involves' (Jackman 2009, p. 13). In short, philosophers' reflective intuitions are often superior because they are largely shaped by philosophers' main goal, that is, the project outlining norms. A philosopher wondering whether a certain application of a concept/use of a world (reasoning, behaviour, and so on) could be explicated by a consistent and accurate framework, must reflect on the practice, consider the situation she is evaluating in light of other cases, and negotiate her intuitions in order to make them consistent (*see* Sect. 6.3).

Hence, whereas non-philosophers' evaluations of a TE will register, in the best-case scenario, their disposition to apply a concept (use a term) in a specific case, in the light of the way in which in everyday contexts and with no particular concern about coherence and precision they apply the concept (use the term), philosophers' judgements will count as normative hypotheses on the application of *X*/on 'X' usage. In other terms, intuitions are verdicts philosophers develop on the basis of their competence and of the reflection that is needed to assess, *prima facie*, the possibility that a certain application of a concept/ use of a term has to be explained by a norm.

In conclusion, if one (1) takes philosophy to be a normative enterprise, (2) adheres to RE, that is, a descriptive and revisionary method, and (3) is persuaded that the way philosophers judge the cases is 'affected' by the kind of project philosophers pursue, then studies showing a mismatch between non-philosophers' and philosophers' judgements do not provide a strong enough reason to question philosophers' methodology.

References

Aronson, E., T.D. Wilson, R. Akert, and (AWA). 1997. *Social Psychology*, 2nd edn. New York: Addison–Wesley Educational Publishers.

Aronson, J., Lustina, M. J., Good, C., Keough, K., Steele, C. M., & Browin, J. (1999). When white men can't do math: Necessary and sufficient factors in stereotype threat. Journal of Experimental Social Psychology, 35(1), 29–46

Cappelen, H. 2012. *Philosophy without Intuitions*. Oxford: Oxford University Press.

Croizet, J.-C., & Claire, T. (1998). Extending the concept of stereotype threat to social class: The intellectual underperformance of students from low socioeconomic backgrounds. Personality and Social Psychology Bulletin, 24(6), 588–594

Croizet, J.-C., Despre`s, G., Gauzins, M.-E., Huguet, P., Leyens, J.-P., & Me´ot, A. (2004). Stereotype threat undermines intellectual performance by triggering a disruptive mental load. Personality and Social Psychology Bulletin, 30(6), 721–731.

Deutsch, M. 2009. Experimental Philosophy and the Theory of Reference. *Mind and Language* 24(4): 445–466.

Gendler, T.S 2011. On the Epistemic Costs of Implicit Bias. *Philosophical Studies* 156: 33–63.

Gonzales, P. M., Blanton, H., & Williams, K. J. (2002). The effects of stereotype threat and doubleminority status on the test performance of Latino women. Personality and Social Psychology Bulletin, 28(5), 659–670.

Hertwig, R., and G. Gigerenzer. 1999. The conjunction fallacy revisited: How intelligent inferences look like reasoning errors. *Journal of behavioural decision making* 12: 275–305.

Jackman, H. 2009. Semantic intuitions, conceptual analysis, and cross-cultural variation. *Philosophical Studies* 146(2): 159–177.

Kauppinen, A. 2007. The Rise and Fall of Experimental Philosophy. *Philosophical Explorations* 10: 95–118.

Ludwig, K. 2007. The Epistemology of Thought Experiments: First Person versus Third Person Approaches. *Midwest Studies in Philosophy* 31: 128–159.

McBurney, D.H., and T.L. White. 2004. *Research Methods*, 6 edn. Belmont, CA: Wadsworth/Thomson Learning.

Marconi D. (2008) Filosofia e scienza cognitiva (Bari-Roma: Laterza).

Martí, G. 2009. Against semantic multiculturalism. *Analysis* 69: 42–48.

Stich, S 2009. Reply to Sosa. In *Stich and His Critics*, eds. D. Murphy and M.A. Bishop. Malden: Wiley-Blackwell.

Steele, C.M., & Aronson, J. (1995). Stereotype Threat and the intellectual test-performance of African-Americans. Journal of personality and Social Psychology, 69 (5): 797–811.

Steele, C.M., & Aronson, J. (1995). Stereotype Threat and the intellectual test-performance of African-Americans. Journal of personality and Social Psychology, 69 (5): 797–811.

Stone, J., Lynch, C. I., Sjomeling, M., & Darley, J. M. (1999). Stereotype threat effects on black and white athletic performance. Journal of Personality and Social Psychology, 77(6), 1213–1227.

Stone, J. (2002). Battling doubt by avoiding practice: The effects of stereotype threat on selfhandicapping in white athletes. Personality and Social Psychology Bulletin, 28(12), 1667–1678.

Sytsma, J., and J. Livengood. 2011. A new perspective concerning experiments on semantic intuitions. *Australasian Journal of Philosophy* 89(2): 315–332.

Tversky, A., and D. Kahneman. 1983. Extension versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review* 90(4): 293–315.

Williamson, T 2007. *The Philosophy of Philosophy*. Malden, MA: Blackwell.

——— 2011. Philosophical Expertise and the Burden of Proof. *Metaphilosophy* 42: 215–229.

References

Ambady, N., M. Shih, A. Kim, and T.L. Pittinsky. 2001. Stereotype susceptibility in children: Effects of identity activation on quantitative performance. *Psychological Science* 12(5): 385–390.

Cohen, L.J. 1981. Can human irrationality be experimentally demonstrated? *Behavioural and Brain Sciences* 4: 317–331.

Cullen, S. 2010. Survey-Driven Romanticism. *Review of Philosophy and Psychology* 1(2): 275–296.

DePaul, M., and W. Ramsey, eds. 1998. *Rethinking Intuition: The Psychology of Intuition & Its Role in Philosophical Inquiry*. Lanham, MD: Rowman & Littlefield.

Deutsch, M. 2010. Intuitions, Counter-Examples, and Experimental Philosophy. *Review of Philosophy and Psychology* 1(3): 447–460.

Devitt, M. 2011. Experimental Semantics. *Philosophy and Phenomenological Research* 82: 418–435.

Earlenbaugh, J., and B. Molyneaux. 2009. Intuitions are Inclinations to Believe. *Philosophical Studies* 145(1): 89–109.

Fisher, E., and J. Collins, eds. 2015. *Experimental Philosophy, Rationalism, and Naturalism: Rethinking Philosophical Method*. London: Routledge.

© The Author(s) 2016

S.M. Nicoli, *The Role of Intuitions in Philosophical Methodology*,
DOI 10.1057/978-1-137-56715-4

Gendler, T.S. 2004. Thought Experiments Rethought—and Reperceived. *Philosophy of Science* 71: 1152–1164.

Goldman, A., and J. Pust. 1998. Philosophical Theory and Intuitional Evidence. In *Rethinking Intuition: The Psychology of Intuition and Its Role in Philosophical Inquiry*, eds. M. DePaul, and W. Ramsey. Lanham, MD: Rowman & Littlefield.

Goldman, A. 2007. Philosophical Intuitions: Their Target, Their Source and Their Epistemic Status. *Grazer Philosophische Studien* 74: 1–26.

Hacker, P.M.S. 2009. A philosopher of philosophy. *The Philosophical Quarterly* 59: 337–348.

Ichikawa, J.J., and M. Steup. 2012. The Analysis of Knowledge. The Stanford Encyclopedia of Philosophy (Spring 2014 Edition), ed. Edward N. Zalta, URL = http://plato.stanford.edu/archives/spr2014/entries/knowledge-analysis/. Date accessed 28 Dec 2015.

Kelly, D., and E. Roedder. 2008. Racial cognition and the ethics of implicit bias. *Philosophy Compass* 3(3): 522–540.

Knobe, J. 2007. Experimental Philosophy. *Philosophy Compass* 2(1): 81–92.

Lam, B. 2010. Are Cantonese Speakers Really Descriptivists? *Cognition* 115: 320–329.

Ludlow, P. 2005. Contextualism and the new linguistic turn in epistemology. In *Contextualism in Philosophy: Knowledge, Meaning, and Truth*, eds. Gerhard Preyer, and Georg Peter. Oxford: Oxford University Press.

——— 2011. *The Philosophy of Generative Linguistics*. Oxford: Oxford University Press.

Machery, E. 2012. Expertise and Intuitions about Reference. *Theoria, an International Journal for Theory, History and Foundations of Science* 73: 37–54.

Machery, E., C.Y. Olivola, and M. DeBlanc. 2009. Linguistic and metalinguistic intuitions in the philosophy of language. *Analysis* 69: 689–694.

Marconi, D. 2010. Wittgenstein and Necessary Facts. In *Wittgenstein: Mind, Meaning and Metaphilosophy*, eds. P. Frascolla, D. Marconi, and A. Voltolini. London: Palgrave Macmillan.

——— 2011b. Wittgenstein and Williamson on conceptual analysis. In Annuario e Bollettino della Società Italiana di Filosofia Analitica (SIFA).

——— 2012. Semantic Normativity, Deference and Reference. *Dialectica* 66(2): 273–287.

Murphy, D., and M.A. Bishop, eds. 2009. *Stich and His Critics*. Malden: Wiley-Blackwell.

Nagel, J. 2012. Intuitions and Experiments: A defence of the case method in epistemology. *Philosophy and Phenomenological Research* 85(3): 495–527.

Nisbett, R., I. Choi, K. Peng, and A. Norenzayan. 2001. Culture and Systems of Thought: Holistic vs. Analytic Cognition. *Psychological Review* 108: 291–310.

Sosa, E. 2005. A defence of the use of intuitions in philosophy. In *Stich and His Critics*, eds. D. Murphy, and M.A. Bishop. Malden: Wiley-Blackwell.

——— 2007. Intuitions: Their Nature and Epistemic Efficacy. *Grazer Philosophische Studien* 74(1): 51–67.

Strawson, P.F. 1992. *Analysis and Metaphysics: An Introduction to Philosophy*. Oxford: Oxford University Press.

Weatherson, B. 2013. The Role of Naturalness in Lewis's Theory of Meaning. *Journal for the History of Analytical Philosophy* 1(10): 1–20.

Weinberg, J.M., C. Gonnerman, C. Buckner, and J. Alexander. 2010. Are Philosophers Expert Intuiters? *Philosophical Psychology* 23(3): 331–355.

Williamson, T. 2004. Philosophical 'Intuitions' and Scepticism about Judgement. *Dialectica* 58(1): 109–153.

Wittgenstein, L. 2009. *Philosophical Investigations, Revised*, 4th edn. Oxford: Wiley-Blackwell.

——— 1967. In *Zettel*, eds. G.E.M. Anscombe and G.H. Von Wright. Oxford: Blackwell, Oxford.

Index

© The Author(s) 2016
S.M. Nicoli, *The Role of Intuitions in Philosophical Methodology*,
DOI 10.1057/978-1-137-56715-4