Paul Smeyers
Marc Depaepe
Editors

EDUCATIONAL RESEARCH 5

Educational Research:

The Ethics and Aesthetics of Statistics

Springer

Educational Research: The Ethics and Aesthetics of Statistics

Educational Research

VOLUME 5

Aims & Scope

Freedom of inquiry in educational research can no longer be taken for granted. Narrow definitions of what constitutes 'scientific' research, funding criteria that enforce particular research methods, and policy decision processes that ignore any research that is not narrowly utilitarian, in many countries, create a context that discourages scholarship of a more speculative, exploratory, or critical sort.

In this series, internationally leading scholars in *philosophy and history of education* engage in discourse that is sophisticated and nuanced for understanding contemporary debates. Thus social research, and therefore educational research, is again focused on the distinctive nature of what it studies: a social activity where questions of meaning and value must be addressed, and where interpretation and judgment play a crucial role.

This educational research takes into account the historical and cultural context and brings clarity to what actually constitutes science in this area. The timely issues that are addressed in this series bear witness to the belief that educational theory cannot help but go beyond a limited conception of empirical educational research to provide a real understanding of education as a human practice. They surpass the rather simple cause-and effect rhetoric and thus transgress the picture of performativity that currently keeps much of the talk about education captive. The authors are united in the belief that 'there is a place within the social sciences in general', and within the discipline of education in particular, for 'foundational' approaches that enable the systematic study of educational practice from a discipline-orientated approach.

Paul Smeyers · Marc Depaepe
Editors

Educational Research: The Ethics and Aesthetics of Statistics

Springer

Editors
Prof. Paul Smeyers
Universiteit Gent
Faculteit Psychologie en Pedagogische Wetenschappen
Dunantlaan 1
9000 Gent
Belgium
paulus.smeyers@ugent.be

Prof. Marc Depaepe
Katholieke Universiteit Leuven
Campus Kortrijk
Subfaculteit Psychologie en Pedagogische Wetenschappen
Sabbelaan 53
8500 Kortrijk
Belgium
marc.depaepe@kuleuven-kortrijk.be

ISBN 978-90-481-9872-6 e-ISBN 978-90-481-9873-3
DOI 10.1007/978-90-481-9873-3
Springer Dordrecht Heidelberg London New York

Library of Congress Control Number: 2010936458

Earlier Volumes in this Series

Educational Research: Why 'What Works' Doesn't Work
Paul Smeyers and Marc Depaepe
2006, VI, 195 p., Hardcover
ISBN 978-1-4020-5307-8

Educational Research: Networks and Technologies
Paul Smeyers and Marc Depaepe
2007, VI, 228 p., Hardcover
ISBN 978-1-4020-6612-2

Educational Research: the Educationalization of Social Problems
Paul Smeyers and Marc Depaepe
2008, VI, 247 p., Hardcover
ISBN 978-1-4020-9722-5

Educational Research: Proofs, Arguments, and Other Reasonings
Paul Smeyers and Marc Depaepe
2009, VI, 199 p., Hardcove
ISBN 978-90-481-3248-5

Contents

Chapter 1
Representation or Hard Evidence? The Use of Statistics in Education and Educational Research

Paul Smeyers and Marc Depaepe

> *The quantitative and antianthropocentric orientation of natural sciences from Galileo on forced an unpleasant dilemma on the humane sciences: either assume a lax scientific system in order to attain noteworthy results, or assume a meticulous, scientific one to achieve results of scant significance.*
>
> (Ginzburg, 1989, p. 124)

A canonical text of the history of science, more in particular of educational research, reads as follows: 'One cannot understand the history of education in the United States during the twentieth century unless one realizes that Edward L. Thorndike won and John Dewey lost' (Lagemann, 2000, p. xi). Apart from whether or not one agrees with this bold claim (see, among others, Depaepe, 2010; Gibboney, 2006; Tomlinson, 1997), one has to admit that the kind of research that uses quantitative, i.e. statistical techniques, has gained most prestige in the 20th century (see, among others, Depaepe, 1993; Wooldridge, 1994; Richardson & Johanningmeier, 1997; Porter & Ross, 2003; Johanningmeier & Richardson, 2008). Various often interrelated factors are responsible for this, such as the belief in and the acceptance of the assumptions of positivism, the institutional growth of the educational market, the so-called scientisation of educational research, the professionalisation and academisation of the training of education(al)ists, the supremacy of meritocratic values in modern societies and the constant need to legitimate these by 'objective' and 'neutral' research. Unlike his colleague Dewey, with whom he worked for more than 40 years at the renowned Teachers College, University of Columbia, New York, Thorndike embraced this 'trendy direction' of educational research. In 1968 Thorndike's biographer admiringly described him as *the sane positivist* (Jonçich, 1968). As a 'cult figure' Thorndike was the sign of the 'new' world with which the old continent could not keep pace: '... while Europeans were exploring the subjective and personal dimensions of experiences – using the eyes and insights of Bergson, Freud and Van Gogh – Americans are keeping

P. Smeyers (✉)
Ghent University, Ghent, Belgium

P. Smeyers, M. Depaepe (eds.), *Educational Research: The Ethics and Aesthetics of Statistics*, Educational Research 5, DOI 10.1007/978-90-481-9873-3_1,

their art representational, their novels realistic, making their philosophy empirical, their historiography scientific, and above all, their psychology behavioral' (Jonçich, 1968, p. 55).

The use of methods of testing and statistics was at the core of this success story, a story based on the unshakable belief that everything can be measured. William McCall (1922) – the residing statistician at Teachers College – immortalised this unbridled trust in quantification with the well-known assumptions (1) 'whatever exists at all, exists in some amount . . .'; (2) 'anything that exists in amount can be measured . . .'; (3) 'measurement in education is in general the same as measurement in the physical sciences (. . .)' (William McCall, 1922, pp. 3–5). On the one hand these assumptions relied to a large extent on Thorndike's educational psychology; on the other hand, once they were made explicit, Thorndike and his followers were eager to adopt them in order to further justify the way they saw research; as the antidote against all societal evils (see, e.g. Travers, 1983).

This is not the first time that the *Research Community* '*Philosophy and history of the discipline of education*',[1] established by the Research Foundation Flanders FWO, Belgium (Fonds voor Wetenschappelijk Onderzoek – Vlaanderen), addresses an area that is paradigmatic for educational research. In both the first (1999–2003) and second (2000–2008) periods, which focused on *Evaluation and evolution of the criteria for educational research*, various positions were scrutinised (see Smeyers & Depaepe, 2003, 2006, and Smeyers, 2008). In the present (third) 5-year period of this *Research Community* (2009–2013), the overall interest is *Faces and spaces of educational research*, which is divided into four subthemes (respectively addressed during the conference in 2009, 2010, 2011 and 2012): the ethics and aesthetics of statistics; the attraction of psychology; institutional space; designs, material culture and the representation of educational research.

The chapters published in this volume were first presented at the 2009 *Research Community* conference in Leuven. Scholars from philosophy and history of education (some of whom are particularly interested in history and philosophy of science) combine their efforts to study statistics as part of both the academic discipline of education and the broader educational context. Statistics are (still) everywhere. Their power and undoubted efficacy in many areas has given rise to the same faith in measurement and metrics. The more statistics we gather, the more we will know. Their use carries with it a number of presuppositions: that reality is being represented, that it can be controlled and the risks can therefore be managed. As case studies, the chapters interpret the ethics and aesthetics of statistics in terms of representation, visualisation and accessibility, the appeal of 'simplicity', of technical languages, numbers, diagrams and pictures, and pay attention to their connection with action plans. At first sight, some of the observations and arguments made by the contributors may give the reader the impression that statistics has only negative connotations and that it should be banned from educational research altogether as its contributions are dubious to say the least, and moreover as one tends to present these results as hard evidence.

This is not what the contributors are trying to say and this is a distorted picture embracing a polarisation that should be opposed. Statistics should neither be seen as

the golden or only road we can follow to understand educational reality nor should their importance be disregarded when the issues that are studied require such an approach. For instance, if one is interested in a phenomenon such as 'bullying' in primary schools, one evidently wants to know how many cases of bullying have been registered. This can be specified further for particular subgroups such as boys and girls, according to age, ethnicity, various living conditions, and so on and so forth. Now it goes without saying that to have an informed estimate of the frequency of the occurrence of a particular problem (as detailed as this can be) is quite essential in educational contexts. Policy needs to take this into account, as it can be an element in the process of determining how relevant the problem is. This not only has implications for what should be done (by whom and at what level), but also for the kind of research that should be carried out (not to mention the quantity of researchers and research funding that should be mobilised). And it is evident that questions about, for example, the relationship between different phenomena in large populations (To what extent is underachievement in school linked to gender, ethnicity, or social class? Which teaching approaches are related to high achievement in reading tests?) – i.e. correlational studies – would certainly require a sample larger than $n = 1$ (see Chapter 6 by Bridges, this volume). Such studies could also point more precisely to what ought to be addressed in this research. They could alert scholars to phenomena that may have gone unnoticed (an example from a different area is, for instance, the prevalence of certain types of cancer in particular geographical areas). Moreover, a 'thick description' of a school or community of the kind normally associated with qualitative research might and perhaps should include quantitative information about the social class make up of the school, the distribution of test scores of pupils, staff student rations, etc. There is, we think, a legitimate place for different kinds of quantitative research within educational research or, even more broadly, within the academic discipline of education (see Chapter 11 by Smeyers, this volume). What we want to underscore here, however, is the more general criticism that can rightly be raised against the exclusive use of randomised field trials and, more generally, experimental or quasi-experimental approaches (by some labelled as the 'Gold Standard'), approaches that have often been used in educational policy to justify certain interventions. There is a plethora of criticism internal to the use of particular statistical techniques, but there is generally a lack of external criticism that takes into account the overall picture of what education and child-rearing should be about. As many of the chapters in this volume show, a crucial element is the way problems are conceptualised. This has far-reaching consequences for the kind of decisions that are taken on the basis of research. This straightforward point is often forgotten when people look at what research 'tells' us.

The contributors to this volume, who are all working in philosophy and/or history of education and who are all particularly interested in philosophical and/or historical aspects of the discipline of education, will point to the lack of appreciation of the relevance of *the concept(s)* and raise questions concerning the application of the research findings (in other words demand attention to the crucial importance of contextualisation). However, even in their own areas, they do not necessarily doubt the contribution statistics can make to particular research questions. This holds even

more for other areas of educational research as long as statistics are not seen as a goal in themselves but as a tool to acquire understanding. Whereas such an approach is gradually winning acceptance within the sociology of knowledge, it is rarely adopted in the historical study of educational institutes including research institutes and universities (see Depaepe, 2010). In our opinion, it is important that this domain should start with statistics so as to acquire a better understanding of the networks that have played a part in the development of various disciplinary matrixes. For example, since the 1970s many educational journals have been the object of qualitative educational research, yet avenues for exploring this further from a quantitative point of view have rarely been mapped. Recently, Tenorth (2010) tried to identify the part played in the 20th century by empirical research in the area of child-rearing and education. Only if such networks of producers of educational knowledge (as well as of those of the gatekeepers and consumers of knowledge) are uncovered, will the cartography or 'social geography' of the discipline of empirical or experimental educational research be shaped from a history of science perspective.

This book explores what made educational researchers dependent on statistics. It deals with topics such as the use of statistics for measuring the prevalence of maltreatment of children, European citizenship and evidence-based happiness, irregular migrants, and for university expansion. The book also explores the drive to boost statistics, which finds its voice in policy initiatives that become slogans and looks at how public opinion polls are used to rationalise political decision making. It questions whether a more limited and modest use can be made of statistics which does not deflect attention away from education's core business and which does not destroy the local practical knowledge that makes the educational area function effectively. The attempts to answer these questions find their expression in 13 case studies from the stance of philosophy and/or history of the discipline of education.

In Chapter 2 David Labaree, explores the historical and sociological elements that have made educational researchers dependent on statistics – as a mechanism to shore up their credibility, enhance their scholarly standing and increase their influence in the realm of educational policy. He begins by tracing the routes of the urge to quantify within the mentality of measurement that arose in medieval Europe and then explores the factors that have pressured disciplines and professions over the years to incorporate the language of mathematics into their discourses. In particular, this pattern has been prominent for domains of knowledge and professional endeavour whose prestige is modest, whose credibility is questionable, whose professional boundaries are weak and whose knowledge orientation is applied. The chapter shows that educational research as a domain – with its focus on a radically soft and thoroughly applied form of knowledge and its low academic standing – fits these criteria to a tee. It then examines two kinds of problem that derive from educational researchers' seduction by the quantitative turn. One is that this approach to educational questions deflects attention away from many of the most important issues in the field, which are not easily reduced to standardised quanta. Another is that by adopting this rationalised, quantified, abstracted, statist and reductionist vision of education, policy-makers risk imposing reforms that will destroy the local practical knowledge that makes the ecology of the individual classroom function

effectively. Quantification, Labaree argues, may be useful for the professional interests of educational researchers but it can be devastating in its consequences for school and society.

In Chapter 3, Marc Depaepe deals with a report published in 1964 where several Flemish intellectuals argued that, 'in order to enhance the quality of university education as much as possible', undergraduate university campuses had to spread out geographically. This would not only reduce the deficiency of certain areas as regards university recruitment but also the social backwardness that was accompanied by this deficiency. Anyone who observes the development of university education over the past 50 years will quickly conclude that this did not occur as directly as the proponents of the 'dissemination of undergraduate education' had hoped. This 'university expansion' – which was accomplished in two phases: in 1965 and in 1971 – was accompanied by a spectacular growth in the number of university students, though that is not to say that this was synonymous with 'democratisation'. Not only did the 'massification' of university education appear to be due more to the success of the traditional campuses, but sociological research also raised doubts about the intended social effect, since children of the less educated made much less use of the university 'expansion' than did the children of the more educated. Historians do not have much difficulty with the plausibility of this last conclusion. That education, as a social institution, primarily bears a 'bourgeois' and 'meritocratic' character has long been recognised in the history of education. Still, the conclusions of the sociological research referred to cannot claim, historically, to be much more than hypothetical and/or heuristic. They are interesting preliminary studies that, with regard to the problem of the democratisation of university education, need both a cultural historical interpretation and more pertinent (*in casu*, primary) source material – a thesis that is further elaborated in the chapter.

In Chapter 4, Jeroen Dekker refers to the fact that after the publication in 1962 of the article on the *Battered Child Syndrome* by the American medical doctors C. H. Kempe, F. N. Silverman and their colleagues (Kempe, 1962), numerous studies were published on abused and neglected children. Moreover, an increasing institutional and legal framework of diagnosing and preventing child maltreatment was set up in many countries in the Western world. In this chapter, the question that is asked concerns whether this increasing world-wide interest in the maltreatment of children resulted in a major diminution of child maltreatment or not. Although the hypothesis that maltreatment of children was diminishing, at least in the economically prosperous Western world, looks strong, the opposite seems to be true when looking more thoroughly into the information available. This chapter focuses on statistical studies on the prevalence of maltreatment of children in the Western world paying special attention to the USA and the Netherlands. Part of the answer to the question relates to the multiplier effect of three phenomena: a broader definition of child maltreatment since the 1970s; the impact of internationally accepted children's rights by prescribing criteria for good parenthood and for child protection; finally, the preference of policy-makers for clear figures or for the aesthetics of statistics, in developing child protection policies – and this notwithstanding the fact that they are confronted with sometimes contradictory figures.

The chapter concludes by maintaining that these three factors vastly contributed to the increase of the reported prevalence of child maltreatment in the Western world.

In Chapter 5, Norbert Grube claims that social sciences create phenomena. Opinion polls construe the formation of public opinion by selective questions within questionnaires. John Dewey saw social psychological analyses as a means to transfer American mass society into the Great Community, whereas Walter Lippmann wanted to put social sciences into the service of the government. The founding fathers of polls in the USA, especially George H. Gallup, referred implicitly to Dewey and regarded polls as a science that would strengthen democracy. This sentiment echoed in the sentiments of Elisabeth Noelle-Neumann and her husband Erich Peter Neumann, the founder of the Allensbach Institute of public opinion research in 1947, which is a central focus of this chapter. Polls do not only aim to establish and improve democracy. They also aim to create national conformity. The nation stands in the focus of polls, especially if national governments are the customers of pollsters. The presentation of data suggests clarity within the findings along with the possibility of reliable predictions, certainty, comparison and competition. The simplification of data in a few charts shall help to rationalise the political decision-making process for a preventative population policy. At the same time, the graphical simplification should support notions about the social and national body that are based on ideals of unity. Furthermore so-called clear diagrams and findings of applied social surveys should be the convincing starting point for political campaigns and for people's instruction. The chapter reveals different models of dualistic questions and dichotomous models of graphical designs. This dualism does not only divide the people into two camps but establishes a new national narrative. The exact and suggestive presentation of findings led to several attempts to instruct and convince the West Germans of the veracity of capitalism, the democratic system and the European unification. But because polls often present the respondents as incomplete human beings, it is questionable if polls could redeem the aim of preventative governmental policy or whether they could engender a democratic community based on communication.

In Chapter 6, David Bridges argues that individual cases can and do have enormous rhetorical and motivational power in public policy debate: at the same time contemporary discussion of 'evidence-based policy' tends to push the individual case study to the margins of policy-makers' interest, which is focussed instead on population studies and randomised controlled trials in which large numbers provide the appearance of validity and a sense of confidence in the results. This chapter seeks to examine the role of the single case (and by extension small numbers of cases) in informing educational policy and practice, asking how, why and under what conditions *should* educators pay attention to such research. It begins by looking at the reasons for the predilection for large numbers in quantitative research, but also the role of the single case in these research traditions in challenging generalisations and inviting closer examination of the particular context in which apparently aberrational results are observed. This chapter then observes the significance of the 'case' in a variety of academic communities – in psychology and the development

of psychoanalytic theory; in ethnography; in auto/biography and life history; in history itself and in law. It identifies the main characteristics of case study, at least in the forms that are most familiar to educational enquiry. A central part of the chapter examines some of the arguments around the possibility of generalising from individual cases starting with the view that educational conditions are so locally defined by time and social context that such generalisation (from any form of research) is doomed from the start. It considers three particular functions of the case study: (i) as a source of conjecture (and grounded theory) and refutation; (ii) as a basis for understanding one particularity by reference to another (without any attempt to generalise); and (iii) as providing a vicarious form of experience and making a contribution to the reader's practical wisdom. A final section of the chapter considers the sense in which the conduct of case study research resembles scientific or artistic approaches to understanding. Following MacDonald and Walker it concludes with the notion that 'case study is the way of the artist, who achieves greatness when, through the portrayal of a single instance locked in time and circumstance, he communicates enduring truths about the human condition'.

In Chapter 7, Elias Hemelsoet argues that for various reasons, irregular migration has become a more frequent phenomenon during the last decades. Humanitarian and social problems related to this fast-growing group of people give a boost to political discussion and subsequently to scientific research. Both politicians and scientists want to 'grab' the situation and acquire an overview of what is happening. Estimating the number of irregular migrants in a country is in most cases the base to deal with emerging problems. A large number of different methods are used to make these estimations. Unfortunately, these bring forth rather feeble results, which is partially a consequence of a large margin of error. Nevertheless, this does not stop people from basing policy with far-reaching implications on these 'findings'. The chapter focuses on a methodological question concerning estimations of irregular migrant numbers, i.e. the problematic character of conceptualisations. This problem will be treated in two ways. First, different definitions of people without legal residence lead to specific ways of conceptualising the problem. This is problematic as different statistical outcomes are generated which will lead to multifarious conclusions and recommendations. Second, there is the intrinsic danger of reducing the complexity of a large amount of data to a limited number of variables by taking away the possibility of making distinctions that are often desirable (which unavoidably implies a tendency towards homogenisation). In the case of irregular migrants, their illegal residence is the benchmark of categorisation. The level of diversity regarding the particular circumstances of these people is often ignored as policy treats them as a monolithic group. To conclude, Hemelsoet considers the extent to which questions put forward by policy-makers and researchers 'make sense' when it comes to dealing with the problems at stake. Why are these questions so attractive and apparently self-evidently valid?

In Chapter 8, Naomi Hodgson focuses on the objects of statistical analysis that provide a current focus for measurement and policy-making in Europe and thereby constitute key indicators according to which states, institutions and individuals become measurable, comparable and governable. She focuses, in particular, on the

concern with happiness and well-being and how this relates to the construction of 'active citizenship'. With reference to policy literature relating to the development of measures of active citizenship, European surveys addressing citizens in terms of their happiness and well-being, and school curriculum recommendations on 'Social and Emotional Aspects of Learning', the chapter seeks to illustrate how we are asked to account for ourselves according to a particular language of active citizenship. It is suggested that attention to the technologies according to which statistical knowledge operates, in light of the increasingly complex relationship between the governmental and the private (commercial) sources of knowledge production, is necessary for understanding how we are made subjects, and what 'citizenship' means, today.

In Chapter 9 Ulrike Stadler-Altmann and Edwin Keiner focus on the contexts empirical educational research knowledge is embedded in. They consider how such knowledge becomes the object of processes of de- and re-contextualisation and functions in accordance with the expectations of various social groups. On the one hand, they identify an aesthetics of educational research knowledge and the rhetorically persuasive power of figures and graphic accounts (in other words an aesthetics of the way 'answers' are provided). On the other hand, they draw attention to an aesthetics of deconstructive scepticism and of epistemological relativism, i.e. an aesthetics of raising questions. The combination of both aesthetic forms, the provocation of both giving answers and raising questions in a research presentation formatted in a particular way, is related to two further aspects of aesthetics that the chapter goes on to address. They describe these aspects as a 'more rhetorical' and a 'more epistemological' aesthetics. Their analysis deals with (a) (oral) presentations of educational research projects which are examined in a micro-analytical way and (b) (written) publications addressing the knowledge society or educational research outcomes based on large-scale assessment studies, especially those carried out by the OECD. These written publications are approached from a macro-analytical perspective. The chapter offers a comparison of a poem with a presentation of educational research. It focuses on the role figures and statistics play in the rhetorical aspect of an exposition of research results. This amounts to a 'grammar' of presentation.

In Chapter 10, Jean Paul Van Bendegem, Karen François and Kathleen Coessens interpret the ethics and aesthetics of statistics in terms of representation, visualisation and accessibility. A specific case study is briefly examined, namely the development of ISOTYPE (International System Of TYpographic Picture Education) by Otto Neurath, as an attempt to represent statistical data in such a way that any citizen could have access to it. They extend this case study into a set of observations concerning the use of diagrams and pictures in, for example, mathematical reasoning and the semiotic perspective that allows one to connect Neurath's work to, for example, C. S. Peirce's approach. In the second part of the chapter they move from the accessibility problem to the action problem: what to do with statistical data and how to connect this data with action plans? It is clearly not enough to 'merely' understand the data if it does not allow the user to transform it into (justified) decisions. An exemplar is presented, namely whether or not, having listened to the weather forecast, one should take one's umbrella, an issue that highlights intrinsic

difficulties. This analysis allows the authors to extend these questions and problems to the present-day status of the concept of 'statistical literacy' as put forward in the PISA reports, where it is clear that the ethical (and political) dimension play(s) an important role. Perhaps the most important conclusion is that the ethical and aesthetic dimensions are present throughout the whole process from the generation of the data to their societal use.

In Chapter 11, Paul Smeyers starts from the observation that statistics are everywhere but that the use of them carries with it a number of presuppositions which should not necessarily be taken for granted. One presupposition is that it is possible to represent reality. Another is the possibility of control. The attraction of statistics lies at the same time in its simplicity as well as in the belief that goes with it that it is thus possible to characterise and control reality. The chapter focuses on the question 'Why we are so eager to turn to one or other kind of statistics when trying to understand and deal with particular social practices'. Smeyers argues that philosophical problems about the structure of language (the particular metaphysical enframing (and longing) that unavoidably seems to take place), haunt us. As a result, the ideals of objectivity (bracketing the performative embeddedness) and rationality that particularly since the Enlightenment characterises our understanding of reality may be seen as emblematic of our unwillingness to live with complexity. Humans do not only long for knowledge they also seem to have an insatiable need to control. Smeyers argues that we always use concepts that invoke something general and that there is no alternative to this even if we take this generality fully into account and desperately try to avoid the victimisation of our prejudices. With the help of crime stories the chapter argues that statistics may not necessarily be the wrong road to take. Their attraction lies in the fact that they make things more simple and answer a 'human all too human need' to have some kind of grip on reality. This too is all about rhetoric and its argumentative force, something that may be easier to exploit now than ever before given the availability of super computers and web-based dissemination of what has been found to be the case.

In Chapter 12, Ian Munday discusses how statistics and the particular kind of discourse that emerges around them serve to suture the wounds in the discourse of effectiveness culture. He begins by showing how performativity operates within the British secondary school and the role statistics play there. Lyotard, who coined the term 'performativity', believed that the only resistance to 'effectiveness' was to turn to absence and silence. Following Gordon Bearn's critique of Lyotard, Munday argues that the latter's vision is hopelessly pessimistic. He therefore turns to Derrida's philosophy of language. Derrida offers a more optimistic metaphysics and his discussion of difference and iterability frames the discussion of statistics that follows. Munday begins with a discussion of numbers as particular kinds of word. Though numbers are, in a sense 'iterable' they are not iterable in a way that is comparable to other words. Indeed, in certain instances when numbers become iterable, when meaning is artificially conferred upon them, they seem to exemplify the false metaphysics of presence that Derrida's philosophy undermines: they become idealised forms of Saussure's linguistic sign which have no equivalents in ordinary language. Munday develops this train of thought in relation to statistics. Statistics

appear as surface signifiers of underlying truth. However, there is another dimension to statistics whereby they also present the promise (or threat) of absolute scepticism. This paradox serves to suture over linguistic slippages/bleeding within the discourse of performativity. It 'legitimates' the farcical 'fixing' of statistics that goes on in British schools. That said, statistics offer a limited form of power. Statistics like all numbers can never operate alone. In British and American education, the drive to boost statistics finds its voice in policy initiatives that become slogans (such as No Child Left Behind). However much slogans (that become mantras) may suture over linguistic instability/creativity this can never be wholly successful. It is within the ordinary/extraordinary operations of language that hope lies.

In Chapter 13, Richard Smith argues that the power of statistics and epidemiology, and their undoubted efficacy in many areas of the physical and human sciences, has given rise to what he calls 'metricophilia', the obsessive devotion to and faith in measurement and metrics of various kinds. He analyses three prominent texts from the year 2009: a newspaper article by Richard Layard, an advocate of measuring happiness as the standard of social progress, the report of the Stiglitz Commission, and Wilkinson and Pickett's book, *The Spirit Level: Why More Equal Societies Almost Always Do Better*. These texts, and the research on which they draw, are in different ways concerned with the failure of increased affluence to bring greater happiness or well-being to the developed world. Smith argues that the metricophilia of these texts and research projects has the effect of levelling and homogenising and requires us to think of all variations of self-esteem, happiness or well-being as essentially forms of the same thing. It leads to over-simplification and reductionism, ignoring or under-valuing crucial philosophical questions in the faith that better (and more) metrics and statistics, will tell us all that we need to know. If we need to grasp the causes of unhappiness, self-esteem and so on, it is a deeper philosophical understanding of these concepts that is needed rather than the drawing of more correlations. Achieving such understanding is an alternative to the search for the bubble referred to in the title. The texts approached in this chapter are, in different ways, coloured by rhetorical techniques and devices that are at odds with the dispassionate objectivity that statistics and science more generally have always offered.

In Chapter 14, Paul Standish argues that a sometimes neglected aspect of Thoreau's *Walden* is his examination of the various ways in which practices of accounting come into our lives and the credibility these have. This raises the question of the value and the limits of numerical accounting, including questions to do with what can and cannot be measured. It is important that Thoreau avoids both excess of faith in numbers as well as mystification of 'what cannot be measured'. Standish's discussion broaches these matters by considering the aesthetic appeal of technical languages and of numbers themselves. It offers a theoretical background in terms of Heidegger's critique of Leibniz's 'principle of reason', which in turn leads back to Thoreau. Some conclusions are drawn in terms of the ways the prominence of statistics and of numerical measures more generally cause some aspects of the world to withdraw, distorting the validity of the accounts we can give. The place of statistics in educational practice is emphasised, and a right assessment of this needs to be achieved, it is claimed, in the context of a different economy of living.

Note

1. For further information about previous work of the *Research Community*, see Smeyers (2008).

References

Depaepe, M. (1993). *Zum Wohl des Kindes? Pädologie, pädagogische Psychologie und experimentelle Pädagogik in Europa und den USA, 1890–1940*. Leuven: Leuven University Press.

Depaepe, M. (2010). Stand und Probleme der Historiographie der empirischen Erziehungswissenschaft. Metareflexionen über eine zwanzig Jahre alte Forschung als Einführung in die Tagung 'Geschichte der empirischen Pädagogik/Erziehungswissenschaft'. In C. Ritzi & U. Wiegman (Eds.), *Beobachten, Messen, Experimentieren. Beiträge zur Geschichte der empirische Pädagogik* (pp. 13–50). Bad Heilbrunn: Klinkhardt.

Gibboney, R. A. (2006). Intelligence by design: Thorndike versus Dewey. *Phi Delta Kappan, 88*(2), 170–72.

Ginzburg, C. (1989). *Clues, myths, and the historical method* (J. Tedeschi & A. C. Tedeschi, Trans.). Baltimore, MD: Johns Hopkins University Press. (Original work published in Italian in 1986).

Johanningmeier, E. V., & Richardson, Th. (2008). *Educational research, the national agenda, and educational reform. A history.* Montgomery, AL: Auburn University.

Jonçich, G. (1968). *The sane positivist. A biography of Edward Lee Thorndike.* Middletown, CT: Wesleyan University Press.

Kempe, C. H., Silverman, F. N., Steele, B. F., Droegemuller, W., & Silver, H. K. (1962). The battered child syndrome. *Journal of the American Medical Association*, 181, 17–24.

Lagemann, E. C. (2000). *An elusive science. The troubling history of education research*. Chicago: University of Chicago Press.

McCall, W. (1922). *How to measure in education?* New York: The Macmillian Company.

Porter, Th. M., & Ross, D. (Eds.). (2003). *The Cambridge history of science. Vol. 7: The modern social sciences*. Cambridge: Cambridge University Press.

Richardson, Th., & Johanningmeier, E. V. (1997). Intelligence testing. The legitimation of a meritocratic educational science. *International Journal of Educational Research, 27*(8), 699–714.

Smeyers, P. (2008). Afterword. In P. Smeyers & M. Depaepe (Eds.), *Educational research: The educationalization of social problems* (pp. 227–237). Dordrecht: Springer.

Smeyers, P., & Depaepe, M. (2003). Introduction. In P. Smeyers & M. Depaepe (Eds.), *Beyond empiricism. on criteria for educational research* (pp. 9–23). Leuven: Leuven University Press.

Smeyers, P., & Depaepe, M. (2006). Introduction. On the rhetoric of 'what works'. Contextualizing educational research and the picture of performativity. In P. Smeyers & M. Depaepe (Eds.), *Educational research: Why 'what works' doesn't work* (pp. 1–16). Dordrecht: Springer.

Tenorth, H.-E. (2010). Forschungspraxis in der Konstruktion der Erziehungswissenschaft. In C. Ritzi & U. Wiegman (Eds.), *Beobachten, Messen, Experimentieren. Beiträge zur Geschichte der empirische Pädagogik* (pp. 195–210). Bad Heilbrunn: Klinkhardt.

Tomlinson, S. (1997). Edward Lee Thorndike and John Dewey on the science of education. *Oxford Review of Education, 23,* 365–383.

Travers, R. M. W. (1983). *How research has changed American schools: A history from 1840 to the present*. Kalamazoo, MI: Mythos Press.

Wooldridge, A. (1994). *Measuring the mind. Education and psychology in England, c. 1860–c. 1990*. Cambridge: Cambridge University Press.

Chapter 2
The Lure of Statistics for Educational Researchers

David F. Labaree

> *Philosophy is written in this grand book, the universe, which stands continually open to our gaze, but the book cannot be understood unless one first learns to comprehend the language and read the letters in which it is composed. It is written in the language of mathematics, and its characters are triangles, circles, and other geometric figures without which it is humanly impossible to understand a single word of it; without these, one wanders about in a dark labyrinth.*
>
> Galileo quoted in Crosby (1997, p. 240)

During the course of the 20th century, educational research yielded to the lure of Galileo's vision of a universe that could be measured in numbers. This was especially true in the United States, where quantification had long enjoyed a prominent place in public policy and professional discourse. But the process of reframing reality in countable terms began eight centuries earlier in Western Europe, where it transformed everything from navigation to painting, then arrived fully formed on the shores of the New World, where it shaped the late-blooming field of scholarship in education. Like converts everywhere, the new American quantifiers in education became more Catholic than the pope, quickly developing a zeal for measurement that outdid the astronomers and mathematicians that preceded them. The consequences for both education and educational research have been deep and devastating.

In this chapter I explore the historical and sociological elements that have made educational researchers dependent on statistics—as a mechanism to shore up their credibility, enhance their scholarly standing, and increase their influence in the realm of educational policy. I begin by tracing the roots of the urge to quantify within the mentality of measurement that arose in medieval Europe and then explore the factors that have pressured disciplines and professions over the years to incorporate the language of mathematics into their discourse. In particular, this pattern has been prominent for domains of knowledge and professional endeavor whose prestige is

D.F. Labaree (✉)
Stanford University, Stanford, CA, USA

P. Smeyers, M. Depaepe (eds.), *Educational Research: The Ethics and Aesthetics of Statistics*, Educational Research 5, DOI 10.1007/978-90-481-9873-3_2,

modest, whose credibility is questionable, whose professional boundaries are weak, and whose knowledge orientation is applied. I show that educational research as a domain—with its focus on a radically soft and thoroughly applied form of knowledge and with its low academic standing—fits these criteria to a tee. Then I examine two kinds of problems that derive from educational researchers' seduction by the quantitative turn. One is that this approach to educational questions deflects attention away from many of the most important issues in the field, which are not easily reduced to standardized quanta. Another is that by adopting this rationalized, quantified, abstracted, statist, and reductionist vision of education, education policymakers risk imposing reforms that will destroy the local practical knowledge that makes the ecology of the individual classroom function effectively. Quantification, I suggest, may be useful for the professional interests of educational researchers but it can be devastating in its consequences for school and society.

2.1 The Roots of Quantification

Alfred Crosby (1997) locates the roots of quantification in Western Europe in the 13th century. What it gradually displaced was a worldview without standardized modes of measurement, which he labels the 'venerable model.' From the perspective of this model, the world was heterogeneous, where differences were qualitative rather than quantitative and thus reality could not be reduced to common units of measurement. Fire rose and rocks fell because that was their nature, with fire returning to the sphere of fire and rocks to the sphere of earth in a four-sphere universe where air and water separated them from each other. Measuring the distance between spheres was as nonsensical as measuring the distance between God and man. Time also had a qualitative character. Years before and after the birth of Christ could hardly be measured in the same manner. And since Jesus said the day had 12 h, the length of the hour shrank as the days got shorter in the fall and then stretched in the spring. Space likewise expanded and contracted in response to spiritual importance, with maps depicting Jerusalem at the center of the world and showing east toward the top because that was the direction of Eden, toward which the world was 'oriented.' It followed naturally that the size of people in a painting was a function of their importance rather than their location in the foreground or background of the scene; saints were big no matter where they stood in the frame. Numbers were difficult to work with, since they were recorded using the first letter of the Latin word for each quantity, which meant that quantities were words and formulas were sentences.

Crosby says that quantification arose in Europe because of efforts by ordinary men to solve practical problems. Leaders and theorists were opposed to viewing the world in standardized measures, so applying the language of mathematics was a task for less elevated folk. Factors like growing trade, increased travel, and an emerging cash economy urged the process forward. As time started to become money, merchants called for reliable measures of distance, time, and accounts, which pushed sailors to develop new measures of navigation, mechanics to develop clocks, and

merchants to develop double-entry bookkeeping. They needed to keep accurate accounts, and they needed figures that could be easily manipulated, so Arabic numbers gradually made headway. But at the heart of this process, according to Crosby, was a fundamental shift in mentality toward thinking of the world in quanta. This was possible largely because of the extreme marginality and backwardness of Western Europe's intellectual and cultural life, as compared to that of the great contemporaneous civilizations of Islam, India, and China. Lacking a centralized state and an intellectual canon, people were relatively free to tinker with measurement for purely practical purposes. The quick spread of factors like cash transactions and church-tower clocks began to educate the populace in a new quantitative world in which things could be measured in fixed units.

2.2 The Adoption of Statistics by Professions

Theodore Porter (1995, 1986) picks up the story in the 19th century, exploring how professions became quantifiers. He argues that what drove the professions to adopt quantification was a growing set of challenges to their professional authority. As a technology of distance, quantification allows a professional community to make arguments that carry weight and establish validity beyond a particular time, place, and community of authorship. Its emphasis is on impersonality (Porter, 1995, p. ix). And as a technology of trust, it helps the profession gain the confidence of key actors in government, courts, and economy, who are seeking objective reasons for the choosing to follow professional advice (p. 225). The move toward quantification, he shows, was not the preferred option for most professionals. Left to their own devices, professional groups over the years have generally chosen to establish their authority through consensus within the professional community itself. But this approach only works if outsiders are willing to cede a particular area of expertise to the profession and rely on the soundness of its judgment.

Democracies in particular are suspicious of claims of elite authority, unwilling to bow to such claims as a matter of professional judgment without an apparently objective body of evidence that establishes their independent credibility. The predisposition toward objectivity that comes from numbers is a natural extension of the concept of a rule of law not men, relying on universal rules rather than personal preferences. And actors such as government bureaucrats in a democracy are especially prone to seek quantitative data to support policy actions because their own status and authority are open to question (p. 8). The United States embraced numbers early in its history for political and moral reasons as well as concerns about elite authority. The decennial census was a central mechanism for establishing the legitimacy of representative government (so that congressmen represented equal numbers of citizens), and in the early 19th century numbers became a means for assessing the state of public morality (through the accumulation of data on pauperism, intemperance, and insanity). By the 1830s, the United States experienced an explosion of the quantification of public data, with the proliferation of statistics societies and quantitative reports (Porter, 1995, pp. 195–197; Headrick, 2000, pp. 78, 87).

For our understanding of the eventual conversion of educational research to the credo of measurement, however, Porter's most salient insight is that the adoption of quantification by a profession is a function of its weakness (1995, pp. xi, 228). If a profession has sufficiently strong internal coherence and high social status, it will assert its right to make pronouncements within its domain of expertise on its own authority. To resort to supporting one's claims with numbers is to cede final authority to others. Only those professions that are lacking in inner strength and outer esteem must stoop to quantify. In particular, Porter notes that the professions and academic disciplines that are most prone to deploying numbers in support of their claims are those whose domain of knowledge is the most applied. Compared with a domain of pure knowledge, where the boundaries between its zone of expertise and the practical world are sharply defined, applied fields find themselves operating in a terrain that is thoroughly mingled with practical pursuits and thus difficult to defend as an exclusive territory (1995, p. 229). Here professionals find themselves subject to the greatest external pressures and the strongest need to demonstrate the credibility of the claims through quantitative means. Such is the terrain of educational research.

2.3 Educational Research as a Soft and Applied Field

American educational researchers in the early 20th century took the plunge into quantification. This was the era of Edward L. Thorndike and Lewis Terman; of the proliferation of intelligence tests and other standardized assessments in schools; and of the development of scientific curriculum, which built on testing to track students into suitable studies. It was the period chronicled by Stephen Jay Gould (1981) in *The Mismeasure of Man* and by Nicholas Lemann (2000) in *The Big Test*. It was when psychology became the dominant discipline within education by embracing quantification more quickly and wholeheartedly than other domains in the field. A useful way to figure out what this happened is to apply to the field of educational research Tony Becher's (1989) analysis of the link between knowledge domains and the social organization of disciplines. In doing so I will draw on my own elaboration of this application to education developed elsewhere (see Labaree, 1998a, 2004, chapter 4).

Becher begins by locating the knowledge domain of individual disciplines on a scale between hard and soft knowledge and between pure and applied knowledge. Then he proceeds to explore the consequences of this knowledge location for the organization of research endeavors within each discipline. In this analytical scheme, educational research is classified as very soft and very applied; and the consequences of this are devastating for the ability of educational researchers to accumulate knowledge, defend it from outsiders, develop a coherent account of the field, build on previous work, and convince policymakers to take their findings seriously.

The difference between hard and soft fields of study is familiarly understood in terms of distinctions like quantitative and qualitative, objective and subjective, and definitive and interpretive approaches. Of course, the hard–soft distinction is

difficult to establish philosophically, since—to pick just one salient example—quantitative studies have a qualitative grounding in the task of assigning particular qualities to a definitive category for purposes of measurement (coding). But one thing we do know for sure, and that is that the so-called hard sciences have an easier job of establishing their claims of validity than their soft knowledge colleagues. In this polarity, the knowledge that educational researchers develop is perhaps the softest of the soft fields, which prevents them from being able to make strong claims about their findings. Like other social science domains, they have to deal with willful actors whose actions do not have the predictable trajectories of billiard balls, and they have to understand these actions within the setting of an extraordinarily complex organizational setting. But in addition, educational researchers are stuck with understanding matters like teaching and learning within an organizational structure that is both loosely coupled (between vertical layers in the system and between units in each layer) and deeply nested (with students nested in classrooms nested in schools nested in school systems, and so on). As a result, it is very hard to generalize across particular contexts, and educational findings customarily contain a pronounced *ceteris paribus* clause: this causes that only if all the other variables are equal. And how often is that condition actually met?

The difference between pure and applied fields of study is generally understood in terms of distinctions like theoretical and practical, intellectually oriented and application oriented. Of course, every pure discipline has claims to usefulness and every applied field as aspirations to theory. But the key difference in practice is that pure disciplines are more in control of their intellectual work and less dependent on context. They are able to pursue issues that proceed from their theories and see where they lead. But applied fields are stuck with the task of pursuing whatever their knowledge domain demands. For educational researchers this means tackling the problems that arise from the professional problems of schooling the populace. Instead of exploring issues for which their theories might be effective, they have to plunge into problems for which their theories and methods are inadequate. Unlike researchers in pure domains, educational scholars have to do what is needed rather than what they are good at.

In making the case for educational exceptionalism, I am naturally somewhat exaggerating the distinctiveness of the educator's situation. After all, social sciences are also confronted with the problem of soft knowledge. But they have the ability to compensate for this difficulty to some extent by shoring up their credibility with a heavy dose of theory. Fields like sociology and psychology can claim that, although they are soft, they are also pure. Likewise, professional domains such as medicine and engineering, like education, are thoroughly applied fields. But they have the ability to compensate for this by stressing how hard the science is that backs up their arguments. They are applied but hard. In contrast, education is distinctive for being in the unwelcome and unnerving position of pursing research in a field that is both highly soft and highly applied. As a result, educational researchers are unable to accomplish things that are within the reach of their colleagues in hard or pure or hard-pure fields. They cannot construct towers of knowledge, building on the work of predecessors, but are stuck with the task of continually rebuilding

their intellectual foundations. They cannot do what they are able to do well but must respond to what is thrown at them. They cannot construct a tight social community around a theoretically focused target, but must disperse themselves across of broad terrain of professional necessity. The boundary between the discipline of educational research and the domains of educational policy and practice is necessarily blurred and easily crossed, which means they lack a distinctive base of knowledge and professional language. As a result, they are broadly vulnerable to kibitzing from outsiders. Everyone is an expert on education, which seems an open book, so the expertise of educational researchers seems like nothing special.

The relentlessly soft and applied character of the knowledge domain that educational researchers occupy helps explain the lowly status of the field within the university; and it also helps us understand why scholars in other academic fields, educational practitioners, and educational policymakers all feel free to offer their own opinions about education without bowing to the authority of its anointed experts. This situation also helps us understand why educational researchers have been so eager to embrace the authority of statistics in their effort to be heard and taken seriously in the realms of university, practice, and policy. If numbers are the resort of the weaker professions, as Porter argues, then educational research has the very strongest of incentives to adopt quantitative methods. So the field has done what it can to cloak itself in the organizational and methodological robes of the hard-pure disciplines. At the most elite universities in particular, educational research has backed away from the applied research role and embraced a more disciplinary approach, styling itself as the discipline of educational studies. It has worked to create large research centers, which mimic the organizational structure of hard science research in an effort to build towers of knowledge. It has developed its own jargon and journals and professional organizations, in an effort to define its boundaries and create a sense of loftiness and obscurity. And it has sought to quantify as much as possible in order to give the field a strong scent of the scientific. It has even invented its own advanced statistical metric for exploring the special challenges of the educational enterprise—hierarchical linear modeling (HLM)—which takes the power of multiple regression and applies it to the Russian doll context of educational practice by framing the analysis of students nested in particular classrooms, schools, and school systems.

So educational researchers have come to the altar of quantification out of weakness, in the hope that their declarations of faith in the power of numbers will grant them newfound respect, gain them the trust of practitioners and policymakers, and enable them to exert due influence in the educational domain. But this 20th-century conversion has not come without cost. I will focus on two major problems that have emerged from the shift to numbers in education. One is that by reducing education to quanta, educational researchers put themselves at high risk of focusing on what they can measure quantitatively instead of what is really important and interesting in their field. Another is that, when drawing on the reductionist map of education provided by quantitative educational research, policymakers are tempted to treat the map as an accurate representation of the field and construct policies from it that, when implemented in the field, will destroy the richly complex and delicate ecology of the classroom.

2.4 One Problem with Quantifying Educational Research: Missing the Point

The first problem with the quantitative turn in educational research is that it provides researchers with a strong incentive to focus on what they can measure statistically rather than on what is important. It is a case of what Abraham Kaplan (1964) called the Drunkard's Search, in which a drunk is looking for lost car keys—not in the dark where he lost them but under the streetlight where he can see better. It allows researchers to be methodologically sophisticated at exploring educational issues that do not matter. Theodore Porter argues that this problem has been particularly pronounced in the United States, where the public policy process unduly demands, and is distorted by, quantitative 'methods claiming objectivity,' which 'often measure the wrong thing.' He goes on to explain:

> As an abstract proposition, rigorous standards promote public responsibility and may very well contribute to accountability, even to democracy. But if the real goals of public action must be set aside so that officials can be judged against standards that miss the point, something important has been lost. The drive to eliminate trust and judgment from the public domain will never completely succeed. Possibly it is worse than futile. (Porter, 1995, p. 216)

As an educational researcher, I have had my own experience with the problem of measuring the wrong thing. This case can serve as a cautionary tale. My first book, which emerged from my doctoral dissertation, was grounded in data with a strong quantitative component. The numbers I constructed in the data-gathering and analysis phase of the study enabled me to answer one question rather definitively. But it took a very long time to develop these quanta, which was particularly galling since it turned out that the most interesting questions in the study were elsewhere, emerging from the qualitative data.

The book, *The Making of an American High School* (1988), was a study of the first public high school in Philadelphia, from its opening in 1838 to its reconfiguration in 1939. Funded as part of a large grant from the National Institute of Education (NIE) to my advisor, Michael Katz, this study called in part for coding data from a sample of students attending the school over an 80-year period and then linking these students to their family's records in the federal census manuscripts, where individuals could be found by name and address. A key reason that NIE was willing to fund this project in 1979 (at $500,000, it was the largest historical grant they had ever made) was the prime place in the grant proposal given to a rigorous quantitative analysis of the relationship between school and work in the city during this period. For the study I selected a sample of Central High School students in every federal census year from 1840 to 1920, a total of 1,948 students. I supervised a crew of 10 research assistants in the tedious task of coding data from school records for these students and then the enormously complex task of locating these students in census manuscripts and city directories for that year and coding that information as well. In the next step, I punched all of the information on IBM cards (those were the days), worked with the data using several statistical packages (especially SPSS and SAS), created a variety of variables, tested them, and finally was able start data analysis. I spent a total of 4 years working on my dissertation, from initial data gathering

to the final manuscript, and fully half of that time I devoted to coding and analyzing my quantitative student data.

These quantitative data proved most useful for examining one issue about high school education in the period, the factors that shaped student attainment. What kind of influence did factors like social class, country of origin, birth order, prior school, and academic grades have on the ability of Central High school students to persist at the school and even graduate? To analyze this issue, I used multiple classification analysis, a form of multiple regression using dummy variables that allows the research to consider the impact of non-interval data such as social class as well as interval data such as grade point average. The dependent variable in the key equations was the chance of graduation (1 = yes, 0 = no) and the independent variables were major factors that might help explain this outcome. The primary finding was that social class had no significant impact on graduation; the graduation rates were comparable for students across the class scale. The only factor that significantly influenced graduation, in fact, was student grades.

This was not an uninteresting finding. It was certainly counterintuitive that class would not be a factor in 19th-century educational success, since it proved to be a very important factor in 20th-century schooling. And I could not have found this out without turning my data into quanta and analyzing them statistically. So in my dissertation I proudly displayed these quantitative data spread across no fewer than 49 tables (Labaree, 1983). As I worked for the next several years to write a book based on the dissertation, I sought to include as much of the number crunching information as I could. This was in line with an old Hollywood adage: If you spent a lot of money on a film, you need to show it on the screen. As a result, I crammed as many tables as I could in the book manuscript and then asked my colleague, David Cohen, to read it for me. His comments were generous and helpful to me in framing the book, but it was his opening line that knocked me back. 'All of these tables and the analysis that goes with them,' he said, 'seem to add up to a very long footnote in support of the statement: Central High School had meritocratic attainment.'

So I had spent 2 years of my life on a footnote. That would not have been such a bad thing if the point being footnoted was the central point of the book. But it was not. It turned out that the most interesting issues that emerged from the book were elsewhere, having nothing to do with the quantitative data I had so laboriously constructed and so dutifully analyzed. The core insight that emerged from the Central case was not about meritocracy but about the way the American high school in the late 19th and early 20th centuries emerged as both object and subject in the larger struggle between the liberal and the democratic for primacy in the American liberal democracy. Established in order to express republican values and promote civic virtue, the high school quickly developed a second role as a mechanism for establishing social advantage for educational consumers seeking an edge in the competition for position. At the same time, that the common school system was established in the 1830s, allowing everyone in the community to gain public education under the same roof, the high school was established as a way to provide distinction for a fortunate few. And when the common schools filled up in the late 19th century and demand for access to high school became politically irresistible,

the school system opened up high school enrollment but at the same time introduced curriculum tracking. This meant that access for the many did not dilute advantages for the few, who ended up in the upper track of the high school where they were on a trajectory for the exclusive realm of the university. So the real Central High School story is about the way schools managed both to allow open access and to preserve exclusive advantage in the same institution. In this way schools were not just responding to social pressures but also acting to shape a new liberal democratic society.

The point of this story for the purposes of the chapter is that statistical work can lure educational researchers away from the issues that matter. In many ways, statistical analysis is compellingly attractive to us as researchers in education. It is a magnet for grant money, since policymakers are eager for the kind of apparently objective data that they think they can trust, especially when these data are coming from a low-status and professionally suspect field such as education. It therefore increases our ability to be players in the game of educational policy, which is critically important in a professional field like ours. At the same time, it enhances our standing at our home institutions and in the broader educational research community, by providing us with the research assistants, doctoral advisees, travel budgets, and prospects for publication that serve as markers of academic status. It provides us with a set of arcane skills in cutting-edge methodologies, which are a major point of professional pride. And once we have invested a lot of time and energy in developing these skills, we want to put them to good use in future projects. The path of least resistance is to continue in the quantitative vein, looking around for new issues you can address with these methods. When you are holding a hammer, everything looks like a nail.

2.5 Another Problem with Quantifying Educational Research: Forcing a Rectangular Grid onto a Spherical World

So quantification is an almost irresistible fact of life in modern professional and disciplinary work. This is especially true in a democracy, and even more so in a democracy like the United States, where both professional and governmental authority are suspect. The bias toward the quantitative is strongest in a professional domain that is low in status and applied in orientation—most particularly in the discounted field of education, afflicted with low status and a radically soft-applied arena of knowledge. So educational policymakers have a preference for data that seem authoritative and scientific and that present a certain face validity. Statistics are best suited to meet these needs, allowing policymakers to preface their proposals with the assertion that ‘research says’ one policy would be more effective than another. Since policymakers want these kinds of data, educational researchers feel compelled to supply them.

This is understandable, but the consequences for educational policy are potentially devastating. One problem with this, as we saw in the last section, is that quantification may deflect researcher attention away from educational issues that are

more important and salient. Another problem, however, is that quantifying research on education can radically reduce the complexity of the educational domain that is visible to policymakers and then lead them to construct policies that fit the normalized digital map of education rather than the idiosyncratic analog terrain of education. Under these circumstances, statistical work in education can lead to policies that destroy the ecology of the classroom in the effort to reform it.

In his book *Seeing Like a State*, James Scott (1998) explores the problems that arise when states instigate social reforms based on a highly rationalized, abstracted, and reductionist knowledge of society. His analysis provides a rich way to understand the kind of damage that educational policy, informed primarily by quantitative data, can do to the practice of schooling. The connection between statistics and the state is both etymological and substantive. The *Oxford English dictionary* (2002) identifies the roots of the word 'statistic' in the term 'statist,' with the earliest meanings best rendered as 'pertaining to statists or to statecraft.' Daniel Headrick (2000, p. 60) shows that statistics first emerged in public view in the 17th century under the label 'political arithmetic,' arising in an effort to measure public health and the mercantilist economy. So over the years, seeing like a state has come to mean viewing society through statistics.

Following Scott, I argue that the statistical view of education takes the form of a grid, which crams the complexities of the educational enterprise into the confines of ledgers, frequency tables, and other summary quantitative representations that are epistemic projections from the center of the state to the periphery of educational practice. Scott notes that part of the appeal of the grid for states is its esthetic: it looks neat, clean, and orderly, like modernist architecture. In part the appeal is power: the ability to impose the unnatural grid on the world is an expression of the state's ability to impose its own order. In part the appeal is rationalization: the grid reflects a plan for how the world should be as the projection of a rational plan instead of as the result of chance, negotiation, or social compromise. And in part the appeal is its utility: the grid allows for easy measurement, identification, and infinite subdivision. The key problem with this attraction to the grid is that policymakers tend to take these representations literally, which leads to reforms in which they seek to impose the rectangular grid of their vision onto the spherical world of education.

The image that comes to mind in thinking of both the utility and the danger of the grid is the Mercator projection (2009) in cartography, which is the example with which Crosby (1997, pp. 236–237) closes his book on *The Measure of Reality*. This map was developed to help solve a particular problem, which was to find a way so that sailors on the open sea could construct the shortest course between two points with a straight line using bearings of a compass. Since such a course in reality would be an arc on the surface of the earth, this posed a real difficulty on a two-dimensional map. Mercator's answer was to make lines of longitude, which converge toward the poles, instead run parallel to each other, thus increasingly exaggerating the east–west dimension as one approached the poles. And in order to maintain the same proportions in the north–south direction, he extended the distance between lines of latitude proportionally at the same time. The result is a map that allows navigators to plot an accurate course with a straight line, which was an enormous practical benefit.

But in the process it created a gross distortion of the world, which has convinced schoolchildren for centuries that Greenland is enormous and Africa is tiny.

Like Mercator's projection, quantitative data on education can be useful for narrow purposes, but only if we do not reify this representation. However, the temptation to reify these measures is hard to resist. So, policymakers routinely use scores in standardized tests as valid measures of meaningful learning in school. They use quantifiable characteristics of teachers (pay, years of graduate education, years of experience, certification in subject) as valid measures of the qualities that make them effective. They use measures of socioeconomic status and parental education and gender and race to 'control for' the effects of these complex qualities on teaching and learning. When they put all of these quantitative measures together into a depiction of the process of schooling, they have constructed a mathematical map of this institution that multiplies the distorting effect of each reification by the number of such variables that appear in the equation. The result is a distorted depiction of schooling that by comparison makes the Mercator projection look like a model of representation. Recall that Mercator only distorted two variables, longitude and latitude, for a real practical benefit in navigational utility; and you can always look at a globe in order to correct for the representational distortion in his map. But the maps of schooling that come from the quantitative measures of educational researchers incorporate a vast array of such distortions, which are multiplied together into summary measures that magnify these distortions. And there is no educational globe to stand as a constant corrective. The only counter to the mounds of quantitative data on schooling that sit on the desks of educational policymakers is a mound of interpretive case studies, which hardly balances the scales. Such studies are easy to dismiss in the language of objective measurement, since they can be depicted

as subjective, context-bound, and ungeneralizable. They can neither be externally validated nor internally replicated. So why take them seriously?

In addition to the danger of reification that comes from the use of statistics for educational policy, there is the danger of giving primacy to technical knowledge and in the process discounting the value and validity of local practical knowledge. Technical knowledge is universalistic, abstracted from particular contexts, and applicable anywhere. It allows us to generalize the real world by abstracting from the gritty complexities of this world according to rationalized categories and quantifiable characteristics that are independent of setting. Its language is mathematics. As a result, technical knowledge is the natural way for the modern state to understand its domain.

In local settings, however, knowledge often takes a less rationalized and more informal character. Understandings arise from the interaction among people, their work, and the contexts within which they live. These understandings incorporate experience from the past, social arrangements that have developed from this experience, and a concern about applying these understandings to recurring problems and serious threats. This kind of knowledge, says Scott, focuses on matters that are too complex to be learned from a book, like when to plant and harvest crops. It arises when both uncertainty and complexity are high and when the urgency to do something is great even though all the evidence is not yet in. It involves not rules but rules of thumb, which means it depends on judgment to decide which rule of thumb applies in a particular case. It relies on redundancy, so there are multiple measures of where things stand, multiple ways to pursue a single end, and multiple mechanisms for ensuring the most critical outcomes, so that life will not depend on a single fallible approach. It is a reliable form of knowledge for the residents of the local ecology because the residents have a stake in the knowledge (it is not 'academic'), it captures the particular experience of the local context over time, and it is the shared knowledge of the community rather than the knowledge of particular experts.

Scott's depiction of local practical knowledge arises in particular from anthropology and the knowledge of peasants tilling the soil in the ecology of a village, but it applies equally well to the work of teaching and learning in the ecology of the classroom. The knowledge required to survive and thrive in that setting is enormously complex, uncertainty is high, rules of thumb (instead of technical laws) dominate, context is everything, and redundancy of measures and actions is essential in order to know what is going on and to avoid doing harm. The craft of teaching in the modern classroom in this sense is similar to the craft of farming in a premodern village. And one key commonality is the incompatibility between the local practical knowledge in this setting and the technical knowledge of the nation state. When the state intervenes in either setting with a plan for social reform, based on a reified quantified vision of how things work, little good can result.

When the state takes the quantified depiction of schooling that educational researchers provide and uses it to devise a plan for school reform, the best we can hope for is that the reform effort will fail. As the history of school reform makes clear, this is indeed most often the outcome. One reform after another has bounced

off the classroom door without having much effect in shaping what goes on inside, simply because the understanding of schooling that is embodied in the reform is so inaccurate that the reform effort cannot survive in the classroom ecology. At worst, however, the reform actually succeeds in imposing change on the process of teaching and learning in classrooms. Scott provides a series of horror stories about the results of such an imposition in noneducational contexts, from the devastating impact of the collectivization of agriculture in the Soviet Union to the parallel effect of imposing monoculture on German forests. The problem in all these cases is that the effort to impose an abstract technical ideal ends up destroying a complex distinctive ecology that depends on local practical knowledge. The current efforts by states across the globe to impose abstract technical standards on the educational village bear the signs of another ecological disaster.

References

Becher, T. (1989). *Academic tribes and territories: Intellectual enquiry and the cultures of the disciplines*. London: Open University Press.

Crosby, A. W. (1997). *The measure of reality: Quantification and western society, 1250–1600*. New York: Cambridge University Press.

Gould, S. J. (1981). *The mismeasure of man*. New York: Norton.

Headrick, D. R. (2000). *When information came of age: Technologies of knowledge in the age of reason and revolution, 1700–1850*. New York: Oxford University Press.

Kaplan, A. (1964). *The conduct of inquiry: Methodology for behavioral science.* San Francisco: Chandler.

Labaree, D. F. (1983). *The people's college: A sociological analysis of Philadelphia's Central High School, 1838–1939*. (Ph.D. dissertation, University of Pennsylvania).

Labaree, D. F. (1988a). *The making of an American high school: The credentials market and the Central High School of Philadelphia, 1838–1939*. New Haven, CT: Yale University Press.

Labaree, D. F. (1998b). Educational researchers: Living with a lesser form of knowledge. *Educational Researcher, 27*(8, November), 4–12.

Labaree, D. F. (2004). *The trouble with ed schools*. New Haven, CT: Yale University Press.

Lemann, N. (2000). *The big test: The secret history of the American meritocracy*. New York: Farrar, Straus, and Giroux.

Mercator Projection Map. (2009). Retrieved October 2009, from http://www.culturaldetective.com/worldmaps.html

Oxford English dictionary. (2002). London: Oxford University Press.

Porter, T. M. (1986). *The rise of statistical thinking, 1820–1900*. Princeton, NJ: Princeton University Press.

Porter, T. M. (1995). *Trust in numbers: The pursuit of objectivity in science and public life*. Princeton, NJ: Princeton University Press.

Scott, J. C. (1998). *Seeing like a state: How certain schemes to improve the human condition have failed*. New Haven, CT: Yale University Press.

Chapter 3
Dazzling Statistics? On the University Expansion in Flanders and the Need for Research into the History of Education that Transcends Quantifying Sociology

Marc Depaepe

> Die Messung selbst konturiert nur einen neuen Datenraum, der erst aus seinem Bezug auf das zu Bemessende interpretierbar wird (. . .) Sie ist nicht einfach das verkürzte Experiment oder die Rohform dessen, was dann die induktiven Verfahren der Naturwissenschaften als objektiv zu kodifizieren meinen
> (Breidbach, 2005, p. 186)

3.1 The Context

The Kortrijk Campus of the *Katholieke Universiteit Leuven*, where I have concentrated my professional activities since 2004, owes its existence to the university expansion of the 1960s (Casselman, 1984). In 1964 a report was signed by more than a hundred Flemish intellectuals – primarily CVP partisans (the then Christian People's Party, which was strong in West Flanders) – it was stated that 'in order to enhance the quality of university education as much as possible and in order to increase the number of university graduates to the full extent that the reserves of the intellectual potential of the nation permit' a geographic spreading of the 'university candidatures' (the university undergraduate structure at the time) was necessary. This would not only reduce the deficit in university recruiting (in West Flemish districts among others) but also the social retardation that accompanied it. Building on the then favourable demographic projections, the same source predicted that the number of Belgian university students – in the 1962–1963 academic year no more than 32,607 (of whom 25,733 were men) – would increase till the end of the decade by almost 50% (*Ruimere kansen*, 1964, p. 12). However, everyone did not recognize the need for university expansion. In particular, leftist intellectuals, who, of course, did not oppose the democratization of university education, were not in favour of expansion. They felt that it would extend the sphere of influence

M. Depaepe (✉)
Katholieke Universiteit Leuven, Leuven, Belgium

P. Smeyers, M. Depaepe (eds.), *Educational Research: The Ethics and Aesthetics of Statistics*, Educational Research 5, DOI 10.1007/978-90-481-9873-3_3,

of the Leuven Catholic University in regions such as West Flanders, which at the time was (and still is) an important recruitment area for the (State) University of Ghent (De Clerck, 1975, 1985). This position was to a large extent shared by the freethinking circles of the Free University of Brussels (Gerlo, 1989, pp. 183–184). In Antwerp, where the Socialist mayor argued for a pluralistic and democratic community university (Craeybeckx, 1962, 1969), the situation was somewhat different. Back in the 1950s there had been strong hopes for the installation of a new state university, partially as a counterweight for the centuries-long presence of the Jesuits in higher education (Lenders, 2002, pp. 182–183).

One who examines the history of the development of university education over the last 50 years will quickly come to the conclusion that, in large measure (for well-founded research is still lacking) it did not directly proceed as the proponents and the opponents of the 'spreading of the candidatures' had hoped or feared. This university expansion ultimately occurred in two phases. The first phase began in 1965 with the creation of candidature campuses in Kortrijk (KULAK) and in Antwerp (RUCA). The second began in 1971 with an analogous formation in Diepenbeek (LUC) and then in Brussels (UFSAL). A superstructure was also created in Antwerp (UIA). University expansion worked out differently than expected. In any event, it brought social problems to the surface in Flanders that, due to the enthusiasm of the 1960s, had not been recognized despite the fact that the spectacular growth of the number of students from the first decade on greatly exceeded the initial estimates of the 'expansion' (Pelleriaux, 1999, p. 3). This increase had less to do with the creation of additional campuses and more to do with the power of attraction of existing universities. It is not the moment to say more about this here.

Instead of going deeper into this matter, let us turn to the current presentation of the problem. If we may believe the newspaper commentators (see e.g. *De Morgen*, 03.01.09 & *De Standaard* 09.09.09), the general public is now wondering whether the enormous expansion of students has not led to a levelling – or even to a decline in quality – of university education. In addition, there are also doubts about the social effects of the democratization of higher education. Allegedly, this democratization seems to have missed its ultimate target, since children of the less educated currently participate two to four times less in higher education than children of the more highly educated. It appears obvious that such a claim, framed by alluring headlines, is fraught with a historicity. However, this is not what I am really interested in here. Rather, my concerns relate to the lack of a multidisciplinary approach to such issues alongside the failure to appreciate, or to appreciate sufficiently, their educational/historical dimension. Due to the fact that sociologists (and this is certainly not to reproach them) have provided the scientific building blocks for this broadly conducted social debate, the generation of statistical data has become the primary methodological tool with which prevailing conceptions are tested for their validity. Indeed, this is particularly pertinent to issues concerning the democratization of higher education because, to my knowledge, there has been no really solid longitudinal research on the phenomenon of levelling. For that matter, where could one obtain a supra-historical standard with which to compare the knowledge levels of various generations of students?

3.2 Problem Statement

The question that the sociologists have posed that I take as my point of departure for historical methodological reflections concerns the social effects of what has gradually come to be called 'the first wave of democratization'. Has the university expansion of, say, 1960 to 1990 led to more social equality or not? At least four studies have attempted to provide a scientifically based answer to this question: (1) that of Barbara Tan (1998) (of the Centre for Social Policy of the University Faculties of St. Ignatius Antwerp published in the *Tijdschrift voor Sociologie*); (2) the study based on it by Koen Pelleriaux (1999, 2001) (available on the Internet; also published as an internal report of the Sociology Department of the Free University of Brussels and also recycled as the first chapter of his PhD thesis); (3) a paper by Steven Groenez (2008) (of the Higher Institute of Labour Studies of the *K.U. Leuven*, presented there on the Sociology Day of 28 May 2008 and also intended as a chapter of his doctorate); and (4) a study (published as a notice and also available on the Internet) by Geert Verbergt, Bea Cantillon, and Karel Van den Bosch (2009) (of the same Center for Social Policy of the University of Antwerp), which was intended to be a kind of replication of Barbara Tan's study.

What these studies more or less have in common is, in addition to their strikingly similar conclusions (more about them later), the empirical numerical data they use and the statistical techniques applied to them. These authors set out, first and foremost, to fill the 'empirical' gap that the existing numerical data have left behind. Indeed, the statistical yearbooks of education, which our research team has assembled into diachronic data series as far as primary (Minten et al., 1991–1996) and secondary education (D'hoker et al., 2006) is concerned, give only the numbers of students, and these numbers do not enable one to determine the social backgrounds of the students. In order to determine the students' social origins, these authors used all sorts of indirect data compiled in surveys of various populations in other sociological research projects. All kinds of advanced statistical techniques were then applied to this (often merged) data. The results did not conform to expected patterns. One might have imagined that the results would reveal a connection between greater opportunities to attend higher education with greater degree of participation among the 'lower' classes. This is not what emerged in the findings.

As regards the evolution of higher education between 1981 and 1991, Barbara Tan stated that, in sharp contrast to the disappearance of sexual inequality, social inequalities remained virtually unchanged: 'Only 16% of the children of less-educated fathers and/or mothers attend higher education. Among children of highly educated parents, this figure is 59%' – a picture that, in her opinion, 'has hardly changed in the last 20 years (1978–1998)' (Tan, 1998, p. 200). She attributes this mainly to the ongoing selection process in primary and secondary education, an issue that has dominated the political agenda of the progressive and democratic forces in Flanders in the past few decades.

As far back as the 1970s, the Socialist ministers of education, in particular, wanted to promote democratization through the introduction of a less selective but more 'comprehensive' educational model in secondary education (see Henkens, 2006). I will return to this point later on, but in the meantime, let us consider Barbara

Tan's final verdict: 'this indicates that the policy that was conducted in the period under consideration with a view to the dismantling of socio-cultural barriers has, in large measure, failed' (Tan, 1998, p. 201). What happened? The concept of the comprehensive middle school, the *nec plus ultra* of all the progressive-pedagogical forces, was rolled back, first by the Liberals and then by the Catholics, who, at the very best accepted a compromise-like '*marche en arrière*' with what was called the 'uniform type' in general secondary education (Henkens, 2004).

Koen Pelleriaux supported Tan's conclusion completely. Even though the number of university students had grown by a factor of 5 in 30 years (and, more specifically, between 1951–1952 and 1983–1984, and the female population by a factor of 11), the correlation between the professional prestige of the father and the educational level of the respondent had, he maintained, remained constant. He writes, 'such a finding places towering question marks about the "democratization" of education in general and of the university in particular. … The laws "concerning various measures for university expansion" passed in 1965 nevertheless explicitly had the intention of giving everyone the opportunity to obtain a university education. Therefore, the candidature years, at the very least, had to be established "close to home". In the light of the data presented here, we must conclude that this 'democratization' of higher education has failed. Many more young people received the opportunity to obtain a university diploma, but the *bias* toward socially better placed groups has remained just as large' (Pelleriaux, 1999, p. 15). For Pelleriaux, this was one more indication that the Flemish society had become very meritocratic after the model of a degenerate meritocracy. Children from well-off families could continue to rely on the advantage of the family fortune so children from economically modest families were not given equal opportunities and this made them very vulnerable. To a significant degree, education reproduced social status.

That Pelleriaux, who later went to work for the study service of the Socialist Party, came to such conclusions – or, inversely, saw precisely in these conclusions an impulse for political involvement – is certainly not a matter of chance. Indeed, the Socialists have made the equal-opportunity policy the spearhead of their education policy in the course of recent years mainly through the efforts of the Flemish minister of education Frank Vandenbroucke. When his chief of staff, Dirk Van Damme, received the opportunity in 2008 to lead the CERI of the OECD in Paris, Pelleriaux succeeded him. He has since become the chief of staff to the present Socialist minister of education Pascal Smet, who recently succeeded Vandenbroucke.

The Higher Institute of Labour Studies – a research institution within the Catholic University of Leuven that has historical roots in the Christian labour movement – has conducted analogous studies. On the basis of these studies, a virtually identical conclusion was reached: 'The great expansion of higher education since the 1960s, therefore, has not led to greater democratization in terms of equal opportunities. The conclusion is clear: more workers' children now go to universities and institutes of higher education than previously. But great expansion of higher education is still not democratization. Other professional groups are progressing just as much. In other words, the position of the workers' children with respect to the others has not improved since the 1960s. The inequality has remained' (Groenez, 2008, p. 10). And here, too, secondary education is painted as the villain: 'The real solution for

the democratization problem in higher education must be sought in secondary education, where alongside the rigid separation between the forms of education, the unequal distribution of the social groups across these educational forms must be tackled' (Groenez, 2008, p. 13).

Finally, the same conclusions are drawn in the study conducted by Verbergt, Cantillon and Van den Bosch. In their update of Tan's data, they observe 'that the link between the educational level of the parents and the participation in higher education between 1992 and 2004 has certainly not weakened. (One might even surmise that there has been a slight increase)' (Verbergt et al., 2009, p. 5). This leads them to the conclusion that educational democratization has been at 'a stubborn "stand-still" for already more than 40 years(!)', which, they argue, is striking and disquieting, 'given the great (and recently sharply increasing) attention for this problem in society and policy making. . ..'. Therefore, 'The future will have to show whether the recent efforts regarding equal educational opportunities will bring about a reversal' (Verbergt et al., 2009, pp. 6–7).

As a historical researcher, I do not immediately have much difficulty in accepting the plausibility of such conclusions. Books about the social-Darwinian nature of the phenomenon of meritocracy and the competitive, *casu quo* selective role that education plays in it fill several bookshelves. That the school as a social institution has a primarily bourgeois character, that it cultivates, transmits and reinforces the language, the norms and the values of the upper classes is a familiar theme in the history of education. Similarly, the fact that the often intensive attempts of policy makers to change anything often gives rise to paradoxical and ironical effects (or side effects?), in part because of the resistance they generate from below, is equally familiar.

Still, to my mind, the conclusions of these social science studies are of a hypothetical and/or heuristic (and therefore provisional) character. In my opinion, they are interesting preliminary studies that, certainly as regards the problem of democratization of higher (and more specifically of university) education, require both more relevant (i.e. primary) source material and a more pluriform (if you will, interdisciplinary, multidisciplinary, or pluridisciplinary) methodological approach. I maintain that this is the case despite the fact that these buzz words have in the meantime been so eroded that they seem to have little meaning. The issue to which it all boils down to is whether statistics, given the generally horizontal approach to social problems, are capable of sufficiently grasping historical questions from a longitudinal developmental perspective.

3.3 Questions from History on the Use of Statistics in Social Sciences

3.3.1 Too General?

Although history, (as a rather narrative or ideographic science) and sociology (as its more theoretical, *in casu*, nomothetic counterpart) have come to show clear signs of convergence since the end of the 1960s, there still remain a considerable number

of methodological bottlenecks in the framework of the desired inter-, pluri- or multidisciplinarity. This is certainly the case when one focuses on concrete research questions and their results as I do in this chapter. The concerns that are being expressed here relate to something I discovered at the beginning of my career. When in 1979 I tried to relate the participation in primary education in Belgium in the 19th century to the degree of industrialization, I used a simple rank correlation coefficient (Depaepe, 1979). This generated harsh criticism from the economic statisticians: at the very least a multiple regression had to be performed in order to investigate the link between one variable to be explained and several explanatory variables. By way of example, the critics made use of a number of dummy variables (these are used to recode some sort of categorial data – most commonly dichotomous variables – in numerical form, see Hudson, 2000, p. 263) in order to be able to cleanse the effect of the industrial horse-power (of the provinces) per child on the participation ratio. Not only did it lead to the standard conclusion that more research was needed, but also – and this was surprising, perhaps, since it came from the statisticians themselves – that the statistical material needed to be supplemented with 'more source criticism and with micro-studies' (Berlage & Standaert, 1981, p. 723).

In my opinion, this conclusion[1] can confidently be applied to the question of the possible success of university expansion. Indeed, whoever considers the four studies cited above can do nothing but conclude that this question was approached in a fairly one-sided fashion, namely in conjunction with its effect on social mobility. Certainly, one of the expectations that emerged from the optimistic years of the 1960s was that a greater supply of university education would help overcome the social disadvantage of the working class. This question was also and primarily formulated in regional terms, which, in addition to political-ideological considerations, gave the impetus to spreading entry opportunities (for example in West Flanders and Limburg). In the studies mentioned, we hear nothing about geographic differentiation. Similarly, an account of the regional effects of the expansion (something that, of course, itself should be placed in a historical perspective) is also absent. What was rural 40 years ago is no longer necessarily so today, and the concepts of city and (im)mobility have since taken on very different meanings. That we learn nothing about this geographical context is bound to the fact that the educational statistics do not permit one to say anything about them at all. However, this is actually not a good argument since the basic data for the hypotheses about the social context of the participants in higher and university education in the studies discussed here were, ultimately, taken from elsewhere.

Moreover, for a good understanding of the effects of university expansion – even the social ones – the categories of 'university' and the various other kinds of 'institutions of higher education' should be clearly distinguished. As general categories, they are, indeed, distinguished from each other in the facts (and in the figures), but there is a tendency in the conclusions to sweep everything together and cover this differentiation. The greatest problem, however, is that the social prestige, let us say, the status of these educational institutions, has not remained the same over the years. Significantly the relations among educational levels have changed profoundly over the last 40 years. First, the ongoing academization of higher education

has probably led to reducing the distance between the university and the other institutions in general. Second, since the introduction of the BaMa decree of 4 April 2003 the differentiations between higher education institutes of the 'long' and the 'short' type have been reformulated in terms of curricula which lead to a bachelor degree with an 'academic' orientation on the one hand and to a bachelor degree with a 'professional' orientation on the other. In concrete terms this means that some of the courses offered in the institutes of higher education with an academic orientation can lead to a master's degree. It is self-evident that these courses were privileged within the recent formation process of 'associations' with universities. In the near future the academization of some of these institutes will definitely result in their being integrated into the universities.

It is just as striking that, in this regard, the evolution of the data on the participation in university and higher education is never related to the evolution of the 'networks'. By 'networks' I mean the pillars or institutional axes around which the school supply is grouped in Flanders, say the public versus the private sector, which obviously has taken on other ideological connotations and denotations over the years. Because of progressive secularization, the mobilizing function that it had in the pre-Conciliar era can obviously no longer be attributed to the Catholic network. Inversely, the ideological labels that were pasted on public education in the course of the 19th century lay little claim to any reality value in a pluralistic society like ours. Nevertheless, the continued institutional existence of these 'networks' seems to have been further secured by the formation of the 'associations'.

Mutatis mutandis, one thing and another also apply to the relationship of the study areas with each other, university or not. Thus, on the basis of my personal experience in conversations with students and colleagues, I find it plausible to assume that the prestige of the arts and humanities in general has diminished (this certainly seems to be the case for young men). Greater prestige is now conferred on the 'hard' sciences (which produce more 'utility' and the associated prospects of a job, money, etc.). Such matters are bound up with the 'economization' of (higher) education. Here, I am not simply referring to 'training' and the greater attunement of job markets but also discourses relating to education and the 'knowledge economy'. This discourse is intrinsic to the notion that is sometimes put forward, namely, that the university and higher education in general are in opposition. Studying industrial engineering, to take just one field, promises to be more lucrative than the study of philosophy or history.

In short, the statistics on which the study of the university expansion is necessarily based are, in my opinion, insufficiently differentiated, and therefore seem to exclude a finely meshed combination of possible variables (this is also the case at a statistical level). For example, with regard to gender differences, it is said that women have overcome their relative deficit with respect to men as regards participation in university and higher education in general (see Tan, 1998, pp. 174–176, 200; Pelleriaux, 1999, pp. 3–4). But what does that ultimately mean for these women? In raising this question, I am not simply referring to individual careers but also to the ways women operate within their families and society. The question, therefore, is whether or not a better understanding of things would require

tapping other sources alongside the rather deficient basic statistical material. To what extent can statistics (here considered primarily as cliometrics) as a technique and/or as a method bring the historicity of studied dynamics to the surface? To what extent, in other words, is historical statistics able to take into account the studied processes and the changes in the social context that accompany these processes and/or to keep them under control?

3.3.2 Too A-Historical?

I have already raised the question about the possible a-historicity of some of the indicators and/or parameters used in order to map social status, the degree of education and the relationship between them. Obviously, 40 years ago, graduating from a university or an institute meant a lot more than it does today. Also the completion of secondary education, which was quasi-universalized since the extension of compulsory education in 1983 (De Ceulaer, 1990), had different connotations in 1960 than it does now. At that time, having completed lower secondary education, one could, for example, easily obtain a position as a clerk in some form of public service (see Depaepe et al., 1992). So should one continue to use these educational levels supra-historically in order to measure the degree of schooling or people's social status? Of course, the same question applies to the profession of the father or the first profession of the respondent, which are taken as indicators of the place of the respondents on the social ladder. Professions and their prestige are themselves subject to sociohistorical fluctuations, and the factors that cause these fluctuations, as we have tried to clarify with respect to teaching profession (Depaepe & Simon, 1997), are not only very numerous but do not necessarily operate in the same upward or downward direction of greater or lesser prestige.

Moreover, it is questionable as to whether or not the social dynamic of intergenerational mobility can be grasped with the rough categorization or hierarchization of the professions that are considered in the studies cited here. Such studies do not go much further than the crude differentiation between blue-collar worker and white-collar worker (see Tan, 1998, 177 ff.; possibly 'higher-level white-collar worker', see Groenez, 2008, 3 ff.), supplemented or not with the distinction between self-employed and having no profession (see Verbergt, et al., 2009, 3 ff.). Do these crude categories permit sufficient determination of the group-specific expansion of the participation of the blue-collar workers in the higher educational levels with which the democratization of education is ultimately operationalized (Groenez, 2008, pp. 3–4)? Not only has the category 'blue-collar worker' taken on different meanings over the last few decades but this specific group has also taken on a totally different quantitative form in absolute but also and primarily in relative terms with respect to the middle-class groups. Even if quantitative adjustments can be made by carefully considered statistical operations, it is still much more difficult to make qualitative adjustments.

When approaching the quantitative aspect, Groenez tries to distinguish relative from structural mobility by means of the application of an odds ratio. This yields some fascinating formulas[2] and the construction of gender-differentiated

graphs concerning the relative chance relationships of the workers versus the higher employees in function of diplomas of higher secondary school and higher education.[3] On the qualitative side, Pelleriaux tries to estimate[4] the regression of the first professional prestige of the respondent with respect to that of the father, the number of years the father attended school for and the number of years of schooling of the respondent via a system of structural comparisons between three groups of respondents (before 1955, between 1955 and 1973, and after 1973 on the labour market) on the basis of the maximum likelihood criterion. This yields various mathematical models that can be tested. With two degrees of freedom resulting in chi-square of only 2.005, the most plausible model is the one in which the covariance between the two independent variables has been kept constant over the three periods.[5] If we advocate this model, this leads to the rejection of the hypothesis that the effect of the educational level on professional prestige of the respondent remains constant during the three distinguished periods. The influence of the professional prestige of the father on the professional prestige of the first job of the respondent is not linear. For the last group (after 1973), the professional prestige of the father has no effect at all on the educational level of the respondent. That is not to say that social origin no longer has an influence on the life opportunities of an individual. It means that the influence proceeds via education (Pelleriaux, 1999, p. 14). This, according to Pelleriaux, underlines the social importance of the educational level and more specifically of the cultural dynamics in secondary education and the cultural specificity of educational forms. Together, they constitute, as it were, the new fault lines of the substantial cultural conflict that has emerged within contemporary society (Pelleriaux, 2001, pp. 239–241).

However attractive these points of view may be from the cultural-sociological point of view, a question still remains as to whether or not the empirical evidence once finally estimated can be deemed sufficient. In fact, for a qualitative analysis, one would have to examine the self-definition and the self-situating of workers over time on the basis of historical research. If one takes such an anthropological view, it could appear that statistics relating to the underprivileged of 1960s' Flanders (the notion of being 'underprivileged' did not even exist then) cannot be 'related' to today's situation regarding privilege. In the 1960s, social and cultural differentiation had much less to do with ethnic, religious, linguistic and national substructures than it does today. At that time Flanders was, culturally, still relatively homogenous and was epitomized by a constrained self-absorption. This took precedent over the development of intercultural openness to the world. Thus, my plea for historical contextualization can in no sense be taken as an alibi for one or other reactionary nationalistic discourse.

Let us return to questions surrounding the identification of social groups and the classes that need to be distinguished. Trying to answer these questions by simply looking at the subject's first profession or the profession of their father is highly suspect. Sociological theories such as those put forward by Bourdieu suggest that cultural capital – which can have an influence on the choice of studies – comes primarily from the side of the mother. Although some evidence contradicts this position (Verbergt et al., 2009, p. 4) from a purely theoretical point of view, it seems sensible to also consider the profession or social and cultural origin of the mother.

In the context of democratization, which is naturally conceived as an upward movement (caused in part by the education enjoyed), the process of social mobility is relative to the historical circumstances in which it takes place. This understanding of social mobility has recently been confirmed by the Stanford historian of education David Labaree at a doctoral colloquium in Bern. In his reading of history, the problem posed in this way: 'In the way that education interacts with social mobility and social equity, both of the measures of social position are purely relative. Both are cases of what social scientists call a zero sum game: $A + B = 0$. If A goes up then B must go down in order to keep the sum at zero. If one person gets ahead of someone else on the social ladder, then that other person has fallen behind. And if the social differences between two people become more equal, then the increase in social advantage for one person means the decrease in social advantage for the other. Symmetry is built into both measures' (Labaree, 2009, p. 6).

If, nevertheless, an absolute supra-historical measure is sought, for example, in terms of participation in higher or university education, then one must take the drainage effect in the category of less educated into account. This is due to progressive 'pedagogizing', which obviously results not only in the growth of the middle classes but also in the perception of the school as a socializing or a civilizing institution of, for and promoted by the middle classes. Because the more educated no longer want to be considered manual labourers, increasing modernization (including the accompanying industrialization, bureaucratization and so on) provides them with all sorts of jobs in the middle classes. Above all, children, who are sent to secondary and higher education but no longer count as children or grandchildren of labourers in the statistics, see to it that these middle classes are indeed able to benefit the most from the increasing supply of education. Historically speaking, one might still wonder whether what is involved here are the enduring effects of democratization in the previous generation. This idea does not necessarily have to rest on the somewhat naïve, possibly nostalgic myopia that the analysis of qualitative source material on social progress, intellectual formation and the creation of individual opportunities for development can readily occasion. It can also be conceived as the result of more structural modernizing processes in which such a determinative role need not be ascribed to the school. It is likely that the 'gastropodal' way (Kossmann, 1987) in which the middle groups – here crudely seen as one single middle category – have taken possession of the modernizing society is the result of economic and industrial development that was not set in motion principally by education or by the school. The idea that the school prepares or anticipates a new society seems primarily inspired by an empathetic reading of the progressive reform proposals that were formulated in the sphere of the Enlightenment and the French Revolution (see below). This idea has been popular in educational circles because it valorizes and legitimates the pedagogical trade (e.g. De Keyser, 1986). This does not alter the fact that, in accordance with present-day economic rationality, education can also and primarily be viewed from a market perspective. From that perspective, it is without doubt a sector, which via the creation of all sorts of jobs, has itself contributed to building up the social middle field. I have nothing to say here about the 'genetic' hypothesis that it is impossible for everyone to be piloted to the middle field via this sector. My silence in this matter stems from a wish to avoid being politically and ideologically misunderstood.

In any event, these speculations point to the benefit of approaching complex educational historical processes, such as the democratization of post-secondary and/or university education, from an individual, biographical perspective. That is to say, it is important to account for the way in which people who were subject to these processes actually experienced them. Such speculations reveal an enduring concern for interpreting these same processes and the concepts, terms and notions in a 'cultural historical' manner. Democratization – consider the thematic issue of *Ons Erfdeel* on the 'subdivided' university (De verkavelde universiteit, 1998) – was an aspiration that emerged from diverse backgrounds and was given voice through visions and political agendas in Flanders. Some saw this process as a potential point of contact with the old ideal of *Bildung*, while others saw an opportunity to develop new networks within the knowledge society. Some saw democratization as an opportunity for high-quality higher education that would be averse to provincialism, while others continued to believe in the need for developing regional spearheads. Some took it as a point of departure for the formation of one single University of Flanders, while others continued to defend the need for competition among the large universities (as forerunners of the associations), and so on. In the meantime, from a politically quite understandable concern for the underprivileged, the emphasis in the debate about the democratization of education has shifted imperceptibly to equal educational opportunities. This is a problem that, as I have noted, is situated primarily at the secondary-school level, which brings me seamlessly to the following point.

3.3.3 Presentist Pitfalls?

That one arrives at secondary education from a pedagogical-progressivist vision is no coincidence. With an appeal to the philosophers of the French Revolution, among whom Condorcet translated the social-democratic starting points the most explicitly as regards pedagogy, the struggle for a comprehensive middle school in the second half of the 20th century became the hallmark of virtually all Socialist ministers of education, not only in Belgium (later in Flanders) but also in other European countries (Wielemans, 1986). Apart from the still unresolved philosophical paradox of how precisely this equality must be achieved in an unequal society (see also Schneider, 2010) – the unavoidable Trojan horse of every Socialist policy within a neoliberal (read capitalistic) perspective – there also remains the problem that policy makers in present-day society readily fall back on social science research (and also statistics) to legitimate their political decisions. This would not be so bad were it not for the fact that the users of these scientific results tend to ignore how such results are generated. Perhaps they are dazzled not only by the aesthetic effects of the numerical material produced but also by the compelling ethical norms that are ascribed to their scientific status.

Let us recall that in the case discussed here the basic data about participation in higher, *casu quo* university education, to which the statistics (the idea of historical correction mechanisms) are applied are ultimately only estimates drawn from samples compiled in other contexts (such as survey of the living situations of Belgian households, time-budget surveys, family surveys, election surveys and so on). These surveys were not only conducted at various times but were also applied in

combination. Moreover, Tan's research functions as the mother, as it were, of all the samples that are regularly recovered in the subsequent studies. Thus, on the basis of data of respondents (from diverse studies and contexts), statements are made about previous generations (questions about educational levels and professions of the parents and grandparents). Consequently, the time span under research is considerably expanded. On the other hand, the principle of synchronicity in regard to questioning risks fading into the mist. This is partly the result of continuous interference of the present and the past. Ultimately, a whole range of statistical techniques and operations (ranging from parameter estimations, calculation of chi-squares and so on) are applied to these collections of data. I am too poor a statistician to judge the validity of all these operations. Nevertheless, as a historian of education, I do wonder about the degree to which any account was taken of the quasi-infinite combination possibilities of variables, during the 'fabrication' of the[se] statistics (Popkewitz & Lindblad, 2001). It is obvious here that, from the well-nigh infinite sample of possibilities used (and published?), only those that gave a significant answer to the policy-relevant question were used. The others (those that were not deemed significant) were simply left aside. In association with the – 19th-century history of statistics in Belgium (Bracke, 2008), this numerical material continues to provide the primary building blocks for a discourse that is eagerly taken over by politicians, journalists and social scientists.

Rather than wanting to accuse authors of manipulation – there can be little doubt that their research has scientific integrity – it seems to me that an over-sophisticated technique is set loose on a shaky basic construction of data in all four studies. I have long sought after an image to capture this insight. First, I thought of the proverbial sledgehammer cracking a nut. On further consideration, I think that rather the inverse is the case. Perhaps, a better image would be of a hippopotamus being pushed under the lens of a high-powered microscope. One would be astonished to observe the same flaccid gray tissue everywhere on the amorphous mass. Over a long period of time the belly of its body would be swelling more than the head or the tail . . . In order to see the richly shaded hues of the educational past, one need not try to reinvent the source material. Rather one might search for the material in university archives. There, on the basis of registration lists, real data about the social origin and amount of schooling of the parents are kept. Obviously the construction or reconstruction of such data series would require much time and labour. Therefore, case studies (preferably by a team) should be designed that would be selected on realistic chronological geographical grounds. The quantitative material generated would have to be enriched by qualitative sources (interviews, life histories, ego documents) for the same cases. Indeed, recalling Breidbach, one who wishes to understand the effects of the expansion of the universities in Flanders will have to look through and beyond the numbers.

Notes

1. See also Hudson (2000, p. 211) who concludes as follows: 'in order to create a new monumentum in economic history, it will be important for historians to avoid the mistakes of the past: carelessness with sources and with questioning the data, preoccupation with oversimplified

neoclassical and macro-level theorising, anachronistic application of present-centered assumptions, [. . .]'.

2. See Groenez (2008, pp. 5, 14):

$$OR = \frac{odds1}{odds2} = \frac{\pi_{1j}/\pi_{1j'}}{\pi_{2j}/\pi_{2j'}} = \frac{\frac{n_{1j}}{n_{1+}}/\frac{n_{1j'}}{n_{1+}}}{\frac{n_{2j}}{n_{2+}}/\frac{n_{2j'}}{n_{2+}}} = \frac{n_{1j} \cdot n_{2j'}}{n_{1j'} \cdot n_{2j}}$$

$$\textit{Relatieve rekruteringskracht } j - j' = OR.\frac{1+\frac{\pi_{1j}}{\pi_{1j'}}}{1+\frac{\pi_{2j}}{\pi_{2j'}}} = OR.\frac{1+\frac{n^*_{1j}}{n^*_{1+}}/\frac{n^*_{1j'}}{n^*_{1+}}}{1+\frac{n^*_{2j}}{n^*_{2+}}/\frac{n^*_{2j'}}{n^*_{2+}}} = OR.\frac{1+\frac{n^*_{1j}}{n^*_{1j'}}}{1+\frac{n^*_{2j}}{n^*_{2j'}}}$$

$$\textit{Relatieve rekruteringskracht } j - j' = OR.\frac{1+\frac{\lambda_1\delta_j n_{1j}}{\lambda_1\delta_{j'} n_{1j}}}{1+\frac{\lambda_2\delta_j n_{2j}}{\lambda_2\delta_{j'} n_{2j}}} = OR.\frac{1+\frac{\delta_j n_{1j}}{\delta_{j'} n_{1j}}}{1+\frac{\delta_j n_{2j}}{\delta_{j'} n_{2j}}}$$

3. See Groenez (2008, p. 11):

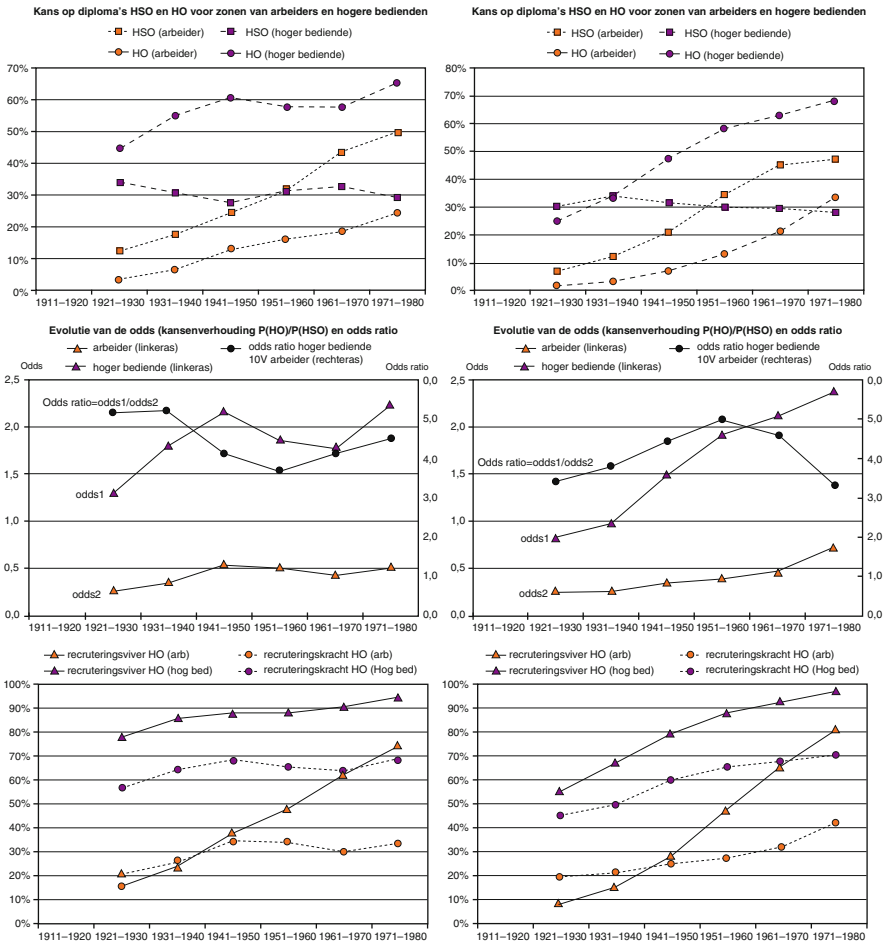


Fig. 7: Constructie van de relatieve kansenverhouding HSO-HO, (mannen, links; vrouwen, rechts)

4. See Pelleriaux (1999, p. 9):

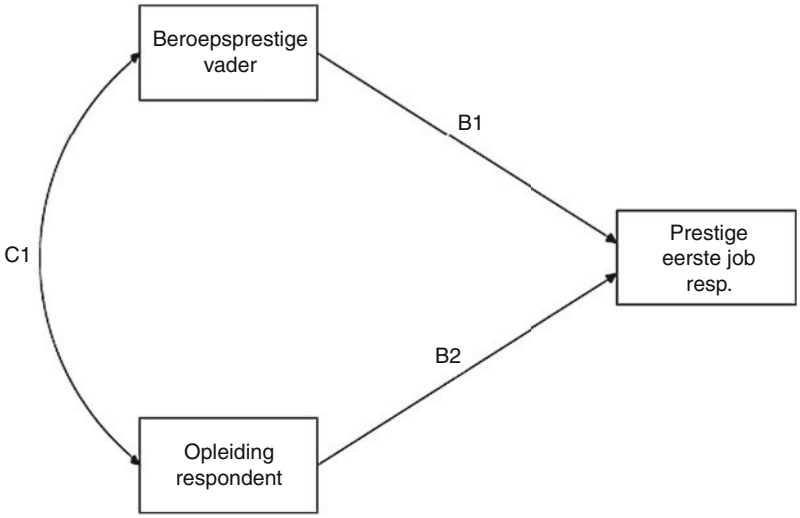


Fig. 4: Regressie van het eerste beroepsprestige van de respondent op het beroepsprestige van de vader en het aantal jaar schooling van de respondent

5. Id., 13: the most acceptable model is the fourth (referring to table 4)

Table 5: Samenvatting: Modelfit voor de vier geschatte modellen

Model	P	χ^2	DF
De drie parameters gefixeerd (Tabel 3)	0,001	22,248	6
Effect beroepsprestige van de vader gefixeerd (Niet opgenomen)	0,001	14,979	2
Effect onderwijs respondent gefixeerd (Niet opgenomen)	0,003	11,783	2
Covariantie tussen onafhankelijken gefixeerd (Tabel 4)	0,367	2,005	2

References

Berlage, L., & Standaert, S. (1981). Deelname aan het lager onderwijs en industrialisatie in België (1846–1911). *Belgisch Tijdschrift voor Nieuwste Geschiedenis. Revue Belge d'Histoire Contemporaine, 12*, 713–726.

Bracke, N. (2008). *Een monument voor het land: Overheidsstatistiek in België, 1795–1870*. Ghent: Academia Press.

Breidbach, O. (2005). *Bilder des Wissens. Zur Kulturgeschichte der wissenschaftlichen Wahrnehmung*. Munich: Wilhelm Fink Verlag.

Casselman, I. (1984). *De lange weg naar Kortrijkse kandidaturen. Universitaire expansie 1957–1965*. Unpublished master's thesis in history, Leuven.

Craeybeckx, L. (1962). *Universiteit Antwerpen nu*. Antwerp: Ontwikkeling.

Craeybeckx, L. (1969). *Universiteit van deze tijd. Zonder schoolstrijd: pluralistisch, democratisch; medebestuur van studenten, volksverbonden; open voor alle begaafden*. Antwerp: De Nederlandsche Boekhandel.

De Ceulaer, D. (1990). *De verlenging van de leerplicht: veertig jaar Belgische onderwijspolitiek.* Leuven: Universitaire Pers Leuven.

De Clerck, K. (1975). *Tien Vlaamse rectoren over tien jaar universitaire expansie.* Ghent: Gakko.

De Clerck, K. (1985). *Rector Bouckaert had toch gelijk.* Ghent: Archief van de Rijksuniversiteit.

De Keyser, C. C. (1986). *Naar een comprehensief Europees basisonderwijs voor het jaar 2000. Vergelijkend historisch essay over de dialectiek tussen twee maatschappij- en onderwijsmodellen*. Faculteit der Psychologie en Pedagogische Wetenschappen, Seminarie voor Comparatieve Pedagogiek, Leuven.

Depaepe, M. (1979). Kwantitatieve analyse van de Belgische lagere school (1830–1911). *Belgisch Tijdschrift voor Nieuwste Geschiedenis. Revue Belge d'Histoire Contemporaine, 10*, 21–81.

Depaepe, M. et al. (1992). *Maatschappelijk profiel van de leerkrachten basisonderwijs tijdens de laatste halve eeuw. Exploratie op basis van een enquête.* Brussel: COV.

Depaepe, M., & Simon, F. (1997). Social characteristics of Belgian primary teachers in the twentieth century. *Cambridge Journal of Education, 27*, 391–404.

De verkavelde universiteit in Vlaanderen. (1998). *Ons Erfdeel, 41*, 14–37.

D'hoker, M. et al. (2006). *De statistieken van het onderwijs in België. Het algemeen secundair onderwijs, 1945–2000. Les statistiques de l'enseignement en Belgique. L'Enseignement secondaire de formation générale, 1945–2000.* Brussels: Archives générales du royaume/Algemeen rijksarchief.

Gerlo, A. (1989). *Noch hoveling noch gunsteling. Een levensverhaal.* Kapellen: DNB/Uitgeverij Pelckmans.

Groenez, S. (2008). *Onderwijsexpansie en –democratisering in Vlaanderen*. Paper presented at the day of the sociologists, May 29th, Leuven.

Henkens, B. (2004). The rise and decline of comprehensive education: Key factors in the history of reformed secondary education in Belgium, 1969–1989. *Paedagogica Historica, 40*, 193–209.

Henkens, B. (2006). *Steekspel tussen traditie en vernieuwing. Dertig jaar polemiek over de structuur van het secundair onderwijs (1963–1992).* Unpublished doctoral thesis in educational sciences, Leuven.

Hudson, P. (2000). *History by numbers. An introduction to quantitative approaches*. London: Arnold.

Kossmann, E. H. (1987). De tijdgeest. Geschiedenis en gebruik van een begrip. In M. Depaepe & M. D'hoker (Eds.), *Onderwijs, opvoeding en maatschappij in de 19de en 20ste eeuw. Liber Amicorum prof. dr. Maurits De Vroede* (pp. 43–53). Leuven/Amersfoort: Acco

Labaree, D. (2009). *What schools can't do: Understanding the chronic failure of American school reform*. Unpublished lecture at a doctoral colloquium, September 4, 2009, Bern.

Lenders, P. (2002). 150 jaar hoger handelsonderwijs in Antwerpen. In H. van Goethem (Ed.), *Antwerpen en de jezuïeten 1562–2002* (pp. 163–184). Antwerp: UFSIA.

Minten, L. et al. (1991–1996). *Les statistiques de l'enseignement en Belgique. L'enseignement primaire 1830–1992* (5 vols.). Brussels: Archives générales du royaume.

Pelleriaux, K. (1999). *De keerzijde van de onderwijsdemocratisering.* Brussels: VUB, Vakgroep Sociologie, Onderzoeksgroep TOR. (Internal report).

Pelleriaux, K. (2001). *Demotie en burgerschap. De culturele constructie van ongelijkheid in de kennismaatschappij*. Brussels: Vubpress.

Popkewitz, T., & Lindbald, S. (2001). Educational statistics, equity problem and systems of reason: Relations of governing education and social inclusion and exclusion. In S. Lindblad & T. S. Popkewitz (Eds.), *Statistical information and systems of reason on education and social inclusion and exclusion in international and national contexts* (pp. 331–350). Uppsala: Uppsala University, Department of Education.

Ruimere kansen voor allen. (1964). Spreiding van het universitair onderwijs in Vlaanderen. Antwerp: De Nederlandsche Boekhandel.

Schneider, J. (2010). *Excellence for all. American school reform, 1983–2008.* A dissertation submitted to the school of education and the committee graduate studies of Stanford University in partial fulfillment of the requirements for the degree of doctor of philosophy.

Tan, B. (1998). Blijvende sociale ongelijkheden in het Vlaamse onderwijs. *Tijdschrift voor Sociologie, 19*, 169–207.

Verbergt, G., Cantillon, B., & Van den Bosch, K. (2009). *Sociale ongelijkheden in het Vlaamse onderwijs: tien jaar later*. Antwerp: Universiteit Antwerpen, Stadscampus, Centrum voor Sociaal Beleid Herman Deleeck. (*Berichten*, January 2009)

Wielemans, W. (Ed.). (1986). *Vernieuwingen in het secundair onderwijs. Een Europese situering*. Leuven/Amersfoort: Acco.

Chapter 4
Child Maltreatment in the Last 50 Years: The Use of Statistics

Jeroen J.H. Dekker

4.1 Introduction

Maltreatment of children as an educational problem received new and on-going attention with the publication in 1962 of the article on the *Battered Child Syndrome* by the American medical doctors C. H. Kempe, F. N. Silverman and their colleagues (Kempe, 1962). Following that publication, numerous studies were published on the abuse and neglect of children. Moreover, an increasing institutional and legal framework for diagnosing and preventing child maltreatment was set up in many countries in the Western world. With the 1989 *United Nations Convention on the Rights of the Child*, in particular article 19, it was accepted that there should be international standards on the protection of children and on the behaviour of their educators. States should exercise more responsibility in this matter and professionals should have more power to tackle the problem. These views were accepted across most of the world with almost all UN member states ratifying the Convention.

For the historian of the early 21st century, looking back at almost 50 years of child maltreatment research and policy, the question rises whether this increasing world-wide interest in the maltreatment of children, this tackling the problem of child maltreatment as a form of educational politics successfully resulted in a major diminution of child maltreatment?

At first sight, the hypothesis that maltreatment of children was diminishing, at least in the economically prosperous Western world, seems strong (see Section 4.2). However, when looking more thoroughly into the information available, with the focus on statistical studies on the prevalence of maltreatment of children in the Western world (particularly when one interprets data drawn from the USA and the Netherlands), we can see that this hypothesis is flawed (see Section 4.3). Part of the answer to the question of why most studies on the prevalence of child maltreatment reject the hypothesis of diminution, notwithstanding the existing controversies between researchers on questions of method and definition, seems to relate to the multiplier effect of a broader definition of child maltreatment since the 1970s

J.J.H. Dekker (✉)
University of Groningen, Groningen, The Netherlands

P. Smeyers, M. Depaepe (eds.), *Educational Research: The Ethics and Aesthetics of Statistics*, Educational Research 5, DOI 10.1007/978-90-481-9873-3_4,

(see Section 4.4). A second multiplier effect on the reported prevalence of child maltreatment seems to be the impact of internationally accepted children's rights by prescribing criteria for good parenthood and for child protection (see Section 4.5). A third multiplier effect seems to consist of the preference of policy-makers for clear figures on this topic so as to develop educational policies, in this case child protection policies. This preference is, according to David Labaree in Chapter 2, this volume, related to the aesthetic appeal of things being neat, clean and orderly. The state, policy-makers, educational professionals and educational researchers all seem to share this desire for "clean" statistics (see Section 4.6). It will be concluded that the combination of the three multiplier effects mentioned above contributed enormously to the increase of the reported prevalence of child maltreatment in the Western world. Diverging statistical outcomes also caused a lot of confusion for policy-makers (see Section 4.7).

4.2 Diminution of the Prevalence of Maltreatment of Children?

Maltreatment of children is not a new issue in the historiography of the history of childhood and education. According to scholars such as Loyd Demause, who built his historical analysis on a linear development of the history of education and childhood, we are moving out of a dark age as regards the maltreatment of children. Until recently, such maltreatment was normal daily practice. Many studies, often inspired by Philippe Ariès' *Centuries of Childhood* from 1960, and written by historians such as Edward Shorter, Lawrence Stone, Simon Schama, Loyd Demause, and Lea Dasberg, defend the idea that the child as *animal educandum* was a new phenomenon in history. It was to be found only among the elite from the end of the middle ages onwards. The majority of people had to wait until the end of the 18th century before they were able to really love and educate their children. Even then maltreatment of children was seen as part of a normal education. However, in reaction to this linear historiography of childhood and education, a series of studies emphasising structure and continuity by historians such as Linda Pollock, Alan Macfarlane, Stephen Ozment, Harrie Peeters, Shulamith Shahar, Emmanuel Le Roy Ladurie, Jacques Gélis and the art historian Jan Baptist Bedaux was published. In these studies the continuity of the idea of childhood, of affectionate relations between parents and their children was stressed. This interpretation of the history of childhood has resulted in the conviction that the history of education and childhood should not be looked at in terms of linear development. This level of conviction is based on many examples of good parenting in early modern Europe (Ariès, 1960; Dekker & Lechner, 1999, pp. 37–49; Dekker, 2006, pp. 21–23; Dekker & Groenendijk, 1991, pp. 317–335).

Although maltreatment of children was either, according to the evolutionists, a normal aspect of education until recently or, according to the structuralists, a deviation from widely accepted educational norms, it was not recognised as a major educational issue until well into the 19th century (Stearns, 2006, pp. 4, 47). Then, with the rise of the Society for the Prevention of Cruelty in Children in 1874 in

New York City, with local SPCCs in many American towns soon after that year, the issue was put on the emerging social work agenda (Giovannoni, 1990, pp. 6–7; Baartman, 1993; Zigler & Hall, 1990, pp. 38–75). Around 1900, adherents of the Progressive Education movement such as Ellen Key emphasised the issue of maltreatment of children in the context of schooling. It was only in the 1980s through the outlaw of corporal punishment in British schools and on schools in a number of American states (Stearns, 2006, p. 105) that this issue was taken seriously. Meanwhile, a number Child Protection Acts were passed in Europe, the USA and Canada (Dupont-Bouchat et al., 2001, chap. 5; Dekker, 2001, 2002, 2000, 2007; Key, 1900). In the first decades of this child protection system, the focus was not so much on child abuse as criminal and neglected children and juvenile justice. Moreover, it was assumed that the prevalence of those problems was very modest indeed. According to the Dutch liberal minister of Justice, P. W. A. Cort van der Linden (1846–1935), speaking in 1901 during the Parliamentary treatment of the Dutch Child Acts, risky parental behaviour and risky childish behaviour were the exception. Moreover, the state only had the right to intervene after negative effects were observed: "Only stopping, not preventing is the responsibility of the state" (Donker, 1955, p. 7). This approach changed after the Second World War. Speaking in 1955 at the 50-year celebration meeting of the Child Acts, Mr. J. Overwater (1892–1958) – president of the National Federation/the Dutch Association for Child Protection, a magistrate and a leading figure in the world of child protection – maintained that child protection was "dominated by cases of various sorts of educational and behavioural problems. Moreover, these problems, in contrast with the majority of former cases on material neglect in the past, are now not limited to one single social group: they do exercise their bad influence upon the society as a whole" (Overwater, 1955, pp. 15–23; Overwater, 1948). This statement reveals a broader understanding of the maltreatment of children, recognising that there are more risks and more families and children at risk.

In 1962, policy focused on child abuse and child maltreatment received an enormous boost with the publication of the famous article *Battered Child Syndrome* by the American medical doctor Kempe. Following this publication, maltreatment of children (understood as an educational problem) was put on the agenda of medical doctors. Moreover, from the 1970s, it got an increasing and lasting attention from policy-makers, educational professionals and educational scientists (Kempe, 1962; Baartman, 1996). Indeed, due to the work of Kempe and his colleagues, "child abuse has become a common household word" (Giovannoni, 1990, p. 10). Carole Jenny wrote an article entitled "Medicine Discovers Child Abuse" during which she praises Kempe's famous 1962 article, calling it one of the *JAMA* (*Journal of the American Medical Association*) classics. According to Jenny, the importance of Kempe's article has little to do with its scientific value. It only includes two case reports and, for some readers, might therefore appear a little thin on statistics. Moreover, *Battered Child Syndrome* "gave speculative psychological explanations about why parents would severely beat their children". Kempe's article became a *JAMA* classic because it established the idea "that physicians have a special responsibility to children – a responsibility to help keep them safe, sometimes even

from their own parents". As a result of this article, mandatory legislation relating to child abuse was established in every state in the USA between 1963 and 1967. This later caught on in Europe. In addition, many scientific reports and articles on the subject were being written (Jenny, 2008, pp. 2796–2797; Kempe & Kempe, 1978; Kempe, 2007).

Growing attention to the maltreatment of children does not automatically mean that it should become more prevalent. On the contrary, when looking back over 50 years of child maltreatment research and policy, we see that maltreatment of children, at least in the Western world, diminished substantially. This was due to two developments: first, a new and almost generally accepted standard of educational behaviour resulted from growing attention to child maltreatment; second, a historically unprecedented increase of the availability of financial means for education resulted from increasing economic prosperity.

First, Kempe's article helped to establish a general standard as regards unacceptable behaviour towards children. Over the last few decades, what counted as unacceptable was made more explicit than ever before. Standards for educational behaviour were explicitly laid down in national and international regulations. This meant that parents could now be expected to be conscious of the fact that, when beating their child, they were acting against almost generally accepted educational standards and breaking the law. The United Nations Convention on the Rights of the Child from 1989 requires member states to protect their children against violence and maltreatment.

Second, increasing economic prosperity in the Western world during the last 50 years meant that more financial means were available for the education of children than ever before in history. This made it easier for a growing number of parents to educate their children and to care for them according to the new educational standards and laws. I want to suggest that this is a convincing reading of the situation despite the occurrence of economic crises in the early 1980s and the crisis that is still going on now. This current crisis might have a temporarily negative effect on the diminishing trend of child maltreatment. It should also be noted that some parents did not benefit from the welfare state. In the USA this became clear in 1997 through the work of American sociologists Paul Amato and Alan Booth in their alarming book entitled "A Generation at Risk" (Amato & Booth, 1997; Dekker, 2009, pp. 17–36). The majority of studies from the 1970s until now seem to reject this hypothesis of diminution. Rather than corroborating the hypothesis of a diminishing trend in child maltreatment, they report an increase of maltreatment. Some of these studies will be examined in the next section. In Sections 4.4, 4.5 and 4.6, I will address the reasons for seeing child maltreatment as an increasing phenomenon.

4.3 No Diminution: Studies on the Prevalence of Child Maltreatment as Historical Sources

After the publication of Kempe's alarming article, numerous articles in an increasing number of academic journals were published on child abuse and neglect. The increase in the number of articles listed under the keywords child abuse in the

MEDLINE database is noteworthy. In 1963, the year in which these keywords were added to the MEDLINE system, only 12 articles were published. In 2006 nearly 600 articles were published (Jenny, 2008, pp. 2796–2797). Also, an impressive institutional framework of diagnosing and preventing child maltreatment was set up (Bancroft & Silverman, 2002, chap. 9; Righthand, Kerr, &. Drach, 2003, chap. 4; Roelofs & Baartman, 1997, pp. 192–211; Baartman, 1996, chap. 5). In addition, from the end of the 1980s, new international juridical standards came into existence and could be used to justify the expanding criteria for child abuse and neglect. We will look at these standards in the next section.

In her aforementioned article from 2008 on Kempe's publication, Jenny concluded: "Over the years there has been acknowledgement that overwhelming numbers of children are abused physically and sexually". In the USA, the National Child Abuse and Neglect Data System reported 12.1/100,000–905,000 children were found "by social service agencies to be survivors of child abuse or neglect". Meta-analysis from 1999 reported "30 to 40% of women and 13% of men experienced sexual abuse during childhood". According to Jenny, it is clear that "understanding of the ways children can be maltreated has expanded greatly". However, she emphasises that "there is no evidence that the actual prevalence has increased, these numbers would have been unbelievable in the 1960s" (Jenny, 2008, p. 2796). The studies to be discussed below, however, report that abuse has become more prevalent during the last decades.

David Finkelhor, one of the most famous experts on the subject after Kempe, published his first comprehensive study on child maltreatment in 1979. In his publications from the 1970s onwards he argued that childhood had become increasingly prevalent (Finkelhor, 1979; Finkelhor, 1984). In a 2005 publication on a representative sample of American children and young people within the age range of 2–17 years, Finkelhor and his team found that "More than one half [. . .] of the children and youth had experienced a physical assault in the study year, more than 1 in 4 [. . .] a property offence, more than 1 in 8 [. . .] a form of child maltreatment, 1 in 12 [. . .] a sexual victimization, and more than 1 in 3 [. . .] had been a witness to violence or experienced another form of indirect victimization. Only a minority (20%) had no direct or indirect victimization" (Finkelhor, Ormrod, Turner, & Hamby, 2005b, p. 5). In an appendix, victimisation is defined as a complex of several dozen activities; Finkelhor and his team define victimisation in a much broader way than Kempe did in 1962. They maintain that "twenty-two percent of the children in this sample had experienced four or more different kinds of victimizations in separate incidents (what we term poly-victimization) within the previous year" (Finkelhor, Ormrod, Turner, & Hamby, 2005, pp. 1297–1312).

Finkelhor is not alone in publishing alarming figures on child maltreatment. Indeed, many publications from the 1990s and early 2000s, published in academic journals such as *Pediatrics*, *Child Maltreatment*, the *European Journal of Criminology*, *Child Abuse & Neglect*, *Child Development*, and *Science*, tell a story of expanding child maltreatment (Sternberg, Lamb, Guterman, & Abbott, 2006, pp. 283–306; Sariola & Uutela, 1992, pp. 823–832; May-Chahal & Cawson, 2005, pp. 969–984). In an article in *Pediatrics* from 2007 on Child Maltreatment in the USA (this study was based on the National Longitudinal Study of Adolescent Health,

a cohort study that uses the method of self-report), Hussey et al. conclude that "self-reported childhood maltreatment was common", while "each type of maltreatment was associated with multiple adolescent health risks". Although the authors warn against making sweeping general statements, because "[d]espite 40 years of sustained research on child abuse and neglect, we are still struggling to answer these basic questions", they conclude that "childhood maltreatment is prevalent, and its adverse consequences are many. Conservative estimates place the number of US children victimized by maltreatment each year at close to 1 million and the annual number of child deaths caused by abuse or neglect at nearly 1500" (Hussey, Chang, & Kotch, 2006, pp. 934, 940). This conclusion confirms the figures generated by Finkelhor. Moreover, they fit the already mentioned alarming "generation at risk" thesis put forward by Amato and Booth. Their book (based on a longitudinal life course Study of Marital Instability beginning in 1980 with interviews of ca. 2000 persons, and repeated in 1983, 1988, and 1992), concludes that "parental marital quality is the key variable" for the well-being of their offspring. Declining marital quality tilted "the balance for offspring making the transition to adulthood [. . .] in a negative direction". This meant that, "the outlook for future generations of youth may be even more pessimistic". Therefore, Amato and Booth recommend family policy that should "be based on creating incentives for parents to act in the best interests of their children" (Amato & Booth, 1997, pp. 215, 221, 223, 239). Studies on other countries, such as Great Britain, resulted in figures comparable to the American ones (Sidebotham & Golding, 2001, pp. 1177–1178, 1189, 1196–1197).

Comparative research published in 2006 by E. Douglas and M. Straus was conducted with university students in 19 countries in which they were asked about their childhood experiences of corporal punishment. This resulted in the conclusion that "over half of the students did not 'strongly disagree' that they were 'spanked or hit a lot' by their parents as child (under age 12)". It should be noted that this figure does not account for major differences occurring among the 19 countries studied. In Washington DC, 72.6% of the students did not strongly disagree with that statement. In Dutch Amsterdam the figure was only 19.7%. The arcadia for children, according to this study, seemed to be Belgium, or precisely Flanders, with only 12.9% not strongly disagreeing. On the other hand, in Freiburg, the percentage was 61.5 and in Leicester, 53.7. It is a matter of fact that major differences were found between university cities. Such differences could also be found within single countries. Let us take Canada as an example. While in Winnipeg, the percentage was 66.5, thus almost the same as in Washington, in Montreal a percentage of 27.3 was found. In other words, according to this comparative study, child maltreatment is particularly prevalent in the USA. Despite the "large differences between the 36 university sites in the prevalence of CP", whereby the "the median rate (56) was high", the authors nevertheless came to the conclusion that "these findings point to an important public health and crime problem among youth from relatively privileged segments of the nineteen countries in this study". They recommend "increased efforts to end all use of CP by parents", consisting of "a change in parent education efforts in the form of unequivocal advice to never smack, analogous to the unequivocal advice to never smoke" (Douglas & Straus, 2006, pp. 293, 302, 303, 311, 314).

Similar research by Straus on neglectful behaviour by parents in the life history of university students in Europe, North America, Latin America, Asia, Australia and New Zealand shows the same variety as regards the prevalence of neglectful behaviour. However, this research shows that it operates at a lower level, ranging "from 3.2 to 36% (median 12%)", with the lowest level being found in New Hampshire, USA. There 3.2% experienced three or more forms of neglectful behaviour. Korea-Pusan came out on the top with 36.4%. These conclusions are alarming: "This study [. . .] found that half of the students experienced at least one of the eight neglectful behaviors as children, and about 12% experienced a pervasive pattern of neglect as indicated by three or more of the eight neglectful behaviors measured". Although the rates of students "who experienced three or more neglectful behaviors as a criterion, [. . .] ranged from a low of 3% to a high of 36%", they conclude: "Even the figure of 3% for the university with the lowest rate is high [. . .] The results show high rates of neglectful behavior in both developed and underdeveloped countries and among a privileged sector of those countries" (Straus & Savage, 2005, pp. 124, 129, 130, 131–133, 134; Berger, Knutson, Mehm, & Perkins, 1988, pp. 259, 260).

Until recently, the prevalence of child maltreatment in the Netherlands has been estimated using US figures. Looking at the Dutch situation it appears that 40,000–80,000 children were being maltreated each year, depending on the definition of maltreatment used. Since 2007, new research, carried out on request of the Dutch government, resulted in substantially higher figures than the earlier extrapolations, varying from 107,200 according to the Leiden University report to 160,700 children according to the Free University of Amsterdam report (Van IJzendoorn et al., 2007; Lamers-Winkelman, Slot, Bijl, & Vijlbrief, 2007; Baartman, Bullens, & Willems, 2005). The difference between these figures can be explained mainly by the method used. The report by the Free University made use of self-report by children aged 11–18 years. This can be compared to the method of sampling used by Straus. What distinguished the methods used by the Free University was the combination of questions on child maltreatment with questions on other nuisance making unpleasant events in the life of the children. Moreover, the questionnaire the children were asked to fill in was not offered as a questionnaire on child maltreatment but as a Nuisance Making and Unpleasant Events Questionnaire, in Dutch "Vragenlijst Vervelende en Nare Gebeurtenissen", abridged as VVNG. As a result, the participating children were not conscious of the fact that they were filling in a questionnaire that contributed to a study on child maltreatment. The report by the Leiden University was not based on child responses, but gained its data from interviewing professionals. A series of professionals, working at schools and child protection institutions, acted as monitors of possible child maltreatment. As expected (if one considers Straus's experiences), the method of self-report resulted in higher figures.

Both reports used the same broad definition of child maltreatment as a framework, namely the definition laid down in the Dutch Law on Youth Care from 2005, containing both physical and psychological harm, emphasising the child's dependence, and including "any form of interaction that is threatening or violent for the

child" (Van IJzendoorn, 2007a, p. 161). In applying this definition in their study, some differences emerged. The definition of maltreatment that features in the Leiden report includes the refusal by parents to obey the advice of professionals when, for example, they did not support the treatment of a child by a professional or when they refused to send the child to a day-care institution when advised to. Thus, the Leiden report is strongly focused on the professional in two ways: on the one hand for the assessment of the prevalence of child maltreatment, on the other in increasing the moral power of the professional in considering a parent's refusal of advice.

Apart from its methodology, the Amsterdam report does not focus on the professional. Its focus is on the child and his/her experiences with child maltreatment. The higher figures of prevalence result both from the well-known effects deriving from the method of self-report and from the fact that the categories of child maltreatment have received further development (Van IJzendoorn et al., 2007, appendix 2; Lamers-Winkelman et al., 2007, p. 2). These two reports were commissioned by the Minister of Youth and the Family to get convincing figures on the prevalence of child maltreatment for its child and family policy. It was felt conclusions should no longer be based on extrapolation of foreign figures, but on Dutch evidence. The result, however, was not a convincing conclusion on prevalence. On the contrary, it generated confusion. Moreover, soon after the publication of the reports a dispute developed between the two research groups. This dispute concerned the methods and the definitions used in the research. This will be discussed in the next section.

According to the studies examined above, prevalence of child maltreatment seemed to increase substantially in the last 50 years. Yet, the prevalence figures from these studies also contribute to major confusion. This is because of changing and expanding definitions of child maltreatment that have supplied policy-makers with strongly varying quantitative maps of child maltreatment.

In the following sections, I will examine how the history of mapping this phenomenon can be explained. Three phenomena with possible multiplier effect will be treated: changing definitions, the impact of children's rights and finally the appeal of statistical mapping for policy-makers, educational professionals and educational researchers.

4.4 From Narrow to Broad: The Changing Definition of Child Maltreatment

When Kempe defined the battered-child-syndrome in 1962, he described physical maltreatment of children in the following way: "The battered-child syndrome, a clinical condition in young children who have received serious physical abuse, is a frequent cause of permanent injury or death. The syndrome should be considered in any child exhibiting evidence of fracture of any bone, subdural hematoma, failure to thrive, soft tissue swellings or skin bruising, in any child who dies suddenly, or where the degree and type of injury is at variance with the history given regarding the occurrence of the trauma [. . .]". As to the reasons for why this occurred, he and his colleagues emphasised psychiatric factors, describing abusers

as "possible 'psychopathic or sociopathic characters'" (Kempe, Silverman, Steele, Droegemuller, & Silver, 1962; Jenny, 2008).

Meanwhile, psychological abuse also became part of the phenomenon of child maltreatment. In 1956 the Dutch psychologist and child protection specialist B. Clemens Schröner published a pioneering book on psychological child abuse. Kempe and his colleagues soon followed, generating a broader concept that included nutritional and emotional abuse (Clemens Schröner, 1956; Helfer, 1974, p. 37). Finkelhor and his colleagues went even further. They developed a concept consisting of many elements, including the categories of (1) Physical assaults, bullying and teasing, (2) Sexual Victimisation, (3) Child Maltreatment, (4) Property Victimisations and (5) Witnessed and Indirect Victimisations. This expansion of the definition was no doubt of influence on the alarming percentages of victimisation among American children and youth (Finkelhor et al., 2005, p. 1297). The process of expansion intensified in the context of studies, such as those by Straus and Savage. These researchers came to alarming conclusions, namely that "half of the students experienced at least one of the eight neglectful behaviors as children, and about 12% experienced a pervasive pattern of neglect as indicated by three or more of the eight neglectful behaviors measured". They also concluded that "high rates of neglectful behavior in both developed and underdeveloped countries and among a privileged sector of those countries" are normal. These conclusions resulted in the assertion that "helping parents avoid neglectful behavior could make a further contribution to the primary prevention of all types of family violence". It is striking how different their definition is from Kempe's. For Straus and Savage, neglectful behaviour encompasses not helping with homework, not comforting a child who is upset, not helping when the child had problems, not making sure the child goes to school, not helping the child to do its best, not giving the child enough clothes to keep it warm, not keeping the child clean and not caring if the child gets into trouble at school. The "most frequent neglectful behavior was 'not helping with homework'". This behaviour was, according to this study, reported by 29% of the students, although the "percentage of parents in the 33 sites who did not help with homework ranged from 10% to 73%" (Straus & Savage, table 3, 2005).

As a result, a great number of definitions of child abuse have been proposed. They can be "arranged in a continuum from very narrow to rather broad". With the narrow definitions generally focused on physical abuse or on sexual abuse, the broadest definitions are related to the optimal development of the child. So, in the 1970s the Child Welfare League of America defined child abuse as "the denial of 'normal experiences' that produce feelings of being loved, wanted, secure and worthy" (Zigler & Hall, 1990, p. 45). The definitions are also, according to Giovannoni, characterised by vagueness. Moreover, they are amorphous and ambiguous. Professional groups, such as medical doctors, social workers and lawyers use them in different ways (Giovannoni, 1990, pp. 10–16).

This continuum of different definitions, used together with different research methods, resulted in another continuum, that of prevalence figures. The diverging figures of the recently published Dutch prevalence studies with prevalence figures from 107,200 to 160,700 are no exception to the rule. Previous observations of the

USA resulted in even bigger differences. According to Zigler and Hall, in 1990 "estimates of the extent of child abuse [. . .] range from 200,000 to 4 million cases annually". That range is due to different definitions, different sampling methods and differences in data sources (Zigler & Hall, 1990, p. 48). It brought Giovannoni, who, by the way, reports a range from 652,000 to 6 million cases annually in the USA for the same period, to the conclusion: "It all depends on to whom you are talking and what they are talking about" (Giovannoni, 1990, pp. 19–20).

The range of prevalence also depends on how strongly researchers rely on the objectivity of the informants who they interviewed. In the Leiden Report, prevalence is partly dependent on the researcher's belief in objectivity of the professionals who were acting as informants for their study on prevalence. This belief can be remarkably strong. For example, the Leiden researchers compare their ca. 700 informants as "video cameras", in other words as elements of an objectively observing system (Van IJzendoorn et al., 2007, p. 57, appendix 13). This is, however, not how historians usually treat their sources. In this specific case, the result of the activity of the informants, the prevalence figures, could also be interpreted as the result of informants not acting as objective, machine-like observers but as individual human beings. When reporting on child maltreatment, these informants undoubtedly do their best to observe, while making use of the broad definition of child maltreatment from the Leiden reporting objective a manner as possible. Yet, they will never be able to avoid making also their own subjective assessment of the emotionally charged topic of child maltreatment. This invariably influences the objectivity of the resulting overall prevalence figures. The researchers working on the VU Amsterdam report opposing the use of professional observers' views when assessing the prevalence of child maltreatment. According to them, professional observers report less child maltreatment than non-professional citizens (Lamers-Winkelman et al., 2007, p. 227). On the other hand, the Leiden researchers do not believe that children and youngsters as adequate reporters of child maltreatment.

In sum, notwithstanding the existing range from very narrow to very broad, the majority of studies on the prevalence of child maltreatment use a very broad definition of child maltreatment. From the 1970s, the battered-child-syndrome as introduced by Kempe turned out to become only one of many categories of child maltreatment. When looking at the prevalence figures for child maltreatment, the range is also impressive. Ultimately, alongside other concerns mentioned above, the belief of researchers in the objectivity of their informants is a main factor in explaining differences in assessing prevalence.

4.5 The Impact of Children's Rights

Next to the multiplier effect of a broadening definition, another multiplier seems to influence the increase of the reported prevalence of child maltreatment. Here, I am thinking of the Children's Rights movement, which offers professionals and scientists a framework for working in the best interests of the child.

According to Stuart Hart in *Child Abuse & Neglect*, "International standards now exist that establish a universal imperative for protecting children from child abuse

and neglect. The UN Convention (treaty) on the Rights of the Child is the pre-eminent international philosophical and legal base in regard to children's human rights" (Hart, 2007, p. 473). Indeed, article 19 from that Convention reads as follows: "(1) States Parties shall take all appropriate legislative, administrative, social and educational measures to protect the child from all forms of physical or mental violence, injury or abuse, neglect or negligent treatment, maltreatment or exploitation, including sexual abuse, while in the care of parent(s), legal guardian(s) or any other person who has the care of the child. (2) Such protective measures should, as appropriate, include effective procedures for the establishment of social programmes to provide necessary support for the child and for those who have the care of the child, as well as for other forms of prevention and for identification, reporting, referral, investigation, treatment and follow-up of instances of child maltreatment described heretofore, and, as appropriate, for judicial involvement". As a consequence, "a new level of heightened awareness and commitment to child protection has been established" (Hart, 2007, pp. 473–474).

Because of the increasing gap between the intentions laid down in the UN Convention and the tough reality of the protection of children at risk, Hart asks for "support for the State to assume its higher than usual legitimate rights and responsibilities to intervene – intrude – in the lives of families and children". He asks "for the application of intervention models that incorporate the qualities of the highly successful Multisystemic Therapy (MST) orientation to working with children at risk, their families and communities" (Hart, 2007, pp. 475–476).

In sum, more power for the state and for the professionals is recommended and justified by the juridical base of international children's rights. These rights function both as a child protection machine and as a multiplier for the number of maltreated children.

4.6 Aesthetics of Statistics, or, the Appeal of Statistical Mapping for Policy-makers

According to David Labaree, policy-makers are sensitive to the aesthetic appeal of statistics when viewing society. So, "educational policymakers have a preference for data that seem authoritative and scientific and that present a certain faced validity" (Labaree, Chapter 2, this volume; Dahlberg & Moss, 2005, pp. 64–85; Paul, 2005, pp. 16–17, 79–83). According to Labaree, mapping educational reality using statistics generates neatness, cleanliness and orderliness. That is appealing for the state and its policy-makers, but also for educational professionals and educational researchers. This is exemplified by the two diverging reports on child maltreatment in the Netherlands treated above. These two reports were made on request of the Minister of Youth and the Family and vice prime minister, André Rouvout, leader of the small orthodox-protestant *Christen-Unie*, who needed convincing figures on the prevalence of child maltreatment for a policy based on Dutch evidence. The Dutch government had a preference for statistical mappings they could develop more adequate policy that they were willing to pay for. In his web log of 25 April 2007, André Rouvout reacts to these reports saying, "we now know the dimension

of the problem". He hoped that the number of child maltreatment reports through the local AMKs, or, Child Maltreatment Report Centres, until now ca. 40,000, would increase substantially. In other words, some hours after receiving the two divergent reports on child maltreatment, the available statistics immediately come to function as a new map of reality. The Minister even hopes that reality will soon be accommodated to these new maps (Rouvout, Weblog 25 April 2007).

In his side letter to Parliament on the same day, he mentioned the divergence in figures between the two reports. His solution for dealing with this was to simply conclude that the Leiden report must indicate a minimum figure. Yet, he also writes that we now understand the dimensions of the problem – the AMKs reports of ca. 40,000 inaccurate and that more policy is necessary in order to show the real dimensions of the problem (Rouvoet, Letter to Parliament, 25 April 2007).

This example shows that due to the pressure exerted by educational professionals (who argue that child maltreatment is an underestimated problem), policy-makers can be easily impressed by new statistical mapping of a phenomenon. In this case, they conclude that figures must be high, that prior reports must have underestimated the problem, and that we now know the true dimensions of the problem. This happened despite the fact that the three prevalence figures, namely 40,000, 107,200 and 160,700, mean that we simply have no way of making any solid conclusions. In other words, the aesthetics of statistics impress policy-makers even when statistics result in confusion instead of clarity. The policy-maker simply says: we now know the problem, and we now can base our policy on statistics.

4.7 Conclusion

In looking back at almost 50 years of child maltreatment research and policy, it became clear that decreasing maltreatment of children in the Western world is not what the majority of studies on the prevalence of maltreatment of children show. It is true that the diminution hypothesis is probably maintained for the prevalence of physical child maltreatment, coming under the definition of child maltreatment as formulated by Kempe in 1962. However, when using other definitions, we must accept that there has been an increase in child maltreatment in the Western world. This increase can be observed by looking at three multipliers (though this is not to discount the influence of other factors).

The first multiplier consists in the constantly broadening definition of child maltreatment since the 1970s. This has meant that the definition of child maltreatment "all depends on to whom you are talking and what they are talking about" (Giovannoni, 1990, pp. 19–20). Behind this development of broadening definitions is the new attention for child maltreatment (that emerged in the 1960s), resulting in more explicit educational standards to be followed by parents and educators.

The impact of internationally accepted children's rights in which criteria for good parenthood and for child protection have been defined forms the second multiplier. Evidence for this impact is to be found in the Dutch Civil Code, where, from 25 April 2007, in article 247#2 we read the following: "When caring and educating

their child, parents do not use any psychological or physical violence or any other humiliating treatment". This change in the Dutch Civil Code resulted from the recommendation in 30 January 2004 by the UN committee that supervises the compliance by member states of the UN Convention on the Rights of the Child (Rouvoet, Letter to Parliament, 25 April 2007, p. 2).

Last but not least, the preference of policy-makers for clear figures on this topic worked as a multiplier. As we saw, the Dutch Minister of Youth and the Family to draw firm conclusions from contradictory figures on child maltreatment. The presentation of figures, notwithstanding the major contradictions between the two reports and the resulting scientific confusion, was enough for policy-makers to conclude that "we now know the dimension of the problem", and that we can now base our policy on a statistical map of the phenomenon, even if this map is not simply representing the world, but making it more confusing.

References

Amato, P. R., & Booth, A. (1997). *A generation at risk. Growing up in an era of family upheaval*. Cambridge, MA: Harvard University Press.

Ariès, Ph. (1960). *L'Enfant et la vie familiale sous l'ancien régime*. Paris: Librairie Plon, 1960.

Baartman, H. E. M. (1993). *Opvoeden met alle geweld. Hardnekkige gewoontes en hardhandige opvoeders*. Utrecht: SWP.

Baartman, H. E. M. (1996). *Opvoeden kan zeer doen. Over oorzaken van kindermishandeling, hulpverlening en preventie*. Utrecht: SWP.

Baartman, H., Bullens, R., & Willems, J. (Eds.). (2005). *Kindermishandeling: de politiek een zorg*. Amsterdam: SWP.

Bancroft, L., & Silverman, J. G. (2002). *The batterer as parent. Addressing the impact of domestic violence on family dynamics*. London: Sage.

Berger, A. M., Knutson, J. F., Mehm, J. G., & Perkins, K. A. (1988). The self-report of punitive childhood experiences of young adults and adolescents. *Child Abuse & Neglect, 12*, 251–262.

Clemens Schröner, B. L. F. (1956). *Psychische kindermishandeling*. The Hague: Nijhoff.

Dahlberg, G., & Moss, P. (2005). *Ethics and politics in early childhood education*. London: Routledge/Falmer.

Dekker, J. J. H. (2000). The century of the child revisited. *International Journal of Children's Rights, 8*, 133–150.

Dekker, J. J. H. (2001). *The will to change the child. Re-education homes for children at risk in nineteenth century Western Europe* (pp. 104–115). Frankfurt am Main: Peter Lang.

Dekker, J. J. H. (2002). Demystification in the century of the child. The conflict between Romanticism and disenchantment in (residential) youth care from the 1830s to 2000. In E. J. Knorth, P. M. van den Bergh, & F. Verhey (Eds.), *Professionalization and participation in child and youth care. Challenging understandings in theory and practice* (pp. 27–48). Burlington: Ashgate.

Dekker, J. J. H. (2006). *Het verlangen naar opvoeden. Over de groei van de pedagogische ruimte in Nederland sinds de Gouden Eeuw tot omstreeks 1900*. Amsterdam: Bert Bakker.

Dekker, J. J. H. (2007). Philanthropic networks for children at risk in nineteenth-century Europe. *Paedagogica Historica, 43*, 235–244.

Dekker, J. J. H. (2009). Children at risk in history: A story of expansion. *Paedagogica Historica, 45*, 17–36.

Dekker, J. J. H., & Groenendijk, L. F. (1991). The republic of God or the republic of children? Childhood and childrearing after the reformation: an appraisal of Simon Schama's thesis about the uniqueness of the Dutch case. *Oxford Review of Education, 17*(3), 317–335.

Dekker. J. J. H., & Lechner, D. M. (1999). Discipline and pedagogics in history. Foucault, Ariès, and the history of panoptical education. *The European Legacy, 4*(5), 37–49.

Donker, L. A. (1955). Rede van de minister van justitie. In L. A. Donker (Ed.), *Toespraken gehouden ter herdenking van de gouden kinderwetten op 1 december 1955, in de Ridderzaal te 's-Gravenhage* (pp. 5–14). The Hague: Ministry of Justice.

Douglas, E. M., & Straus, M. A. (2006). Assault and injury of dating partners by university students in 19 countries and its relation to corporal punishment experienced as a child. *European Journal of Criminology, 3*(3), 293–318.

Dupont-Bouchat, M.-S., Pierre, E., Fecteau, J.-M., Trépanier, J., Petit, J.-G., Schnapper, B., et al. (2001). *Enfance et justice au XIX^e siècle. Essais d'histoire comparée de la protection de l'enfance 1829–1914, France, Belgique, Pays-Bas, Canada*. Paris: Presses Universitaires de France.

Finkelhor, D. (1979). *Sexually victimized children*. New York: Free Press.

Finkelhor, D. (1984). *Child sexual abuse. New theory and research*. New York: Free Press.

Finkelhor, D., Ormrod, R., Turner, H. A., & Hamby, S. L. (2005a). Measuring poly-victimization using the Juvenile Victimization Questionnaire. *Child Abuse & Neglect, 29*, 1297–1312.

Finkelhor, D., Ormrod, R., Turner, H. A., & Hamby, S. L. (2005b). The victimization of children and youth: A comprehensive, national survey. *Child Maltreatment, 10*(1), 5–25.

Giovannoni, J. (1990). Definitional issues in child maltreatment. In D. Cicchetti & V. Carlson (Eds.), *Child maltreatment: Theory and research on the causes and consequences of child abuse and neglect* (pp. 3–37). Cambridge: Cambridge University Press.

Hart, S. N. (2007). Reflections on the implications of re-victimization patterns of children and youth as clarified by the research of Finkelhor, Ormrod and Turner. *Child Abuse & Neglect, 31*, 473–477.

Helfer, R. A. (1974). The responsibility and role of the physician. In R. E. Helfer & C. H. Kempe (Eds.), *The battered child* (pp. 25–39). Chicago: University of Chicago Press.

Hussey, J. M., Chang, J. J., & Kotch, J. B. (2006). Child maltreatment in the United States: Prevalence, risk factors, and adolescent health consequences. *Pediatrics. Official Journal of the American Academy of Pediatrics, 118*(3), 933–942.

IJzendoorn, M. H. (2007a). Kindermishandeling in Nederland: Ongeveer 107.200 slachtoffers in 2005. *Pedagogiek, Wetenschappelijk Forum voor Opvoeding, Onderwijs en Vorming, 27*, 160–175.

IJzendoorn, M. H. van, et al. (2007). *Kindermishandeling. Leiden attachment research program*. The Hague: Ministry of Justice/WODC.

Jenny, C. (2008). Medicine discovers child abuse. *The Journal of the American Medical Association (JAMA), 300*(23), 2796–2797.

Kempe, A. (2007). *A good knight for children*. Bangor, ME: Booklocker.com, Inc.

Kempe, R. S., & Kempe, C. H. (1978). *Child abuse*. Cambridge, MA: Harvard University Press.

Kempe, C. H., Silverman, F. N., Steele, B. F., Droegemuller, W., & Silver, H. K. (1962). The battered-child syndrome. *The Journal of the American Medical Association (JAMA), 181*(1), 17–24.

Key, E. (1903/1900). *Das Jahrhundert des Kindes*. Berlin: S. Fischer.

Lamers-Winkelman, F., Slot, N. W., & Bijl, B. (2007). Scholieren over kindermishandeling: Het VU-PR research onderzoek. *Pedagogiek, Wetenschappelijk Forum voor Opvoeding, Onderwijs en Vorming, 27*(3), 220–235.

Lamers-Winkelman, F., Slot, N. W., Bijl, B., & Vijlbrief, A. C. (2007). *Scholieren over mishandeling. Resultaten van een landelijk onderzoek naar de omvang van kindermishandeling onder leerlingen van het voortgezet onderwijs*. The Hague: Ministry of Justice/WODC.

May-Chahal, C., & Cawson, P. (2005). Measuring child maltreatment in the United Kingdom: A study of the prevalence of child abuse and neglect. *Child Abuse & Neglect, 29*, 969–984.

Overwater, J. (1948). *Kinderrechtspraak en kinderbescherming*. Alphen aan den Rijn: Samsom.

Overwater, J. (1955). Rede van de voorzitter van de Nationale Federatie "De Nederlandse Bond tot Kinderbescherming". In *Toespraken gehouden ter herdenking van de gouden kinderwetten*

op 1 december 1955, in de Ridderzaal te 's-Gravenhage (pp. 15–23). The Hague: Ministry of Justice.

Paul, J. P. (Ed.). (2005). *Introduction to the philosophies of research and criticism in education and the social sciences*. Upper Saddle River, NJ: Person.

Righthand, S., Kerr, B., & Drach, K. (2003). *Child maltreatment risk assessments. An evaluation guide*. New York: The Haworth Press.

Roelofs, M. A. S., & Baartman, H. E. M. (1997). The Netherlands. Responding to abuse: Compassion or control? In N. Gilbert (Ed.), *Combatting child abuse. International perspectives and trends* (pp. 192–211). Oxford: Oxford University Press.

Rouvoet, A. (2007). *Aard en omvang van kindermishandeling in Nederland* [Letter to the President of the Lower House of the Dutch Parliament, 25 April 2007]. Retrieved April 4, 2007, from http://www.jeugdengezin.nl/nieuwsberichten/2007/kindermishandel-omvangrijk.asp

Rouvout, A. (2007). *Weblog 25 April 2007. Kindermishandeling: vaak een verborgen probleem*. Retrieved April 4, 2007, from http://www.jeugdengezin.nl/organisatie/minister-rouvoet/weblog/overzicht/kindermishandeling

Sariola, H., & Uutela, A. (1992). The prevalence and context of family violence against children in Finland. *Child Abuse & Neglect, 16*, 823–832.

Sidebotham, P., & Golding, J. (2001). Child maltreatment in the "Children of the Nineties". A longitudinal study of parental risk factors. *Child Abuse & Neglect, 25*, 1177–1200.

Stearns, P. N. (2006). *Childhood in world history*. New York: Routledge.

Sternberg, K. J., Lamb, M. E., Guterman, E., & Abbott, C. B. (2006). Effects of early and later family violence on children's behavior problems and depression: A longitudinal, multi-informant perspective. *Child Abuse & Neglect, 30*, 283–306.

Straus, M. A., & Savage, S. A. (2005). Neglectful behavior by parents in the life history of university students in 17 Countries and its relation to violence against dating partners. *Child Maltreatment, 10*(2), 124–135.

Zigler, E., & Hall, N. W. (1990). Physical child abuse in America: Past, present, and future. In D. Cicchetti & V. Carlson (Eds.), *Child maltreatment. Theory and research on the causes and consequences of child abuse and neglect* (pp. 38–75). Cambridge: Cambridge University Press.

Chapter 5
Constructing Social Unity and Presenting Clear Predictions: The Promise of Public Opinion Pollsters to Measure and Educate Society

Norbert Grube

Social sciences create phenomena. This thesis, put forward by Thomas Osborne and Nicolas Rose (1999), means that social sciences do not only construe reality in a manipulative way but detect interrelations which were previously unknown.

Here I want to focus on polls as an instrument to measure people's attitudes and opinions of. I will concentrate on the Allensbach Institute of Public Opinion Research, the oldest German institute that was founded by Erich Peter Neumann (1912–1973) and Elisabeth Noelle (1916) in 1947. It was a private business venture with scientific ambitions. In this study I want to deal with two main areas. In the first part I analyse the self-legitimisation of pollsters as observers of the social in order to strengthen democracy. Here I refer to John Dewey's plea that social sciences should provide the means to create the great community and to the modification of this idea put forward in 1935 by George Horace Gallup (1901–1984), the founder of the American Institute of Public Opinion. Following on from this, I look at post-1945 West Germany. Keeping in mind the critiques of polling made by the likes of Pierre Bourdieu (1930–2002), Vance Packard (1914–1996) and Theodor Adorno (1903–1969), I will examine ethical issues surrounding polling. In the second part, I will show how poll data generate a new national narrative in the era of neo-liberal governments (Foucault, 2004a, 2004b). Do polls create public communication and advance the establishment of democratic participation? Do they merely construe one national body (community) by creating a kind of clear and rational discourse about national topics in order to avoid irrational, radical, or even violent political conflicts? If this is so does it mean that polls intensify the dualistic political scheme of friend versus enemy (Schmitt, 1932/2002, p. 26), thus dividing the population in two or more camps of opinion and excluding minority groups who cannot be integrated into these camps? Opinion research is seen here as a permanent observation of the population, which monthly or annually creates updated trends of opinion formation. Referring to these series pollsters claim to predict future opinions and suggest the

N. Grube (✉)
Institut für Historische Bildungsforschung Pestalozzianum, Zürich, Switzerland

P. Smeyers, M. Depaepe (eds.), *Educational Research: The Ethics and Aesthetics of Statistics*, Educational Research 5, DOI 10.1007/978-90-481-9873-3_5,

precise and rational forecast of future developments. In addition to these claims, pollsters have ambitions to educate the people. Therefore, as I shall show in the last section, they construe respondents as incomplete beings.

5.1 The Creation of National Community: Dewey's and Gallup's Ethical Legitimation of Social Sciences and Polls

Around about 1900, in the period that saw the emergence of empirical social research, expectations of the social sciences rose out of a desire for social coherence. In 'The Public and its Problems' (1927) the pragmatist philosopher John Dewey (1859–1952) worried about the eclipse of 'the public'. Dewey compared public participation and communication in American mass society with rural 18th century communities. Writing in the spirit of historical Romanticism, Dewey is saddened by the gradual disappearance of traditional face-to-face communities and local neighbourhoods. He feels that public participation in problem solving is beginning to fade into the ether. At the beginning of the 20th century, many people were being excluded from the benefits of technological inventions, economic progress and the self-regulation of public affairs. The increasing social anonymity and impersonal voting system weakened the will of average Americans to engage in public affairs (Dewey, 1927/1984, pp. 305, 309–311, 319–320). Instead of local self-regulation powerful special interest groups, who exerted a great influence on politicians, businessmen and bureaucratic experts (Dewey, 1927/1984, p. 321; Lippmann, 1915), governed urban affairs. Like Dewey the American journalist and political spin-doctor Walter Lippmann (1889–1974), registered the alienation of the 'outsiders'. The 'outsiders' were the 'great mass of wage-earners and farmers, of clerks and small managers'. For Lippmann, these groups were alienated from the sphere of political decision-making. This sphere was dominated by well-informed 'insiders' and experts (Lippmann, 1915, p. 35). As a consequence the 'necessarily ignorant' outsiders used their power as voters only irrationally, sometimes even destructively. Lippmann (1925, p. 150) complained that such groups were 'often meddlesome'. This fatalistic behaviour endangered the necessary mobilisation of human resources necessary for both the development of international and economic competition and the drive to make the world safe for democracy. Lippmann and Dewey accused the mass media of seducing the masses (Dewey, 1927/1984, pp. 321–322). Media coverage presented solely selective information, stimulated a prevalent desire for amusement, entertainment and distraction, helped to maintain ignorance all of which led to the inability to overview and judge complex public affairs (Dewey, 1927/1984, p. 306; Lippmann, 1925, pp. 43–44).

Both Lippmann and Dewey hoped that the social sciences could create knowledge about thoughts and attitudes of people, their perceptions of diverse society and their experiences with complex interdependent aspects of modernity. Lippmann wanted to put social sciences, or at least polls, into the service of the government for an efficient and improving government of the masses. In post-war Germany,

Max Horkheimer and Theodor W. Adorno (1969, p. 272) and Jürgen Habermas (1990, p. 321) criticised poll-based governance. They argued that this kind of arcane policy would endanger and erode democratic participation. In contrast to Lippmann, Dewey (1927/1984, p. 339; 1939/1988, p. 156) demanded the 'freedom of social inquiry' and 'full publicity' for its results and interpretations. The equitable diffusion of knowledge should establish scientific attitudes as a precondition for the people to rule public affairs on their own. In Dewey's view (1927/1984, p. 345) 'the communication of the results of social inquiry' forms public opinion and organises popular communication that 'alone can create a great community' (Dewey, 1927/1984, p. 324; Igo, 2007, p. 282). Communication meant reification of subjective experiences. Common shared or communicated experiences were the common ground to integrate individuals into a great community. Finally, Dewey (1939/1988, pp. 167–168) hoped that the creation of scientific attitudes and school education would establish social coherence. Deliberative public opinion should depend on rational comparison and proofs of facts and opinions and not on dogmatic statements and propagandistic camouflage.

5.2 An Instrument for Democracy? Polls in the Public and Governmental Sphere

George Gallup followed Dewey's footsteps. In his 1940 book, *The Pulse of Democracy*, Gallup outlined a utopian view of the potential of polls. He maintained that polling would become the national equivalent of the New England town meeting. It would give a voice to the views of the common man (Igo, 2007, pp. 121–124; Converse, 1987/2009, pp. 122–124). In 1940, Gallup presented poll data in reports that were published up to four times a week in 106 major metropolitan newspapers across the country. These reports were named 'America Speaks!' (Igo, 2007, p. 117; Raupp, 2007, p. 33; Keller, 2001, p. 38). The programmatic title referred not only to notions of establishing a democratic community, but first and foremost to ambitions for creating a new national community. This national ambition became even more relevant following the entry of the United States into the Second World War. Then Gallup and other pollsters surveyed the national will and support for military engagement. At the same time these data were published in order to show the formation of a national war community (Raupp, 2007). Publishers and journalists were the pollsters' first customers. This was also true for another American pioneer of polling, Elmo Roper (1900–1971), who published data in the magazine *Fortune*. It is obvious that the bond between pollsters and the media was not, as recent studies suggest, formed in the latter part of the 20th century (Kruke, 2007, p. 495). Like Gallup and Roper in America, West-German leading poll agencies such as EMNID, Infratest or the Allensbach Institute often published their data in magazines, newspapers, journals, and books from the late 1940s onwards (Bacherer, 1987; Noelle & Neumann, 1956; Köcher, 2010). Allensbach (1969) was the first institute that cooperated with a German TV channel (Zweites Deutsches

Fernsehen) in order to present data during an election campaign. According to Elisabeth Noelle-Neumann's (1980) concept of the 'public opinion research correspondent' the Allensbach Institute monthly publishes results and interpretations in the 'Frankfurter Allgemeine Zeitung', one of the most important newspapers in Germany. Big publishers, magazines, newspapers or broadcast channels ordered polls for studies looking at readership and audience research. In the last decade almost 10% of the annual turnover of the German poll agencies was generated by research for media and publishers (ADM, 2008, p. 11). Today, every important German polling agency surveys the people due to contractual obligations with newspapers, magazines like *Der Spiegel*, *Focus* and TV channels. But the selective presentation of poll data in the media does not only enlighten and strengthen the public discourse. Poll agencies are financially dependent on powerful media companies and consumer goods industries. Therefore, polls create a pseudo-conversation with people. I use the term pseudo-conversation because the questions accord with dominant discourses. Dominant discourses will be enforced if polls strongly refer to or direct attention to them. Interviewees are treated like customers. This clashes with Dewey's idealist vision of a new community characterised by free and transparent communication as polls construct and delimit social reality. They divide the population into social/consumer groups.

However, pollsters, such as those from Allensbach, defined empirical social research as a science that would enhance democracy (Demokratiewissenschaft) (Plé, 1990, pp. 258–259; Neumann, 1952). Post-war Germany provides an interesting historical context for looking at the relationship between polls and democracy. This is partly due to the western allies' attempts to re-educate. Within this process of reeducation the pollsters themselves had educational ambitions. They felt that statistics played an important role in educating the people in accordance with governmental guidelines. Between 1945 and 1949, special sections of allied military governments used polls for the democratic transformation of Germany. These groups included the survey section of the office of the US-Military Government (OMGUS) (Merritt & Merritt, 1970, 1980), the British Public Opinion Research Organization (PORO) and the French Education Publique (Grube, 2007; Kutsch, 1995). In a few instances the reeducation process was influenced by readings of Dewey (Schlander, 1981).

Though it appears that polls help to strengthen democracy, to what extent could pollsters reconstruct Dewey's ideal of free public communication? In the German context, the military administration or the new government were the customers of pollsters. They belonged to the political scene at the beginning of the Cold War era. Therefore, in accordance with government agendas pollsters were prevented from publishing all the data. Furthermore, the rise of polls ran parallel to the consolidation and extension of the national welfare state (Raphael, 1996, p. 182). National governments and bureaucracies became important customers of pollsters. Questioning a sample of the national population can help to reconstruct the national society but not transcending its discursive limits (Igo, 2007).

In post-war West Germany findings from the social sciences were put into the service of the government, and the pollsters implicitly followed Walter Lippmann's suggestions. For example, two or three times a month the Allensbach Institute of

public opinion research sent confidential reports to German chancellors. This began in the autumn of 1950 with poll analyses about the popularity of the reconstruction of the new national army in Germany.[1] This cooperation between the Allensbach Institute and the German chancellery continues to this day. Polls had to calculate the consequences of fundamental political decisions. Prospects about the ongoing support of the rearmament and the establishment of the social market were surveyed.[2] With the support of the Allensbach Institute, the German minister for economy, Ludwig Erhard, observed the support for monetary reform in 1948. Pollsters should throw a prospective view on future developments. Erhard (1962, p. 13) asked these questions, 'How will the increasing wealth strengthen the character, inner power, the soul, and personality of every citizen (. . .)? Will the consciousness and will of freedom increase? Will the human being get rid of the obligations and dependencies?' These questions call to mind Michel Foucault's analysis of German neo-liberalism.

5.3 Contradictory Functions of Poll Statistics: Service for Efficient and Rational Government, National Coherence and the Gaining of Political Power

Although the previous section outlined the strengthening of national government by polls, one can also regard polls as a necessary means to generate knowledge for a self-restricted government that is reliant on a preventative national population policy. This perspective ties in with Erhard's statement and Foucault's writings on German liberalism (2004b, pp. 188, 190, 197–201). Because governments in diverse mass societies cannot guide souls in ways analogous to the pastoral dominated regimes of the 18th century, the market provides the criterion for the regulation of society. The overall aim is the circulation of goods and the mobilisation of human resources in order to improve national welfare. Risks of disease, economic crash and other unforeseeable things shall be avoided because they endanger economic success. Therefore, statistics and data collection on birth rates and mortality rates, epidemics, bumper crops, crop failure and price trends generate facts that help to prepare a government for social and economic crises. Data series should enable the national governors and the scientific experts to minimise or even prevent risk (Foucault, 2004a, pp. 38–41, 71–73, 93–98). Numbers offer possibilities of comparison, give an overview about probabilities of further developments and indicate solutions. In the 20th century the western neo-liberal governments confined themselves to constructing preventative social frameworks. A better exploitation of human capital should lead to increased economic benefits. Quantitative research was no longer limited to the measurement of price trends, harvests and economic conjuncture. Instead, the population as a whole became the focus of psychologists, sociologists, marketing experts and communication experts. Attitudes, opinions or even the mentality of the people could be explored as a precondition for successful, demand-oriented and preventative policy. Jean Jacques Rousseau (1762/2006, pp. 47–48) pointed out that wise legislation depends on verification by the people.

Pollsters therefore ask the opinions of a representative sample of the population. This data is used to survey these opinions in order to govern the people efficiently, reliably and rationally. This will supposedly generate security and reduce risk. With this in mind, the German press officer of the chancellery, Werner Krüger (1969), outlined the need for poll-driven government. In 1948, an advertising leaflet from the Allensbach Institute put forward the view, 'Only the one who knows the market, will avoid [economic – NG] mistakes. Every mistake of production can destroy the basis of subsistence. (. . .) In this [risky – NG] situation we offer to you our service. We are able to create reliable poll analyses with scientific methods of market research that gives you advice of right behaviour. We know the consumers and we discover the wishes you are interested in.'[3]

Though pollsters served governmental policy by predicting the outcome of elections in order to calculate the political future and help avoid the shock of unexpected changes in the balance of power (Schmidtchen, 1965, p. 96), this was not their only function. At the same time polls were seen as an instrument to gain and retain political power, as an instrument for orchestrating as well as controlling the political conflict during election campaigns. Pollsters such as Erich Peter Neumann from the Allensbach Institute became increasingly active as advisors in election campaigns. They collaborated with political propagandists who were using rhetoric to try and sway public opinion (Spicka, 2007). In this context polls have been both instruments of the right and left camps. With implicit reference to the political theory of Carl Schmitt (1932/2002), the Allensbach pollsters classify the voters into these two camps. The respondents have to rate themselves on a scale from zero to hundred, in order to identify the strength of the right and left camps in the 'battle for the climate of opinion' (Noelle-Neumann, 1998, p. 324). This dualistic division of the national population should serve for an efficient election campaign and relates to the dichotomic political propaganda in the bipolar Cold War. In this sense pollsters contrasted the advantages of the social market economy or the European unification with the socialist command economy and national neutrality (Grube, 2007). In contrast to Dewey's vision of a community of public communication, Neumann (1966) supported governmental and even dualistic propaganda that instructed the people in order to close the gap between political leaders and the masses therefore creating acceptance and trust towards the government.

In the period of the Cold War pollsters and the conservative government regarded the creation of social coherence as necessary to compete with the socialist nations. In their view socialist collectivism was superior to the social diversity of the western world. Therefore, 'political polls could serve to minimize currently dominating misunderstandings between the government and the governed people and add to agreement and a process of integration' (Noelle-Neumann, 1957, p. 17). Despite the various classifications of the respondents into different social groups, polls should generate national unification.

At first glance it seems to be a paradox that the dualistic division of the population into left and right, workers and employees, democrats and non-democrats, protestant and catholic should help to form the nation as a whole. But it is exactly

Table 5.1 Marking the enemy in the dualism of the Cold War

'Do you think the Western democracies and Communism in the East can, in the long run, live together in peace?'

	Yes (%)	No (%)	No opinion (%)
1954, December	20	66	14
1955, June	32	49	19
1955, September	34	49	17
1955, December	21	61	18
1956, July	32	52	16
1956, November	36	54	10

Noelle and Neumann (1967, p. 598)

this poll driven simplified dual division of the social and of the bipolar world order (Table 5.1) that is the basis for the formation of national conformity. Dualistic questions about the superiority of America against The Soviet Union or about the coexistence of both imperia should strengthen the western camp and West German conformity.

The creation of national unity – 'to lead the minority to the opinion of the majority' – is the task of poll-based political leadership (Neumann, 1964, p. 23). In this sense, Neumann (1963, p. 25) stated, 'information about public opinion [. . .] (is) the basis, that he [the politician – NG] is able to direct it [public opinion – NG] to the point of his belief and conviction'. The regulation and harmonisation of popular attitudes was the main aim of polls. In regard to this point the German sociologist Helmut Schelsky (1912–1984) (1959, pp. 115, 121) partially agreed with Neumann. If the government renounces socio-psychological findings in the process of guiding and leading the people, then anarchy will be the result (Neumann, 1964, p. 27). Neumann's sorrows were generated by experiences of violence and social conflicts in 1932 and 1933. Therefore, he and other German sociologists, like Helmut Schelsky, were sceptical about public controversy. In their view people were not ready or prepared for democracy.

This double function of polls, which served bio-political aims in the long run and national or political coherence in the short term, led to criticism of the polls. They were regarded as an undemocratic and unethical instrument of empowerment ('Bemächtigungsinstrument' – Sonnemann, 1963, p. 164) or as an instrument of rational demagogy (Bourdieu, 1997, pp. 129, 132). Pollsters were presented as hidden persuaders (Packard, 1957/1992) who would stabilise the existing and established political system and elite by forming the political language and the understanding of political messages (Frevert, 2005, p. 15). But the estimation of pollsters as almighty persuaders is an exaggeration. According to Bourdieu (1997, pp. 132–133), polls only construct artefacts and it's doubtful that they can capture the self-will of the so-called inside dopester (Riesman, 1961, pp. 196–198), the other-directed respondent.

5.4 The Suggestive Aesthetics of Charts: The Clearness of Social Coherence or Camps, the Security of Prediction Due to Reliable Observation and Permanent Comparison

In Section 5.5, I will focus on some statistical charts in order to analyse how pollsters try to create a new dualistic national narrative that aims to unify the nation. I will focus on some techniques of questioning and on the presentation of data. The translation of analyses of opinion formation into numerical data (Manhardt, 2008; Keller, 2001) suggested non-scrutinised evidence and often divided the population into two contrary camps. Since the late 1940s, pollsters have observed the process of democratisation of Germany. To this end they created questions that divided the population into two parts: one part that was ready for democracy and the other part that was not prepared to participate in political decisions. This division often becomes manifest in questions that contain short pseudo-dialogues. Let us consider examples of questions pertaining to the values of freedom and equality. Two contrary opinions are presented in two silhouettes. One says, 'I think that freedom and equality are equally important. But if I had to choose between the two, I would consider personal freedom more important, that is, for people to be able to live in freedom and not to be restricted in their personal development'. The other says, 'Certainly both freedom and equality are important. But if I had to choose between the two, I would consider as much equality as possible to be more important, that is, for no one to be underprivileged and for class differences not to be so strong' (Noelle-Neumann, 1995, p. 33, 1998, p. 331). The respondents have to rate themselves in accordance with one of these two statements. Differentiated opinions or precise definitions of the two terms freedom and equality are not presented. Another example of this dualistic approach to questioning demands that the respondent decide if he or she is in favour of authoritative or parliamentary government (Noelle-Neumann, 1995, p. 28). Other dichotomic questions forge the respondents to confess their happiness or their sorrows about German unification. In another case the respondents have to come out in favour of democracy or of a different form of government. These divisions indicate that only a selected portion of citizens is authorised to participate, whereas the remainder has to obey and adapt to governmental guidelines or prevalent social norms. The complexity of political attitudes is reduced to questions that evoke dualistic answers: 'Yes' or 'No', 'Agree' versus 'Not agree'. For example, the Allensbach pollsters judged responses to the questions 'Do you think most people can be trusted' and 'Are you for or against capital punishment' as indicators for democratic capabilities in the era of the Cold War and German unification (Noelle-Neumann, 1995, pp. 38, 39). Both questions construed two artificial camps (Fig. 5.1, 5.2 and 5.3):

The democratic camp is the one that rejects the death penalty and trusts most people. The other mistrustful part endorses the death penalty and seems to be unprepared for democracy – it favours authoritarian forms of government. This division of the confused post-war Germans into two clear groups suggests the possibility of rational, efficient and successful government of the people as a whole. This is because these charts and numbers eliminate social diversity by merging

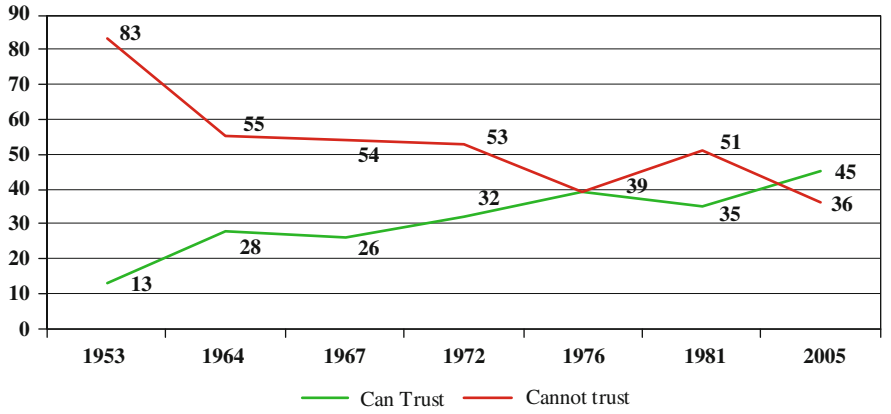


Fig. 5.1 Mistrustful Germans = undemocratic Germans? 'Do you think that most people can be trusted?'

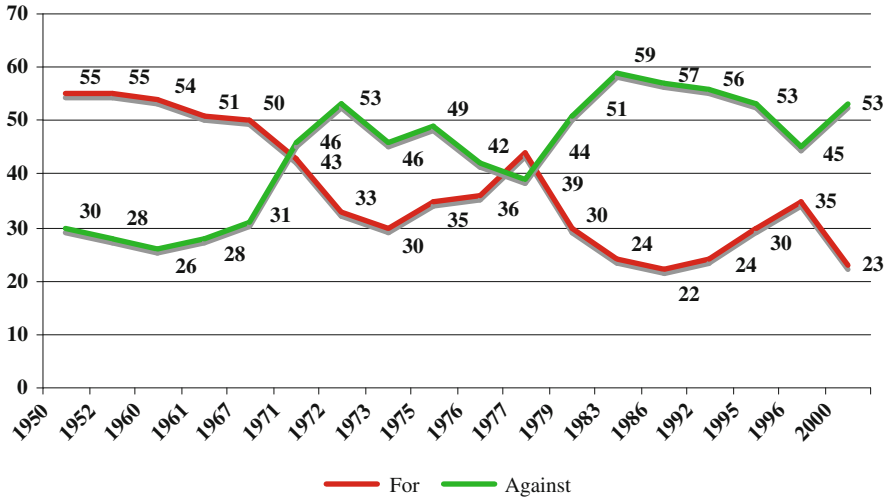


Fig. 5.2 'Are you, on principle, for or against capital punishment?'

different opinions and attitudes into two camps. At the same time these data series seem to guarantee a future in which mass democracy can be observed, controlled, and formed. But this simplification of public affairs and communication does not improve public participation in the sense put forward by Dewey.

It should be noted that even the Allensbach pollsters doubted if this dichotomy correctly represents the complexity of opinion formation. Therefore, for a long time they preferred so-called open questions because the answers to closed questions are 'directly impregnated with the atmosphere of the interrogation'.[4] Up until the middle of the 1950s, many reports by the Allensbach Institute contain a collection

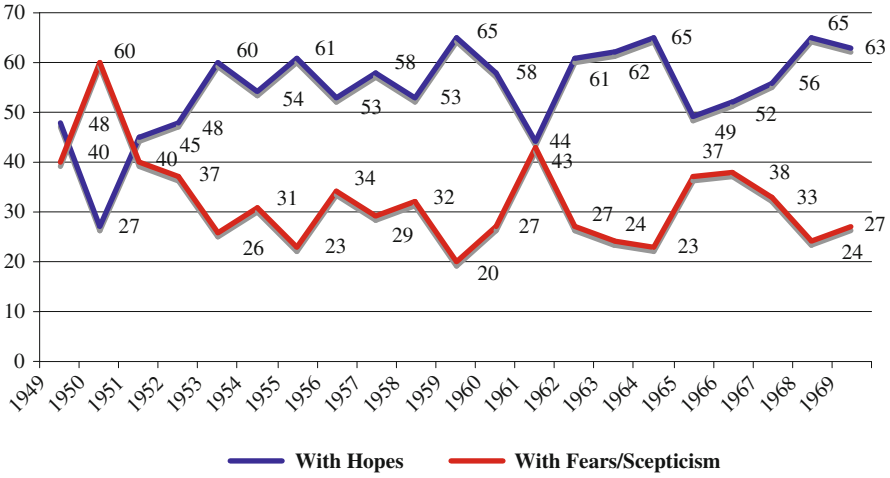


Fig. 5.3 'Is it with hopes or with fears that you enter the coming year?' N. sample: West Germany includes West-Berlin, population 18 years and older Allensbach Institue

of answers towards open questions. In contrast to multiple-choice questions, open questions allow for the collection of various answers. These answers expose the artificial construction of the dualistic classification of opinion because they show that attitudes concerning the death penalty, democracy or trust are based on different ideologies and evoke a wide range of estimations that cannot be detected and presented by polls. For example, the experience of National Socialism provided the background against which religious and pacifist respondents saw the death penalty as an adequate instrument to punish war criminals. But these respondents favouring the death penalty could just as well be democrats. Today in the U.S., two-thirds of all respondents of a Gallup poll sanctioned the death penalty in autumn 2009 (Newport, 2009).

Nevertheless, the simplified presentation of two opposite trends and statistical series has been established as a comprehensive overview that is suitable for publication in the mass media. It seems paradoxical that this sort of publication of poll data does not offer a wide range of information, but supports the simplification of complex results (Institut für Demoskopie Allensbach, 1969, pp. 26, 33). The graphical aesthetics do not only suggest credibility and plausibility. Statistical facts suggest an ostensible traceability and an insight in the formation of opinion. But these dualistic classifications of popular opinions offer only selective results instead of distinguished interpretations. They do not comply with Dewey's demand for the full publicity of results from the social sciences.

The presentation of two clear data trends suggests the scientific omnipotence to overview the complex formation of opinion. The graphical aesthetics of two trends demonstrate the permanent observation of the social and therefore promise security in the process of political planning and decision-making. For example, in post-war West Germany the Allensbach Institute measured the public support

for policy decisions related to rearmament or the popularity of the social market economy. This was conducted on an almost monthly basis. Questions included 'Do you agree or disagree with Adenauer's policy on the whole?' or 'Do you think that prices have generally remained constant during the last three months, or that they have risen or fallen?' (Noelle & Neumann, 1967, pp. 256–264, 368–369, 438–439). These questions again serve to form two camps of supporters and opponents and implicitly refer to Carl Schmitt's division of friend and enemy as the main category of the political. The dualistic data series should enable the government to consider the political atmosphere and calculate the possible public reactions towards political decisions. This governmental use of polls promotes the taming of controversial public communication and the control of democracy. Nothing unforeseen should happen to the government. Polls and election forecasts should ensure that voters do not surprise the politicians.

The data series suggest both the possibility of prediction and the prevention and/or containment of an uncertain future. Data series allow temporally comparisons. Looking back to the past, the pollster can interpret the present. By the same token, she can also predict future trends. Clear predictions can, it is argued, be generated from answers to the question, 'Is it with hopes or with fears that you enter the coming year?' (Fig. 5.3). The results reveal the aesthetic of a reflected image, as if the two lines of opinion are conditioned by each other and exclude other possible opinions.

Allensbach pollsters, like Elisabeth Noelle-Neumann (1989), are of the opinion that answers to this question do not only reveal optimism or pessimism. In addition to this, they help to predict economic growth in the following year. These answers are, she argues, more reliable than the predictions put forward by business researchers. Interestingly, the assumption of the public is a prophet contradicts the view that human beings are undemocratic and incomplete beings in need of political and educational leadership. We can therefore detect inconsistencies in pollsters' assumptions. For example, Gerhard Schmidtchen (1965, pp. 81, 98, 213), a former member of the Allensbach Institute, describes human beings as ignorant and resigned. However, at the same time he maintains that they are as interested in and well-informed as regards political affairs. These inconsistencies alongside the perceived need to renew models for questioning may explain the recent unreliability of polls. In 2002 and 2005, election forecasts made by the Allensbach Institute were wrong. Pollsters failed to anticipate the last minute swing in the vote. The emergence of new media like the Internet, different forms of mass communication and new kinds of voting behaviour generated factors that could not be accounted for by the polls. With these changes it would seem that individuals have adapted elements of the aforementioned other-directed character (Riesman, 1961) that permanently observes his environments in order to form and change his opinion.

Though a data series may remind us of a medical fever chart, the empirical social sciences could not keep up with the natural sciences in regards to validity. This did not stop pollsters from Allensbach adopting the language of medical science by using terms such as 'diagnose' and 'therapy' (Neumann, 1956).[5] George Gallup even claimed that he could check 'the pulse of democracy'.

Data series reduce the complexity of social reality. Politicians, journalists and some scientists do not think about the genesis of data and the context of production. The data often provide the bedrock for various discourses, but debates on the construction of data are rare. There are only a few examples of occasions on which poll data are scrutinised in the public sphere. One is the medial critique about bad election forecasts, but this critique becomes ritualised. Another example is the 1986 debate about the genesis of a study by Allensbach on unemployment (Noelle-Neumann & Gillies, 1987). Even the German parliament discussed the reliability of the data. That said this rare degree of critical interest in polling results arose due to the parallel election campaign.

Despite these critiques polls initiate and dominate public discourses because they enable intra- and international comparisons. Counting and measuring are the preconditions for comparison and this is the task of statisticians. As Jean Jacques Rousseau (1762/2006, p. 92) exclaimed, 'Now, it's your turn statisticians: count, measure and compare'.

Comparisons can be seen as an instrument to calm political emotions and to eliminate alternative solutions so that the supposed 'best group' should serve as a model for others.

This can be illustrated by 2002 polls about people's satisfaction with the German school system following the publication of the PISA results. In the PISA test, pupils from Bavaria and Baden-Württemberg did better than pupils from Berlin and Northern Germany. In correlation with the PISA result, nationally published polls from Allensbach revealed agreement with PISA's findings on Bavaria (Fig. 5.4),

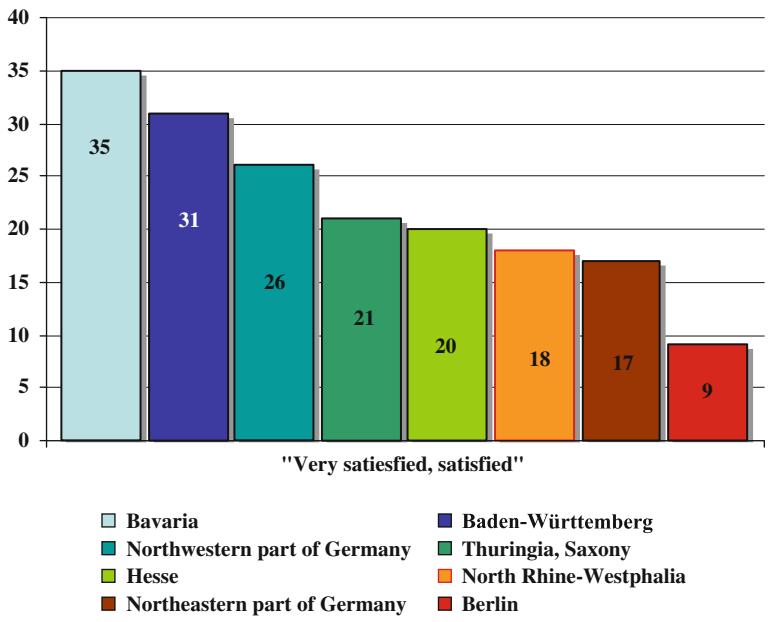


Fig. 5.4 Satisfied with the German school system?

Table 5.2 International comparison of national pride

'Are you proud to be German/French/English/American? Would you say you are . . .'							
	West Germany		East Germany		France	England	USA
	1991 (%)	2001 (%)	1991 (%)	2001 (%)	2001 (%)	2001 (%)	2001 (%)
'Very proud'	16	20	19	12	35	52	80
'Rather proud'	41	42	43	45	46	35	18
'Not so proud'	22	17	17	21	6	6	2
'Not at all proud'	11	8	8	8	3	1	*x*
Undecided	10	13	13	14	10	6	*x*

Institut für Demoskopie Allensbach 2001/2002

but differed with regard to other federal states. The message in the year of the federal election was clear: the conservative school policy in Bavaria is successful and should serve as a model for others.

Data charts enable comparison and comparison shall initiate competition. Therefore, one can say that polls make knowledge available for neo-liberal government policy. In the period of competing nation states and competing ideological systems during the Cold War, the international comparison was an important political instrument to show national efficiency, strength and performance. Various data were used to underline national superiority or to initiate reforms. This can be illustrated by the German discussion (that took place in the 1980s) on whether or not Germany was a weak nation compared to others (Noelle-Neumann & Köcher, 1988). This discussion became heated due to the alleged lack of national pride in comparison to other nations such as the United States, England and France (Table 5.2).

Here the comparison of poll data led to an analysis of national deficits. Similar polls in the 1950s had construed national deficiencies in the process of reeducation and democratisation as I outlined before.

5.5 The Limits of Polls as an Instrument to Create the Great Communication Community

Polls often present the respondents as incomplete human beings. This is because since the 1950s the interviews sometimes tended to examine elementary knowledge, especially when they were conducted for the government or politically engaged entrepreneurs. Such tests intensified the simplification of social reality. The respondents were examined on their mathematical, geographical or orthographical skills. They had to solve arithmetical problems like one-half plus one-tenth and were asked to indicate Germany on the map. They were tested to see if they could write words such as 'telephone', 'credit', 'director' and 'rhythm' correctly.[6] Recently pollsters

Table 5.3 Human beings – incomplete beings?

	'Do you know what. . .	No (%)
1952	. . . a functionary is'?	29
1954	. . . an emigrant is'?	40
	. . . opposition means'?	57
1951	. . . "subsidies" are'?	58
1952	. . . national sovereignty is'?	60
1954	. . . "to ratify" means'?	71
1954	. . . federalism means'?	87
1961	. . . the Federal Council is'?	90

Noelle and Neumann (1967, p. 119)

from Allensbach have maintained their ambitions as regards the verification of adult citizens' educational levels (Petersen, 2008). In political affairs polls underline and construct the widespread ignorance of the people, but the pollsters do not consider the situation of respondents who must react to surprising questions. Polls even present experts as ignorant, because they are not as prepared to answer the questions as the interviewers are (Schmölders, 1966, pp. 190–215). Most of the respondents could not explain terms such as 'national sovereignty', 'NATO', 'grand coalition', 'Federalism' or 'opposition' (Table 5.3).

Furthermore, many Germans did not know Charlemagne's date of birth or if Martin Luther had lived before or after Thirty Years' War (Noelle & Neumann, 1967, pp. 116–120). These artificial tests were supposed to reveal clear results about deficits in political knowledge. But these interrogations only served the interests of political leaders. This sort of questioning presupposes that the respondent is simultaneously rational and ignorant/irrational. The pollsters did not consider the outlook on life of most of the respondents and could not detect capacities and the potential of people who were overwhelmed with the tasks of everyday life. Many of the respondents recognise that polls transform into tests and keep silent. This is revealed by both declining response rates (Koch & Porst, 1998). But these results were grist to the mill of pollsters and politicians who have often stressed that 80% of West Germans had only visited an elementary school. They then demanded the improvement of political education.

Because pollsters from Allensbach observe the population and explain the data they feel competent to give advice on education. Referring to poll data Neumann complained, 'Bad school education, insufficient information by a psychological ineffective edited press, superficial political education, distraction by material sorrows: these are the preconditions that stand against a lively relationship between the mass and political and economic leaders. The deficiency of systematic explanation must lead to disgust with the existing system (. . .)'.[7]

Pollsters and other scientists had various educational ambitions. Paul F. Lazarsfeld, who founded the 'Bureau of Applied Social Research', supported the educational ambitions of applied social research. Such ambitions included the elimination of racism, a change in consumer behaviour and greater comprehension of

international issues (Lazarsfeld, Berelson, & Gaudet, 1944/1969, p. 34). In a similar way Karl Mannheim (1951, pp. 13–23) and Lippmann (1922/1990, pp. 264–275) demanded that a social psychological diagnosis of people's education should act as a basis for 'planning for freedom'. Hartmut von Hentig, the famous reform pedagogue from Bielefeld, aimed to prepare people in West Germany for citizenship and responsibility (von Hentig, 1963, quoted from Von Hodenberg, 2006, p. 55). In contrast to this, Erich Peter Neumann (1965, p. 29) from Allensbach pointed out that politics had to fulfil important mass educational tasks: 'Actually systematic examinations have generated almost unambiguous indications that the relationship between the citizen and the state are more at risk the fewer the citizen knows about the state' (Neumann, 1955). Indeed political skills and knowledge should be improved – especially among the youth. But Neumann judged political education as an important precondition to strengthen public support for governmental leadership in West Germany.

Allensbach pollsters did not see data per se as an instrument to create public communication or the kind of education that Dewey wanted. In contrast to Dewey they thought that statistical data, series and charts are too abstract for most superficially educated people. This seems to be a paradox because the presentation of data and charts should suggest clear regulation of the national body as a means to create social coherence. However, Neumann was of the opinion that the publication of pure data could neither create acceptance of the government decisions nor convince the population of political programs: 'It is evident that there is no easier way of going over people's heads than by using statistics. The majority of the population thinks that the proportion between 100,000 marks and 100,000,000 marks differs only in the different quantity of zeros'.[8] Therefore, Neumann wanted to reserve poll data exclusively for the political elite and public intellectuals. Pollsters or special journalists should comment and interpret quantitative results to avoid the misreading of the data. Here, we discover a contradiction that cannot be resolved: the legitimisation of polls derives from their capacity to improve democratic or at least national mass communication. Polls seem to be a means to restore the great community. However, they often present the respondents as incomplete and ignorant human beings who are not able to deal with public affairs. It seems that polls aggravate the gap between experts, governors and the people even though the results are often published.

5.6 Conclusion

Neumann's statement reveals a bizarre and paradoxical mistrust of the possibility that polls can create and strengthen public communication in Dewey's sense. Poll data should create clarity. It should help with prediction and comparison. However, its main purpose ended up being the elaboration of political or governmental strategies by political experts. The publication of data should not and did not serve to establish free communication among people but convinced them of the veracity of governmental aims and dominant norms. If the majority of respondents comply with

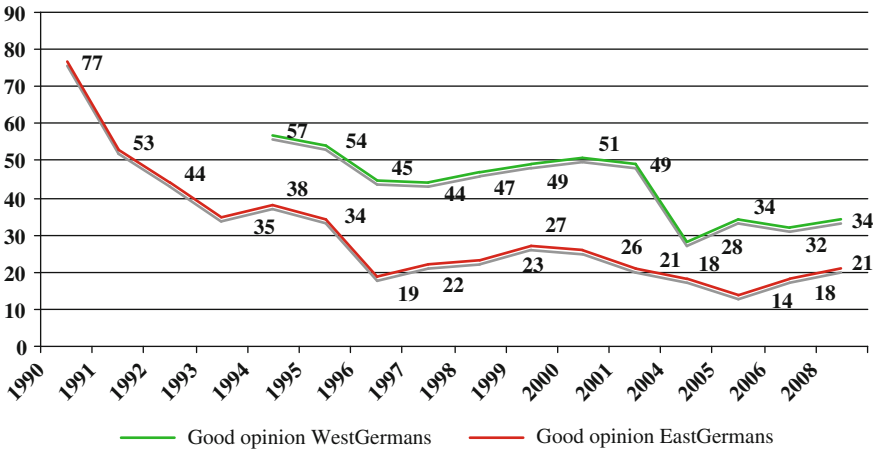


Fig. 5.5 'Do you have a good or bad opinion of the Federal Republic's economic system?'

these norms, the data attest its rightness. If not, the minority part is interpreted as *pars sanior* and the pollsters aim to convince the seduced majority through propaganda and educational efforts. Even the mediatisation of polls is interwoven with medial, economic and political power.

Nevertheless, with the creation of clear, almost dualistic ideological camps, pollsters aimed to level social diversity in favour of a coherent national body. The construction of a national narrative should enable intra- and international comparisons and competition that intensify the national narrative. The simplification and reduction of social complexity by charts and curves should serve to rationalise political decision-making processes. But in contrast to this ambition, pollsters themselves contribute to new hypes by writing and publishing alarming reports if unexpected outliers confuse the expected course of the chart. Little support for the government, little trust in democracy and/or social market economy motivate pollsters to exaggerate warnings about an incalculable future (Köcher, 2008a, 2008b). And it's grotesque that these alarming warnings don't question the pollsters' abilities to make predictions. Furthermore, these alarmist reports contradict the Anspruch to rationalise the political decision-making process and the formation of opinion. Unfortunately, they get a lot of public attention because the audience shall take data and the pollsters' promise to predict the future for granted.

Notes

1. Allensbach Archives, IfD-Report 'Die Stimmung im Bundesgebiet', No. 1, November 1950. IfD-Report 'Die Stimmung im Bundesgebiet' No. 2: Anwachsen der Opposition. Die Entscheidung über die Wiederbewaffnung. November 1950.
2. Allensbach Archives, IfD-Report 2: Sonderumfrage Währungsreform I, 26.-30.6.1948. IfD-Report 3: Sonderumfrage Währungsreform II, 17.-22. Juli 1948.
3. Allensbach Archives, History of the Institute: Advertising leaflet of the institute, 1st July 1948.

4. Allensbach Archives, IfD-Report 205/II Die soziale Spannung. Poll Report in order of the Gesellschaft für Gemeinschaftswerbung. November 1952, p. 64. Compare IfD-Report 112/II Die Haarpflege. Schaumpon – Haarwasser – Öl und Fixativ. Die Markenwahl/Kaufgewohnheiten. Umfrage vom November 1950. IfD-Report 153/I Jugend zwischen Ost und West. Eine Umfrage unter ostdeutschen FDJ-Angehörigen in West-Berlin im Auftrag der Deutschen Korrespondenz. November 1951. IfD-Report 153/III 'Was in Deutschland anders werden sollte'. IfD-Report 37: Ist Deutschland antisemitistisch. Ein diagnostischer Beitrag zur Innenpolitik. Herbst 1949. IfD-Report 22: Goethe 1949. Funkbearbeitung einer Massen-Umfrage in drei Teilen. Von Hans Georg Brenner, Erich Peter Neumann und Elisabeth Noelle im Auftrag des Nachtprogramms des NWDR.
5. Allensbach Archives, IfD-Report 37: Ist Deutschland antisemitisch. Ein diagnostischer Beitrag zur Innenpolitik. Herbst 1949. Allensbach Archives, Dep. History of the Institute: Letter of Elisabeth Noelle-Neumann to Max Horkheimer, 7th July 1963.
6. Allensbach Archives, IfD-report 645: Studien zu Schulfragen, 1958, chart A11-A21.
7. Allensbach Archives, IfD-report 154/I: Das soziale Klima, S. 3.
8. Allensbach Archives NL EPN, Korrespondenz I, Exposé 'Die Frage, welche Möglichkeiten für eine erfolgsversprechende Wahlkampagne' (1953) bestehen, o.V. [Neumann], o.D. IfD-Report 552 Das Zentrum der Krise. März 1957, pp. 5, 8, 10.

References

ADM – Arbeitskreis Deutscher Markt- und Sozialforschungsinstitute. (2008). Annual report 2008 [online report]. Retrieved February 5, from, http://www.adm-ev.de/fileadmin/user_upload/PDFS/Jahresbericht_08.pdf

Bacherer, K. (1987). *Geschichte, Organisation und Funktion von Infratest*. Salzburg: Instituts für Publizistik und Kommunikationswissenschaft.

Bourdieu, P. (1997). *Die verborgenen Mechanismen der Macht. (Schriften zu Politik & Kultur 1)*. Hamburg: VSA-Verlag.

Converse, J. (2009). *Survey research in the United States. Roots and emergence 1890–1960*. New Brunswick, NJ: Transaction Publishers.

Dewey, J. (1927/1984). The public and its problems. In J. A. Boydston & B. A. Walsh (Eds.), *John Dewey: The later works, 1925–1953* (Vol. 2: 1925–1927, pp. 235–374). Carbondale, Edwardsville: Southern Illinois University Press.

Dewey, J. (1939/1988). Freedom and culture. In J. A. Boydston & B. Levine (Eds.), *John Dewey: The later works, 1925–1953.* (Vol. 13: 1939, pp. 63–188). Carbondale: Southern Illinois University Press.

Erhard, L. (1962). *Fragen an die Meinungsforschung*. Allensbach: Verlag für Demoskopie.

Foucault, M. (2004a). *Geschichte der Gouvernementalität. Vol. 1: Sicherheit, Territorium, Bevölkerung.* Frankfurt/Main: Suhrkamp

Foucault, M. (2004b). *Geschichte der Gouvernementalität. Vol. 2: Die Geburt der Biopolitik*. Frankfurt/Main: Suhrkamp.

Frevert, U. (2005). Neue Politikgeschichte: Konzepte und Herausforderungen. In U. Frevert & H.-G. Haupt (Eds.), *Neue Politikgeschichte. Perspektiven einer historischen Politikforschung* (pp. 7–26). Frankfurt/Main: Campus.

Gallup, G. H., & Rae, S. (1940). *The pulse of democracy. The public opinion poll and how it works*. New York: Simon & Schuster.

Grube, N. (2007). Das Institut für Demoskopie Allensbach und die "Deutschen Lehrerbriefe" als Instrumente staatsbürgerlicher Erziehung? Ansprüche und Umsetzungen 1947 bis 1969. *Jahrbuch für Historische Bildungsforschung, 13*, 267–288.

Habermas, J. (1990). *Strukturwandel der Öffentlichkeit. Untersuchungen zu einer Kategorie der bürgerlichen Gesellschaft*. Frankfurt/Main: Suhrkamp.

Horkheimer, M., & Adorno, T. W. (1969). *Dialektik der Aufklärung. Philosophische Fragmente*. Neuausgabe, Frankfurt/Main: Fischer Verlag.

Igo, S. E. (2007). *The averaged american. Surveys, citizens, and the making of a mass public*. Cambridge; MA: Harvard University Press.
Institut für Demoskopie Allensbach (Ed.). (1969). *Wählermeinung – nicht geheim*. Allensbach: Verlag für Demoskopie.
Institut für Demoskopie Allensbach (Ed.). (2001/2002). *Internationale Wertestudie 2001/2002. Ein Überblick über die wichtigsten Ergebnisse*. Allensbach: Author.
Keller, F. (2001). *Archäologie der Meinungsforschung. Mathematik und die Erzählbarkeit des Politischen*. Konstanz: UVK.
Koch, A., & Porst, R. (Eds.). (1998). *Nonresponse in survey research*. Mannheim: ZUMA.
Köcher R. (2008a, June 11). *Akzeptanz der sozialen Marktwirtschaft sinkt. Wirtschaftswoche* [online report]. Retrieved February 5, from http://www.wiwo.de/politik-weltwirtschaft/akzeptanz-der-sozialen-marktwirtschaft-sinkt-295474/
Köcher R. (2008b, October 2). Ost und West. Auch 18 Jahre nach der Wiedervereinigung grosse Unterschiede. *Wirtschaftswoche* [online report]. Retrieved February 5, from, http://www.wiwo.de/politik-weltwirtschaft/ost-und-west-auch-18-jahre-nach-wiedervereinigung-grosse-unterschiede-308606/
Köcher, R. (Ed.). (2010). *Allensbacher Jahrbuch der Demoskopie, Bd. 12: Berliner Republik (2003–2009)*. Berlin: Saur.
Krüger, W. G. (1969). Wissenschaftliche Beratung der Politik – Theorie und Praxis. In *Politisches Seminar der Staatsbürgerlichen Vereinigung 1954 e.V. Wissenschaftliche Beratung in Wirtschaft und Politik*. Bergisch Gladbach: Joh Heider, *17*, 50–61.
Kruke, A. (2007). *Demoskopie in der Bundesrepublik Deutschland. Meinungsforschung, Parteien und Medien 1949–1990*. Düsseldorf: Droste.
Kutsch, A. (1995). Einstellungen zum Nationalsozialismus in der Nachkriegszeit. Ein Beitrag zu den Anfängen der Meinungsforschung in den westlichen Besatzungszonen. *Publizistik, 40*(4), 415–447.
Lazarsfeld, P. F., Berelson, B., & Gaudet, H. (1944). *The people's choice. How the voter makes up his mind in a presidential campaign*. New York: Duelle, Sloan, Pearce. (German translation, 1969. *Wahlen und Wähler. Soziologie des Wählerverhaltens*. Neuwied, Berlin: Luchterhand).
Lippmann, W. (1915b, November 13). Insiders and outsiders. *The New Republic, 54*, 35–36.
Lippmann, W. (1922/1990). *Die öffentliche Meinung*. Bochum: Universitätsverlag Brockmeyer.
Lippmann, W. (1925). *The phantom public. A sequel to "Public Opinion"*. New York: Macmillan.
Manhardt, S. (2008). Vermessene Moderne. Zur Bedeutung von Maß, Zahl und Begriff für die Entstehung der modernen Kultur. In D. Baecker, M. Kettner, & D. Rustemeyer (Eds.), *Über Kultur. Theorie und Praxis der Kulturreflexion* (pp. 191–208). Bielefeld: Transcript.
Mannheim, K. (1951). *Diagnose unserer Zeit. Gedanken eines Soziologen*. Zürich: Europa-Verlag.
Merritt, A. J., & Merritt, R. L. (Eds.). (1970). *Public opinion in occupied Germany. The Omgus surveys, 1945–1949*. Urbana: University of Illinois Press.
Merritt, A. J., & Merritt, R. L. (Eds.). (1980). *Public opinion in semisovereign Germany. The HICOG surveys, 1949–1955*. Urbana: University of Illinois Press.
Neumann, E. P. (1952). Politische und soziale Meinungsforschung in Deutschland. In Institut zur Förderung öffentlicher Angelegenheiten (Ed.), *Empirische Sozialforschung. Meinungs- und Marktforschung. Methoden und Probleme* (pp. 44–51). Frankfurt: Author.
Neumann, E. P. (1955, September 5). Studien über die Verankerung der Demokratie in Deutschland. ESOMAR/WAPOR-Kongress, Konstanz.
Neumann, E. P. (1956, February 16). Politische Umfragen als Informationsmittel. Vortrag vor dem Bundespresseamt, Bonn.
Neumann, E. P. (1963, May 5). Meinungsforschung und Politik. Gesellschaft für Wehrkunde, Bonn.
Neumann, E. P. (1964). Wiedervereinigung in der öffentlichen Meinung. *Die politische Meinung, 9*(91), 19–31.
Neumann, E. P. (1965). Probleme unseres Frankreich-Bildes. *Die politische Meinung, 10*(September), 25–29.

Neumann, E. P. (1966). Der gute Wolf. Über den Verzicht auf Propaganda. *Die politische Meinung, 11*(114), 3–5.

Newport, F. (2009). In *U.S., Two-Thirds Continue to Support Death Penalty. Little change in recent years despite international opposition* [online report]. Retrieved from February 5, from, http://www.gallup.com/poll/123638/in-u.s.-two-thirds-continue-support-death-penalty.aspx

Noelle, E., & Neumann, E. P. (Eds.). (1956). *Jahrbuch der öffentlichen Meinung. Bd. 1: 1947–1955*. Allensbach: Verlag für Demoskopie.

Noelle, E., & Neumann, E. P. (Eds.). (1967). *The Germans. Public opinion polls 1947–1966*. Allensbach: Verlag für Demoskopie.

Noelle-Neumann, E. (1957, December 3). *Politische Meinungsforschung und demokratische Praxis*. (Vortrag vor der Arbeitsgemeinschaft Sozialdemokratischer Akademiker, München.).

Noelle-Neumann, E. (1980). The public opinion research correspondent. *Public Opinion Quarterly 44*(4), 585–597.

Noelle-Neumann, E. (1989). The public as prophet. Findings from continuous survey research and their importance for early diagnosis of economic growth. *International Journal of Public Opinion Research, 1*, 136–150.

Noelle-Neumann, E. (1995). Juan Linz's dissertation on West Germany: An empirical follow-up, thirty years later. In H. E. Chehabi & A. Stephan (Eds.), *Politics, society, and democracy. Comparative studies* (pp. 13–41). Boulder: Westview Press.

Noelle-Neumann, E. (1998). A shift from the Right to the Left as an indicator of value change: A battle for the climate of opinion. *International Journal of Public Opinion Research, 10*(4), 317–334.

Noelle-Neumann, E., & Gillies, P. (1987). *Arbeitslos. Report aus einer Tabuzone*. Frankfurt/Main: Ullstein.

Noelle-Neumann, E., & Köcher, R. (1988). *Die verletzte Nation. Über den Versuch der Deutschen, ihren Charakter zu ändern* (2. Auflage). Stuttgart: DVA.

Osborne, T., & N. Rose (1999). Do the social sciences create phenomena? The example of public opinion research. *British Journal of Sociology, 50*(3), 367–396.

Packard, V. (1957/1992). *Die geheimen Verführer. Der Griff nach dem Unbewussten in jedermann*. Düsseldorf: Econ.

Petersen, T. (2008, November 19). Der Bildungsstand der Erwachsenen. Die Kenntnisse der Schüler sind gründlich erforscht. Was aber die Erwachsenen wissen und für wissenswert halten, ist nur mit Umfragen zu erkunden. *Frankfurter Allgemeine Zeitung,* p. 5.

Plé, B. (1990). *Wissenschaft und säkulare Mission. 'Amerikanische Sozialwissenschaft' im politischen Sendungsbewußtsein der USA und im geistigen Aufbau der Bundesrepublik Deutschland*. Stuttgart: Klett-Cotta.

Raphael, L. (1996). Die Verwissenschaftlichung des Sozialen als methodische und konzeptionelle Herausforderung für eine Sozialgeschichte des 20. Jahrhunderts. *Geschichte und Gesellschaft, 22*, 165–193.

Raupp, J. (2007). *Politische Meinungsforschung. Die Verwendung von Umfragen in der politischen Kommunikation*. Konstanz: UVK.

Riesman, D. (1961). *Die einsame Masse. Eine Untersuchung der Wandlungen des amerikanischen Charakters*. Hamburg: Rowohlt.

Rousseau, J.-J. (1762/2006). *Vom Gesellschaftsvertrag oder Grundsätze des Staatsrechts*. Stuttgart: Reclam.

Schlander, O. (1981). Der Einfluss von John Dewey und Hans Morgenthau auf die Formulierung der Re-Educationpolitik. In M. Heinemann (Ed.), *Umerziehung und Wiederaufbau. Die Bildungspolitik der Besatzungsmächte in Deutschland und Österreich* (pp. 40–52). Stuttgart: Klett-Cotta.

Schelsky, H. (1959). *Ortsbestimmung der deutschen Soziologie*. Düsseldorf & Köln: Eugen Diederichs.

Schmidtchen, G. (1965). *Die befragte Nation. Über den Einfluss der Meinungsforschung auf die Politik*. Frankfurt/Main: Fischer.

Schmitt, C. (1932/2002). *Der Begriff des Politischen. Text von 1932 mit einem Vorwort und drei Corollarien*. Berlin: Duncker & Humblot.

Schmölders, G. (1966). *Psychologie des Geldes*. Reinbek bei Hamburg: Rowohlt.

Sonnemann, U. (1963). Machen die Meinungsforscher die Meinung, die sie erforschen? In U. Sonnemann (Ed.), *Das Land der unbegrenzten Zumutbarkeiten. Deutsche Reflexionen* (pp. 162–175). Hamburg: Rowohlt.

Spicka, M. E. (2007). *Selling the economic miracle. economic reconstruction and politics in West Germany 1949–1957*. New York: Berghahn.

Von Hodenberg, C. (2006). *Konsens und Krise. Eine Geschichte der westdeutschen Medienöffentlichkeit 1945–1973*. Göttingen: Wallstein.

Chapter 6
$n = 1$: The Science and Art of the Single Case in Educational Research

David Bridges

This chapter is concerned with educational research that focuses on the single case (or, by extension, a very small number of cases) and with the question as to whether the single case can "properly" inform educational policy and practice beyond its own boundaries. The force of "properly" here is to frame the question as primarily an epistemological one, though the argument will extend beyond a narrow interpretation of the epistemological in exploring the way in which any kind of research might inform judgement.

In psychological and rhetorical terms there are plenty of examples of the power of the individual cases to shape decisions. Journalists are taught to ground big issues in the stories of individual human beings as a means of communicating their significance. Politicians know well the rhetorical power of the single exemplary story to persuade opinion. For example, in the UK the story of Laura Spence, a state school pupil with excellent qualifications, who was not accepted by an Oxford College symbolised Oxbridge's failure to address issues of privilege in their admissions policies.[1] Contemporary legislation is increasingly, it appears, shaped by individual tragedic events that seem to evidence faults in the system (cf. the terrible saga of the abuse of Victoria Cimbié, which became a driving force for far-reaching changes in the social care system in the UK – Laming, 2003; HM Government, 2003), and laws are even named after the individuals whose cases have come to symbolise the cause which the legislation addresses (cf. "Megan's Law" requiring sex offenders to notify their communities of their presence – passed by the State of New Jersey in 1994 NJSA 2C and ff. and as Federal Law in 1995 42 USC par 13701). That single cases *can* have enormous persuasive power seems to me to require little argument: the question is, *should* they have this power and under what conditions? Is it rational to change policy or practice in the face of evidence from a single case?

D. Bridges (✉)
University of Cambridge, Cambridge, UK

P. Smeyers, M. Depaepe (eds.), *Educational Research: The Ethics and Aesthetics of Statistics*, Educational Research 5, DOI 10.1007/978-90-481-9873-3_6,

6.1 Quantitative Research Methods and the Predilection for Large Numbers

There are clearly a lot of educational research questions which it would be nonsensical to answer on the basis of a sample of one. These would include any question which was concerned with the *scale* of an educational phenomenon. (How many children begin school without being able to write their names? What proportion of those entering Oxford and Cambridge have been educated in private schools? To what extent has the number of children being excluded from schools grown in the last 10 years?)

Similarly, questions about, e.g. the relationship between different phenomena in large populations (To what extent is underachievement in school linked to gender, ethnicity or social class? Which teaching approaches are related to high achievement in reading tests?) – i.e. correlational studies would certainly require a sample larger than $n = 1$.

Some of these questions can be answered on the basis of something close to a full set of data (even if these are not necessarily entirely reliable). Given, for example, that all those entering Oxford and Cambridge have to record the schools in which they have been educated, it is possible for n to equal all those in the relevant population ($n = N$). More commonly in research we have to be satisfied with obtaining data from a sample of the population in which we are interested, and we use these observations to make inferences about the larger population. Here we enter a battery of requirements on this process of sampling, the satisfaction of which is a key condition for the "scientific" acceptability of the research and hence its service to educational policy and practice.

These requirements would normally include measures designed to ensure that the population sampled "represents" or mirrors as nearly as possible the characteristics of the population as a whole, both in the diversity of that population and the proportional representation of that diversity. Among the factors that determine the size of the sample that is required in this sort of research are

- The extent to which the sample studied is carefully selected to match the significant characteristics of the wider population (which might permit a smaller sample of the kind employed, for example, in focus groups), or
- The extent to which the sample is randomly selected (which would require a larger sample in order to ensure representativeness and reduce the margin of error)
- What margin of error is acceptable in the results (is the researcher or the researcher's sponsor content with a rough and ready picture of what is happening or do they require something much more precise?). Broadly speaking, the margin of error decreases as the sample size increases, but there is always a trade-off between the margin of error accepted and the cost of the research.

This last principle is sometimes expressed (in terms which have particular significance for this chapter) of *the law of large numbers*, which is a theorem in probability

that describes the long-term stability of the mean of a random variable. Basically, the law dictates that the more observations you make of a particular variable, the more the sample mean will tend to approach and stay close to a particular value. If, for example, we toss a (fair) coin just two or three times, then it is quite possible that it will land as heads every time; if we continue for 10 or 20 tosses, then the proportion of heads will start to approximate to 50%; and it will get closer and closer to 50% the more times you toss the coin.

This law of large numbers tends to be extended (perhaps incorrectly, since it is essentially a principle applicable to a random variable with a finite expected value) as a general principle of research based on quantitative measure to indicate that, other things being equal, the larger the sample, the closer to a true picture, i.e. the more valid, will be the results. "The larger the sample size, the greater its accuracy… The sampling error – the difference between the sample and the population which are due for sampling – can be reduced by increasing sampling size" [though] "after a certain level, increases in accuracy tend to trail off as sample size increases" (Bryman & Cramer, 1990, p. 104). This principle applies in general to both correlational research (e.g. To what extent is entry to higher education a function of ethnicity, gender or social class?) and experimental research (How effective are mentoring schemes aimed at under-represented groups in raising entry to higher education in these groups?) In any research which is looking at the relationship between different variables, for example, if there are very few observations, then there are also few possible combinations of the values of the variables, and thus the probability of obtaining *by chance* a combination of the values indicative of a strong relation is relatively high. It is a feature of such research that when the enquiry has to be based on a sample rather than the entire relevant population, we are nevertheless interested in the confidence with which we can extrapolate from the sample to that whole population – and, other things being equal, a larger sample reduces the risk of error and or pure chance and increases the confidence with which one can draw inferences from the sample to the whole population.

It is against these sorts of expectations from quantitative research traditions that a sample size of a single case ($n = 1$), or even three or four cases can look faintly ridiculous. But is it?

6.2 The Function of the Single Case *Within* a Quantitative Research Tradition

It is worth noting, first of all, that the single case (and I shall use this also to include a small number of cases such that would not normally be thought to be statistically significant) can play an important and indeed devastating role even within the broadly quantitative research tradition. I have thus far accepted without comment the supposition which underpins the logic of a large part of empirical research in education as elsewhere, that one can draw inferences from the an appropriately constructed sample which has been the object of study to the wider population from which it is selected: in other words, one can generalise from the particular

cases. However, Popper, most notably among philosophers of science, has pointed out the logical gap that always lies between any number of singular or "particular" statements, such as accounts of the results of observations or experiments, and "universal" statements, such as hypotheses or theories or generalisations that are held to be true for all instances (Popper, 1959, 1963). Instead Popper turns the logic of scientific method on its head and argues that the task of empirical enquiry is to attempt to *refute* conjecture, i.e. such hypotheses or theories, and that the strength of such hypotheses lies in the measure of their ability to resist such attempts at their refutation (Popper, 1963). But what then becomes significant is that it takes not a thousand but just a single case of convincing counter-evidence to achieve this refutation. In the often quoted example, however many observations of white swans you make, these will not be enough to demonstrate conclusively that "all swans are white". On the other hand, a single observation of a black swan will suffice to refute the same proposition. This is not just a technical philosophical point, it is a common feature of scientific experience that it is the occasional aberrational result that signals that there is something not quite right with the existing theory.

Significantly, Stake situates case study in this Popperian tradition in the sense of offering the power of the negative example: "Case study is part of scientific method, but its purpose is not limited to the advance of science. Whereas single or a few cases are a poor representation of a population of cases and poor grounds for advancing grand generalization, a single case as a negative example can establish limits to grand generalization. For example, we lose confidence in the generalization that a child of separated parents is better off placed with the mother when we find a single instance of resultant injury. Case studies are of value in refining theory and suggesting complexities for further investigation, as well as helping to establish the limits of generalizability" (Stake, 2000, p. 445).

Harvey Goldstein argues, however, that the application of Popperian ideas to social science can be problematic since many (most?) hypotheses cannot be framed in terms of a few simple logically distinct states. Stake's example really needs to be rephrased along something like the following lines: "Children of separated parents, on average, are better off placed with the mother". The counter-example would be valid only if we could carry out a study that demonstrated that *on average* they were not better placed. Of course, it is not that simple, since one would need to specify the circumstances of the case, the particular features of the situation or context which might explain the departure from the norm. So Popper's notions are often fine in the natural sciences where it may be reasonable to assume universally applicable relationships so that universal statements or hypotheses make sense. In the social sciences these ideas have less utility. Of course, finding, for example, that some families or schools appear to behave differently from others is interesting and leads to further questions and data collection, and the key role for case studies is to try and understand why this might be happening. However, in the social sciences a single case study will rarely refute a generalisation because they are rarely offered as universally applicable laws.[2]

Further, as Elliott and Lukes argue, "case-focussed reasoning" is an intrinsic part of the development and application of quantitative studies:

> . . . the construction of scenarios and cases is essential in the design of quantitative experiments and research projects both in the form of discursive assumptions and the actual case-based pilot studies. But it is no more absent in the policy implementation of the resulting data. Cases like stories have to be told before a policy based on quantitative "evidence" can be translated into prescriptive documentation, but they are also essential in the political justification of policy. (Elliott & Lukes, 2009, pp. 98–99)

In other words, when we are designing a quantitative experiment we have to construct a story about what might (or might not) happen in a particular situation given certain limited different conditions; and when we report the research we feel greater confidence in telling a story about what did happen in those situations given certain (limited) conditions and what would be likely to happen in other similar situations (this being the inferential turn). The trouble with such studies is they can only ever deal with a very limited set of descriptors of the situations abstracted from what we know to be a very much more complex set of realities – which is precisely what draws some people towards not just case reasoning but case study.

6.3 Different Forms of the Single Case

On the whole, when we talk of research in education focussed on a single case, we are not referring to quantitative research. There is, of course, a long tradition of case study in psychology especially in child psychology. For many teachers from the 1950s and 1960s a "child study" was a key component of their teacher training. The psycho-analytic tradition was almost entirely constructed around a mixture of theory and case study (and the cases would not have withstood much scrutiny against the criterion of their representativeness either). Subsequently in educational circles, case study became especially associated with ethnographic research in an anthropological tradition. In the last two decades, encouraged in particular by its espousal by feminist researchers, case study has often taken the form of biography and autobiography, narratives of people's educational lives and experiences presented either as the outcome of research or as sources to inform the writing of a research account (see Griffiths & Mcleod, 2009). More widely, "cases" have played a central role in the development of medical understanding and, whatever might be the general recommendations for medical practice defined, for example, by the National Institute for Clinical Excellence, the examination of the individual case and case history remains central to medical decisions. English common law or, as it is sometimes referred to, "case" law is indeed constructed around decisions about individual cases that have been determined in the courts, the historical collection of which constitute the basis for contemporary legal argument, reference and decision.

In the context of a book which is concerned with historical as well as philosophical approaches to inquiry, it is relevant to note too Stenhouse's argument that a great deal of history is best conceived of as case studies in the sense of studies of particular defined phenomena bounded by time and place (Stenhouse, 1977). He gave as examples Girouard's *The Victorian Country House*, *The working life of women in the seventeenth century*, *English wayfaring life in the middle ages* and, to focus

rather more on education, *English grammar schools in the reign of Queen Elizabeth*. Of course, the plethora of historical biographical works would also constitute cases of a slightly different kind. But to take a more recent example Carlo Ginzburg's *The cheese and the worms: the cosmos of a 16th century miller*[3] explicitly positions itself as a qualitative and singular study and is contrasted by its author with the practice in which "virtual teams of scholars have embarked on vast projects in the quantitative history of ideas" (Ginzburg, 1992, p. xx). It is perhaps worth recording Ginzburg's rationale for this approach, which presages some of the arguments which will follow:

> Menochio [the miller in question] articulated the language that history put at his disposal. Thus it becomes possible to trace in his disclosures in a particularly distinct, almost exaggerated form, a series of convergent elements… A few soundings confirm the existence of traits reduceable to a common peasant culture. In conclusion, even a limited case (and Menocchio certainly is this) can be representative: in a negative sense, because it helps to explain what should be understood, in a given situation, as being "in the statistical majority"; or, positively, because it permits us to define the latent possibilities of something (popular culture) otherwise known to us only through fragmentary and distorted documents, almost all of which originate in "the archives of repression". (Ginzburg, 1976/1992, p. xxi)

In the context of educational research "cases" may be constructed on the basis of a variety of forms of enquiry and hence of data including descriptive writing by an observer or a participant (or a participant observer), interview data (typically derived from unstructured or semi-structured interviews rather than questionnaires), observational logs or diaries, documentary sources (including correspondence, official documents, newspaper reports, minutes of meetings); photographic images. In this sense "case study" is not a research method: rather it is (i) a way of focussing or putting boundaries around a piece of research (cases are often referred to as "bounded systems" – see Adelman, Kemmis, & Jenkins, 1976, p. 141) and (ii) a form of representation of the research, a choice in favour of a particular form of representation of observation, evidence and experience.

The outcomes of these various forms of enquiry have at least some of the following features in common:

- They focus on a single case or on a small number of cases
- Such cases are "an instance drawn from a class" (Adelman et al., 1976, p. 141) or, as it is sometimes put, "of the type but not necessarily typical", so in orthodox sampling terms they can only claim a very modest degree of "representativeness"
- They deal holistically with educational phenomena and involve "a myriad of not highly isolated variables" (Stake, 1980, p. 71)
- They provide a detailed or "thick" picture of a situation or series of events (and consequently having narrative form). Of course, such description does not exclude quantitative information where relevant. Stenhouse (1980, p. 4) emphasised in his 1979 Presidential address to the British Educational Research Association "an acute need for attention to be paid to quantitative aspects of case study"

- They are typically written in ordinary, "naturalistic" rather than technical language
- They do not in themselves present results or conclusions. "Themes and hypotheses may be important, but they remain subordinate to the understanding of the case" (Stake, 1980, p. 71).

The first thing to note about this picture of the single case is that it is qualitatively different (as well as different in scale) from what might be a single case in, for example, research based on a survey or randomised controlled experiment. In either of these approaches the actual information contained in a single case would be tiny by comparison with, e.g., a life history or an ethnographic case study. Of course, if the 1 in $n = 1$ is a response to a questionnaire, then the single item is not a lot of use. "To know particulars fleetingly", wrote Stake, "is to know next to nothing. What becomes useful understanding is *a full and thorough knowledge of the particular*, recognising it in new and foreign contexts" (Stake, 1980, p. 69, my italics). The single case may even in this tradition be a book length treatment of a single school – as in Stephen Ball's influential study of Beachside Comprehensive (Ball, 1981).

Nevertheless, it would be difficult to find anything much further removed from the requirements described earlier for the credibility of correlational or experimental research than this account of case study. The two really do not admit of sensible direct comparison though a lot of the literature on case study strives to show its consistency, or at least comparability, with natural science methods and methodology (cf. Kemmis, 1980). As MacDonald and Walker argued: "We might say that case study is that form of research where $n = 1$, only that would be misleading, because the case-study method lies outside the discourse of quantitative experimentalism that has dominated Anglo-American educational research" (MacDonald & Walker, 1975/1977, p. 2).

Is it any surprise that the proponents of the related discourse of "evidence based" policy and practice and those excited by "systematic reviews" of educational research tend to leave it aside as "unscientific" and irrelevant? But are they right to do so – or are they confused about the different ways in which different genres of research – even those based on a single case – may inform (again, legitimately) educational policy and practice? In the remainder of this chapter I want to explore possible approaches to the question "How may a single or small number of case studies inform educational policy and/or practice?"

6.4 Generalising from the Single Case?

> "To generalise" is to be an idiot. To particularise is the alone distinction of merit. General knowledge are those that idiots possess. (William Blake in *Annotations to Sir Joshua Reynold's Discourses* Reynolds, 1992, p. 36)

If the case is "an instance drawn from a class" (Adelman et al., 1976, p. 141) – for example, a comprehensive school, a disaffected teacher, a multi-ethnic group

of children – and is chosen because it has at least a *prima facie* representativeness, then it is tempting to suppose that one might generalise about the class from which the instance is drawn. However, apart from any of the standard critiques of such attempts from a statistical point of view, Adelman et al. have their own account of why it is not that simple.

> During the conduct of the study the description of the case will increasingly emphasise it uniqueness. Loosely speaking this means that the study will reveal increasingly some case bound features of the instance vis-à-vis the class. More strictly speaking, the study will transcend the principle of selection (i.e. selecting the instance as representative of a given class) and become a study of a unique case. It does not follow that the researcher, by virtue of this shift, abandons the hope of generalisation, though he may do so. Rather the basis of generalisation itself has changed. (Adelman et al., 1976, p. 141)

Quite what might constitute the alternative basis for generalisation here, I shall explore shortly. For the moment, however, it is worth noting a different message which might also be drawn from this account of what happens when you start to look at individual cases in a class, i.e. that they serve to remind the research or policy-maker of the unique mix of conditions and circumstances which have significance in the individual case – and thus of the dangers of seeking to apply *any* generalised prescription as to how people should act under these particular conditions. Even quantitative researchers with any sophistication share this scepticism. David Byrne, for example, explains in *Interpreting Quantitative Data* that:

> Statistical inference, although often reified beyond its limited but important value in relating the properties of a sample to the contextual and local universe from which it is drawn, is only a component in the far more important and scientifically significant issue of generalisation… The real world of the social/natural is composed of evolutionary and interacting systems. The epistemological consequence of this is that knowledge is inherently local. We cannot appeal to universal laws applicable everywhere … as the basis for generalisation. (Byrne, 2002, pp. 74–75)

Writing in the context of the evaluation of the Humanities Curriculum Project, Barry MacDonald came to a similar conclusion: "No two schools are sufficiently alike in their circumstances for prescriptions of curricular action to be able adequately to supplant the judgement of the people in them" (Macdonald, 1971 cited in Simons 1971, p. 118). Perhaps it is, then, the very particularity of educational practice which needs to be understood. It was precisely for this reason that Lawrence Stenhouse presented a picture of the curriculum, not as a prescription to be followed unquestioningly, but as a hypothesis which needed to be tested against the particular conditions which applied in a unique setting. "Theory is useless to the practitioner unless he can subject it to situational verification, that is, test it in the situation in which he currently finds himself" (Stenhouse, 1977, p. 3). This is why he emphasised the need for teachers to be researchers, which was more a call for the adopting of a questioning, enquiring approach to generalisations about how and what to teach than a requirement on teachers to adopt a particular set of research practices. There were, in any case, always ethical constraints on the teacher's assumption of a research role: "Given the ethics of the situation, the teacher

can test theory only by taking educational action, that is action which can be justified in educational rather than experimental terms so that he could conscientiously take it without experimental intent" (ibid).

Cronbach, similarly, acknowledged the complex particularity of social (including educational) settings and drew an important conclusion for the direction of social science research: its primary aim should become "interpretation in context" not "generalisation" (Cronbach, 1975, p. 123).

St. Clair (2005) has taken the argument a stage further, invoking the notion of "superunknowns" as an obstacle to *any* possibility of transferring empirically based knowledge between educational settings:

> In any social situation there are an infinite number of factors that could influence the interaction between humans (such as weather, language, height, breakfast food, and so on). Any infinite group of factors, even if there are millions taken into account, leaves an infinite number unaccounted for – these are the superunnowns. . ..[Further] there is no truly effective strategy to deal with superunknowns. This is because there are an infinite number of them, and not only do we not know what they are, we are also unaware exactly how important they are. . . [Consequently] If similarity is accepted as the basis for induction, and the importance of superunknowns is recognised, then the conclusion has to be that the justification for induction in educational research is weak – there is no systematic, logically coherent justification for the assumption that what we learn from setting A applies to setting b. (St. Clair, 2005, p. 446)

The problem about St. Clair's argument is that it appears to remove the possibility not only of transferring understanding from one situation to any other but even of understanding a particular case, since here, too, there will presumably be an infinite number of unknowns as well as some things which are known. Or, perhaps, it is simply that one could never understand the particular case fully, which while not entirely a trivial qualification, would be more than anyone might expect to do any way.

Nevertheless, St.Clair's and the other cautions indicated here are ones which I think need to be taken very seriously, not least because they have important implications with respect to the level in any educational system at which practical decisions about teaching need to be taken. They argue that, though the broad values and principles which a system might be expected to realise might appropriately be determined through some democratic process at a higher level, decisions about how in the end to realise these values in practice have to be made on the basis of a detailed understanding of the particular context in which they are to operate.

This caution has not, however, persuaded all advocates of case study to abandon the language of "generalisation" even if, as Mejia (2009, p. 2) has pointed out, this is a very different kind of generalisation from that associated with randomised controlled trials. "Various authors have rushed to assign it an adjective so that it is made clear that there is a distinction": *analytic* (Yin, 1984), *retrospective* (Stenhouse, 1980) or *naturalistic* (Stake & Trumbull, 1982), among others. "Other authors", continues Mejia, "have preferred to drop the term 'generalisation' and replace it with another one: for instance *relatability* (Bassey, 1981), or *transferability* (Lincoln & Guba, 1985)" (Mejia, 2009, p. 2, my italics) – though Bassey did, of course, later (2001) add *fuzzy* generalisations to the repertoire.

How, then are we to respond to the question about how a single case can legitimately inform educational practice or policy. In the next sections I want to explore three approaches: (i) the single case as a source of conjecture and refutation; (ii) relating the particular to the particular; and (iii) the single case as an extension of experience and as a contribution to practical wisdom.

6.5 The Single Case as a Source of Conjecture and Refutation

In a way this is perhaps the easiest case to advance because it is the most modest in its claims. Popper (1963) has offered a framing of science as a matter of conjecture (i.e. theory building or hypothesis formulating) and refutation (the negation of these through discussion, critique and empirical testing) – and even if this has drawn its critics it has survived as at least a partial picture of what goes on in scientific enquiry. I have already discussed the role of the single case in the process of refutation. The single case (and *a fortiori* the single case richly described) is enough to make it clear that what might have been thought to be the case is not necessarily the case, need not be the case, might not be the case if particular conditions obtained.

It is the refutation side of Popper's twin concepts that tends to get more attention, but conjecture is at least as important, and it would be part of my argument that a well-described or narrated single case, and perhaps even more the sort of accumulation of cases that Stenhouse favours, provides a fertile source for such conjecture, moreover one in which the active imagination is "grounded" in a graphic description of reality or in the multi-perspectival takes on reality that are characteristic of the modern ethnographic case. In the Cambridge Accountability Project, for example (Elliott et al., 1981), five researchers spent 18 months studying six secondary schools and their relations with their parents and communities. Through the process and on the basis of the evidence generated the researchers kept asking themselves, "What general principles about school accountability can we offer of these studies?" (the conjecture) and "Can we find any evidence in the studies to challenge or gainsay these principles?" (the attempted refutation). In other terms, familiar to anyone engaged in case research, we were generating grounded theory (see, for example, Glaser & Strauss, 1967). In such an example, the small scale of the sample is off-set to some degree by the detailed examination of the cases in question.

In this case the researchers were not technically able to answer the question "How typical is this experience?" though at no stage in the subsequent conferences and events for headteachers and local authority officers was this seen as an issue; they could readily identify with the picture that was provided. In a sense the claim that I am more interested in making for this sort of research relates to the interest, the freshness, the fecundity of the insights and ideas that it can generate, and especially if these insights are sufficiently well grounded and well described that they command the response "I can see that that is how things were" and perhaps (but not essential to this part of my argument) "... and I had not thought of it like this before, but I can see that that is what is also happening in my own school" – which brings me to my second argument.

6.6 Relating the Particular to the Particular

> In its most significant form, generalisation about the case promotes generalisation from case to case. . . as in art, which teaches by example rather than precept. . . (Adelman Kemmis, & Jenkins, 1980, p. 142)

When Kennedy asked teachers what persuaded them about the utility of a particular piece of research, she began to build up a picture that applied, in fact, independently of the genre of research which they encountered: "teachers *forged analogies* between the studies they read and their own situations or practices" (Kennedy, 1999, p. 537, my italics). "Forging analogies" is one way of expressing the processes which are involved in coming to understand one situation through an understanding of another. It is this process that Stake refers to a "naturalistic generalisation":

> Often the situation . . . is one in which there is need for generalisation to a similar case rather than generalisation to a population of cases. Then the demands for typicality and representativeness yield to need for assurance that the target case is properly described. As readers recognise essential similarities to cases of interest to them, they establish the basis for naturalistic generalisation. (Stake, 1980, p. 71)

However, I think it is misleading to call this process "generalisation", because there is no generalisation and none is required. It is not a matter of a logical sequence which starts with a particular instance, moves inductively to a general theory about the class to which the instance belongs and then deduces from the general theory what might be the case in another particular instance that fall within the class. It makes no general claims. It simply affirms that this single instance A is sufficiently like another instance B that I can gain some understanding of B through my understanding of A. "The general idea is that rich description of a single case or of a reduced number of cases, if of good quality, will help other practitioners see their own cases reflected and judge for themselves what is applicable in their own practice" (Mejia, 2009, pp. 2–3).

Such – let us call it – "application" of the single case is especially significant in the context of the relationship between research and practitioners, who are on the whole not too bothered about whole populations or systems of children, classrooms or schools (and will probably regard their own situation as unique any way) but only in their own situation. They are not so interested in general theory but in principles and practices which they can identify with and which "fit" with their experience in their own classroom.

I strongly suspect that, notwithstanding their public subscription to the cult of large numbers, policy-makers do not respond very differently to practitioners in this respect, though it would need a different kind of study to establish this. Elliott and Lukes' account of "case reasoning" fits with a model in which, even faced with, for example, the quantified results from a randomised controlled experiment, a policy-maker still needs to envisage the forces at work in terms of a picture of what might be happening in a situation with which he or she is familiar or, perhaps, one which they have experienced vicariously or one they can construct in their imagination.

6.7 The Case as Offering a Vicarious Experience

> Experimental research "guarantees" the veracity of its generalisations by reference to formal theories and hands them intact to the reader; case study research offers a surrogate experience and invites the reader to underwrite the account, by appealing to his tacit knowledge of human situations. The truths contained in a successful case study report, like those in literature, are "guaranteed" by "the shock of recognition". (Adelman et al., 1980, p. 143)

Adelman et al. were here echoing, in particular, the views of Stake (who was a participant in the conference from which their paper emerged). There are four elements to this argument, which it may be helpful to distinguish.

First, there is a claim about what case studies can offer in the form of vicarious experience:

> One of the more effective means of adding to understanding – for all readers – will be by approximating through the words and illustration of our reports the natural experience attained in ordinary personal involvements. (Stake, 1980, p. 65)

> Certain descriptions and assertions are assimilated by readers into memory. When the researcher's narrative provides opportunity for vicarious experience, readers extend their memories of happenings. Naturalistic, ethnographic case materials, to some extent, parallel actual experience, feeding into the most fundamental processes of awareness and understanding. (Stake, 1994, p. 240)

Second, there is a view about how such vicarious experience engages with human understanding: extending their memories of happenings, entering their tacit understanding. In a sense, then, the argument rests on a theory about how practitioners and policy-makers learn (come to extend their understanding) rather more than an epistemological theory about what authority can be claimed for what they learn. Kemmis does indeed make a direct connection with such learning theory when he writes:

> In Piaget's terminology, the reader must be able to assimilate the situation-as-reported to his present experience and accommodate the present forms of experience to the new forms of possible experience offered by the report. It is in this sense that case study often works by Verstehen (empathetic understanding) feeding imagination and experience as much as propositional knowledge in discourse. (Kemmis, 1980, p. 128)

By extension Adelman et al. suggest that there is a particular pedagogy attached to the use of case study "which teaches by example rather than precept" (Adelman et al., 1980, p. 142).

Third, and perhaps more controversially, contributors like Stenhouse and Stake have tried to address the claims of natural sciences to the predictive power of their hypotheses, though their language shifts from "prediction" to "expectation" and "anticipation":

> The responsive diagnostic *predictions* which interplay with action may be called anticipations. Surprise in a phenomenology of understanding is equivalent to falsification in scientific theory. (Stenhouse, 1977, p. 4, my italics)

> I am concerned here with how we can improve the professional common sense understanding of the educator which supports *anticipations* through diagnostics – in short with the improvement of judgement. (Stenhouse, 1977, p. 5, my italics)

> Naturalistic generalisations develop within a person as a result of experience. They form from the tacit knowledge of how things are, why they are, how people feel about them, and how these things are likely to be later or in other places with which this person is familiar. They seldom take the form of predictions but lead regularly to *expectations*. They guide action. (Stake, 1980, p. 69, my italics)

In short, then what we have so far is a picture of the way in which the single case serves to extend, enrich and inform personal experience and understanding, much of this at a tacit level. This is a very grounded, concrete understanding packed with detailed description and personal accounts of what it is like to be there "where the action is". It is in this sense perhaps the kind of understanding that Aristotle identifies as *phronesis* or "practical wisdom". Being thus grounded it positions the actor to anticipate (if not confidently to predict) what might happen if. . ., what to look out for, even if this always comes with the ultimate requirement to test out such anticipation in the particular circumstances of their own setting. Can any form of educational research really offer more than this to the practitioner, or even the policy-maker?

Fourth, what is clearly acknowledged here (cf. the reference to the Piagetian notion of assimilation above) is that neither practitioner nor policy-maker is an educationally empty vessel. Both come to any research report with a huge load of mental and emotional baggage drawn from past experience (personal and professional), past reading or encounters with research and past reflection. Further, as Stake has reminded us, much of this is buried in their tacit understanding of the world. Any contribution that research makes to their understanding will only be a small part of the total, and its significance will be hugely influenced by what is already there and by their current felt needs. Thus *any* research results (randomised controlled experiments included) will be interpreted by the reader in ways which reflect their pre-existing understanding and preoccupations.

> Even in factual terms every research report is open to multiple, more or less reasonable, interpretations and usually is interpreted in different ways by different people. It is not necessary to go to Derridean extremes about the "dissemination" of meaning to recognise that there is a sense in which readers construct the meaning of any research report and may do so in diverse ways. (Hammersley, 2002, p. 46)

This is another way of reminding ourselves that receiving research (of any kind) is a kind of learning and is subject to all that we know about the individual's construction of that learning. We need perhaps not so much a theory of generalisability as a theory of learning – which might be informed by Elliott and Lukes' suggestion that "The policy-maker's grasp of actualities is not so much enhanced by the straightforward application of general principles as by judgements that are tutored by the comparative study of cases" (Elliott & Lukes, 2009, p. 89; Bridges & Watts, 2009). Hence – and importantly if we care about practitioners and policy-makers taking our research seriously – "If the readers of our reports are the persons who populate our houses, schools, governments and industries; and if we are to help them understand social problems and social programs, we must perceive and communicate in a way that accommodates to their present understandings" (Stake, 1980, pp. 64–65).

6.8 Case Study: Science or Art?

This collection of papers comes under the title of *The Ethics and Aesthetics of Statistics*. I am not going to say here anything about the ethics of research in which $n = 1$, though, of course, such research carries particular responsibilities since, case study both makes particularly heavy demands on participants (no mere "just 10 min to complete a questionnaire" here) and because the detailed description which is typically a feature of a case study makes it difficult to guarantee anonymity, if that is what is required. But what of the aesthetics of the single case? Is case study more akin to science or art?

Helen Simons edited a seminal collection of papers, from which I have drawn extensively, that appeared under the no doubt consciously provocative title *Towards and Science of the Singular* (Simons, 1980), and one of its contributors, Stephen Kemmis was unequivocally committed to "the proposition, so heavily validated by lay and professional usage that it hardly needs defending, that case study work is science" (Kemmis, 1980, p. 100 and pages following for his justification). He argues that

> In [three] crucial ways ... case study work is like all forms of science: it is an empirical process of truth seeking; it is a social, cultural and cognitive process; and it resolves in its practice the double problem of justifies true belief – the justification of belief and the belief in justification. (Kemmis, 1980, p. 106)

But even if case study satisfies these three conditions (and I wonder how many contemporary researchers would concede – even if they should – that they are "truth seeking"), is this sufficient to demonstrate that case study is a "science"? More importantly, perhaps, does it matter? Why should one legitimate genre of research need to conform to the requirements of another – unless one is conceding that only the scientific, out of all the intellectual resources of the academy, can claim legitimacy? Kemmis seems to be in danger of falling victim to the very form of "scientism" that in the same paper he seeks to attack.

It seems to me better to acknowledge that case study has its own strengths rooted in the arts and humanities rather than in the natural sciences – that this carries with it requirements for thoroughness, accuracy, attention to detail, but also imagination, social sensitivity, contextual understanding, and quality requirements in the structure and writing of the report – in other words its own discipline and rigour which provide both the warrant for its credibility and the pedagogical or rhetorical power to affect the reader's understanding.

To free case study research from the particular requirements of the natural sciences is not by any means to suggest that anything goes. Simons argues that "The case study must not only be authentic and detailed ..., it must also be rigorously accurate and impartial" (Simons, 1971, p. 122). For Stenhouse these principles were honoured in case study by his requirement that any such study should be linked to a set of case data, i.e. all the materials assembled by the case worker. These, he acknowledged would be too bulky for repeated handling and too sensitive for immediate release, but they were (on an analogy which he drew with historical research) the primary sources against which, in principle at least the case study

could be checked and challenged. More accessible would be a case record "a theoretically parsimonious condensation of the material of the case data produced by selective editing without explicit comment" (Stenhouse, 1977, p. 20). The purpose of the case record, as Elliott and Lukes explain, is to enable others "to critique the case study and to judge the extent to which it could be verified as a reasonable interpretation of the available evidence" (Elliott & Lukes, 2009, p. 90).

But we can go further than this in locating case study within an artistic or, since the main but not exclusive tools of the case researcher are words, a literary tradition. In so far as case study pays attention to rich description and the evocation of place and time, to personal experiencing, to reflecting complexity (including moral complexity) rather than abstracting a few variables from complex situations, to "portrayal" rather than (or as well as) analysis, to narrative form, to ordinary rather than technical language, to a good read and to the essential indeterminacy of the text – does it not locate itself rather more in the stable of literature or at least the humanities rather than science.

> The kind of case-studies which we believe education needs have characteristics which call for a fusion of the styles of the artist and the scientist. When Freud said, "It still strikes me myself as strange that the case histories I write should read like short stories and that, one might say, they lack the serious stamp of science" he caught the unease of the researcher who, disdaining the "safety of numbers", discovers that his data is most effectively exposed in a mode which is generically associated with the artist. (MacDonald & Walker, 1975/1977, p. 3; Freud, 1955, p. 160)

And not only should this not matter: Why do we find ourselves constantly having to apologise because not everything in the academy meets the requirements of science (or what these are supposed to be)? Might not the difference, after all, be a source of excitement? Science may offer the *general* (though there are many reservations which need to be entered into even this claim – see, e.g. Smeyers, 2009) but the arts offer the possibility of the *universal*:

> To see a world in a grain of sand
> And heaven in a wild flower
> Hold infinity in the palm of your hand
> And eternity in an hour.
> (William Blake: Auguries of Innocence in Nicholson & Lee, 1916)

In a sense the greater the particularity of the case study and the more fully humanity is displayed in all its complexity and ambiguity – and there is a direct analogy here with a great novel or drama – the more opportunity there is for the reader to discern not just what relates to their own condition or to a sub-class of subjects like them but to humanity as a whole.

> Case study is the way of the artist, who achieves greatness when, through the portrayal of a single instance locked in time and circumstance, he communicates enduring truths about the human condition. (MacDonald & Walker, 1975/1977, p. 3)

In this process, too, the case study does something equally important, which is to restore that humanity to its place at the centre of educational practice and policy from which the crude abstractions of scientific measurement constantly threaten its removal.

6.9 Historical/Biographical Note

In a volume which celebrates the contributions of history as well as philosophy to educational enquiry, perhaps I may be allowed a short historical note on the sources which are especially prominent in this chapter.

This chapter is in a sense a homage to a community of people who contributed substantially to the theory and practice of case study in a very creative period which began in the early 1970s, nearly all of whom had an affiliation of some kind with the Centre for Applied Research in Education at the University of East Anglia. Lawrence Stenhouse was its founder and Barry MacDonald, John Elliott and Rob Walker, among the founding or early members, were all subsequently its directors. Clem Adelman, David Jenkins, Helen Simons, Stephen Kemmis, Harry Torrance and Malcom Parlett all had periods of employment with or attachment to CARE, and Bob Stake and Lou Smith were among the regular international visitors and collaborators.

The group found common causes (as well as strongly voiced differences of opinion) at the conferences which were convened from time to time in Cambridge and which from the beginning were reported in papers published by the *Cambridge Journal of Education* in 1976, 1981, 1986, 1995 and, most recently, 2007. This last conference issued in the development of an archive of material relating to the use of case study in educational evaluation developed by this network of scholars which can be found at www.ensemble.ac.uk/projects.edeval (accessed 7 February 2010).

I was editor of the *Cambridge Journal of Education* between its foundation in 1971 and 1982 and served my apprenticeship in case study research under John Elliott and later Barry MacDonald even before joining UEA in 1990, and I attended all but the first of the Cambridge conferences, hosting the most recent one at St Edmund's College.

It is *de rigeur* in some academic circles to cite only the most contemporary references. I hope to have illustrated the rich insights that are still to be mined from what are by now almost historical sources.

Notes

1. The Laura Spence affair in 2000 enraged Gordon Brown (who was then UK Chancellor of the Exchequer). The fact that Laura, of Monkseaton Community High School, was rejected by Oxford's Magdalen College despite her straight-A predictions, seemed so deeply unfair to Mr Brown that he resolved to make Oxbridge mend its elitist ways and admit more state school pupils. A White Paper presented to Parliament in January 2003 accordingly proposed the "rapid expansion" of measures intended to "widen access", and the government has kept up the pressure on both Oxford and Cambridge ever since (see Boss, 2006 in the *Spectator*).
2. I am grateful to Harvey Goldstein for the argument in this paragraph which he offered in response to an earlier draft of this chapter.
3. I am very grateful to Lynn Fendler for drawing this work to my attention.

References

Adelman, C., Kemmis, S., & Jenkins, D. (1976). Re-thinking case study: Notes from the second Cambridge conference. *Cambridge Journal of Education, 6*, 139–150. (Also in Simons H. (Ed.). (1980). *Towards a science of the singular* (CARE Occasional Publication no.10, pp. 45–61). Norwich: CARE/University of East Anglia).

Ball, S. (1981). *Beachside comprehensive: A case study of secondary schooling*. Cambridge: Cambridge University Press.

Bassey, M. (1981). Pedagogic research: On the relative merits of search for generalisation and study of single events. *Oxford Review of Education, 7*, 73–93.

Bassey, M. (2001). A solution to the problem of generalisation in educational research: Fuzzy prediction. *Oxford Review of Education, 2*, 5–23.

Boss, C. (2006, December 2). Why so many state school pupils drop out of Oxbridge: The political gameplay behind high number of university dropouts in UK. *Spectator*.

Bridges, D., & Watts, M. (2009). Educational research and policy: Epistemological considerations. In D. Bridges, P. Smeyers, & R. D. Smith (Eds.), *Evidence based education policy: What evidence? What basis? Whose policy*? (pp. 36–57). Oxford: Wiley Blackwell. (Previously published as Supplementary Issue of *Journal of Philosophy of Education, 42*, 2008, 41–62).

Bryman, A., & Cramer, D. (1990). *Quantitative data analysis for social scientists*. London: Routledge.

Byrne, D. (2002). *Interpreting quantitative data.* London: Sage.

Cronbach, L. J. (1975). The two disciplines of scientific psychology. *American Psychologist, 30*, 116–127.

Elliott, J., Bridges, D., Ebbutt, D., Gibson, R., & Nias, J. (1981). *School accountability*: London: Grant MacIntyre.

Elliott, J. & Lukes, D. (2009). Epistemology as ethics in research and policy: The use of case studies. In D. Bridges, P. Smeyers, & R. D. Smith (Eds.), *Evidence based education policy: What evidence? What basis? Whose policy*? (pp. 82–114). Oxford: Wiley Blackwell. (Previously published as Supplementary Issue of *Journal of Philosophy of Education, 42*, 2008, 87–120).

Freud, S. (1955). *Standard Edition* (Vol. 2). London: Hogarth Press and the Institute of Psychoanalysis.

Ginzburg, C. (1992). *The cheese and the worms: The cosmos of a sixteenth century miller* (J. Tedeschi & A. Tedeschi, Trans.). Baltimore: John Hopkins University Press.

Glaser, B. G., & Strauss, A. L. (1967). *The discovery of grounded theory: Strategies for qualitative research*. New York: Aldine.

Griffiths, M., & Mcleod, G. (2009). Personal narratives and policy: Never the "twain?". In D. Bridges, P. Smeyers, & R. D. Smith (Eds.), *Evidence based education policy: What evidence? What basis? Whose policy*? (pp. 115–137). Oxford: Wiley Blackwell. (Previously published as Supplementary Issue of *Journal of Philosophy of Education 42*, 2008, 121–144).

Hamilton, D., Jenkins, D., King, C., McDonald, B., & Parlett, M. (1977). *Beyond the numbers game: A reader in educational evaluation*. London: Macmillan.

Hammersley, M. (2002). *Educational research, policy making and practice*. London: Paul Chapman.

HM Government CM 5861. (2003). *Keeping children safe: The Government's response to the Victoria Cimbie Inquiry Report and the joint Chief Inspectors' report "Safeguarding children"*. Norwich: HMSO.

Kemmis, S. (1980). The imagination of the case and the invention of the study. In H. Simons (Ed.), *Towards a science of the singular* (pp. 97–140, CARE Occasional Publication no.10). Norwich: CARE/University of East Anglia.

Kennedy, M. M. (1999). A test of some common contentions about educational research. *American Educational Research Journal, 35*, 511–541.

Laming, L. (2003). *Report of the Victoria Cimbié inquiry*. Norwich: HMSO.
Lincoln, Y., & Guba, E. (1985). *Naturalistic enquiry*. London: Sage.
MacDonald, B. (1971). *Evaluation of the Humanities Project: A holistic approach*. Paper presented to the Annual Meeting of the American Educational Research Association, New York, 1971.
MacDonald, B., & Walker, R. (1975). Case study and the social philosophy of educational research. *Cambridge Journal of Education, 5,* 2–20. (Also in Hamilton, D., Jenkins, D., King, C., McDonald, B., & Parlett, M. (1977). *Beyond the numbers game: A reader in educational evaluation*. London: Macmillan).
Mejia, A. (2009). *On why there is generalisation in research on singe cases.* Paper presented to Annual Conference of the Philosophy of Education Society of Great Britain, New College, Oxford, 3rd to 5th April 2009.
Nicholson, D. H. S., & Lee, A. H. E. (Eds.). (1916). *The Oxford book of English mystical verse*. Oxford: Oxford University Press.
Popper, K. R. (1959). *The logic of scientific discovery*. London: Hutchinson.
Popper, K. R. (1963). *Conjectures and refutations*. London: Routledge and Kegan Paul.
Reynold, J., Sir (1992). In P. Rogers (Ed.), *Discourses: Including "Annotations to Reynold's Discourses" by William Blake*. Harmondsworth: Penguin.
Simons, H. (1971). Innovation and the case study. *Cambridge Journal of Education, 1*, 118–124.
Simons, H. (Ed.). (1980). *Towards a science of the singular* (CARE Occasional Publication no.10). Norwich: CARE/University of East Anglia.
Smeyers, P. (2009). On the epistemological basis of large population studies and their educational use. In D. Bridges, P. Smeyers, & R. D. Smith (Eds.), *Evidence based education policy: What evidence? What basis? Whose policy*? (pp. 58–81). Oxford: Wiley Blackwell. (Previously published as Supplementary Issue of *Journal of Philosophy of Education, 42*, 2008, 63–86).
St. Clair, R. (2005). Similarity and superunknowns: An essay on the challenges of educational research. *Harvard Educational Review, 75*, 435–453.
Stake, R. (2000). Case studies. In N. Denzin & Y. Lincoln (Eds.), *Handbook of qualitative research* (2nd ed., pp. 435–454). Thousand Oaks, CA: Sage.
Stake, R. E. (1980). The case study method in social enquiry. In H. Simons (Ed.), *Towards a science of the singular* (CARE Occasional Publication no.10, pp. 62–96). Norwich: CARE/University of East Anglia.
Stake, R. E. (1994). Case studies. In N. Denzin & Y. Lincoln (Eds.), *Handbook of qualitative research* (2nd ed., pp. 435–454). Newbury Park, CA: Sage.
Stake, R. E., & Trumbull, D. (1982). Naturalistic generalisations. *Review Journal of Philosophy and Social Science, 7*, 1–12.
Stenhouse, L. (1977). *Case study as a basis for research in a theoretical contemporary history of education.* Mimeo: Norwich CARE/University of East Anglia. (Archive and as Case study and case records: Towards a contemporary history of education. *British Educational Research Journal, 4*(2), 21–39).
Stenhouse, L. (1980). The study of samples and the study of cases. *British Educational Research Journal, 6*(1), 1–6.
Yin, R. (1994). *Case study research: Design and methods*. London: Sage.

Chapter 7
To Frame the Unframable: Quantifying Irregular Migrants' Presence

Elias Hemelsoet

7.1 Introduction

For various reasons, irregular migration has become a more frequent phenomenon during the last decades. Without going deeper into the globalising context of growing and changing migration tendencies in different parts of the world, it should be remarked that this topic is gaining attention. Until recently, there was only a limited amount of scientific evidence on irregular migration (with a focus on clandestine activities such as human trafficking and smuggling). But these days, in Western societies there are humanitarian and social problems related to the growth of this group, problems which have stimulated political discussion. This has subsequently led to scientific research on the subject. Irregular migrants have become visible in our everyday lives. They defend their rights in self-help groups, protest marches and hunger strikes and they are visible in the streets: we are all familiar with the salesman in the Pakistani night shop, the East-European or South-American cleaning lady in the hotel, the gipsy woman begging for money. These clichés provide prototypical examples of the role irregular migrants play in the public imagination. As a consequence of this situation, both politicians and scientists want to 'grab hold' of what is happening and to acquire an overview of the state of affairs. Predominantly short-term government-driven research is flourishing.

It is interesting to have a closer look at this research. In 2005, the Government of UK attempted to estimate the numbers of irregular migrants and published the UK's first official estimate of irregular migration (Woodbridge, 2005). The aim of the study was to review methods that have been used in different countries to estimate the number of illegally resident persons, followed by an assessment of the applicability of these methods to the UK. The presupposition of this Government-directed initiative is the idea that 'sizing the illegally resident population' will contribute in some way or another to migration policy. The Institute for Public Policy Research

E. Hemelsoet (✉)
Ghent University, Ghent, Belgium

P. Smeyers, M. Depaepe (eds.), *Educational Research: The Ethics and Aesthetics of Statistics*, Educational Research 5, DOI 10.1007/978-90-481-9873-3_7,

(IPPR) tries to look at things in a broader perspective. This organisation defines itself as the UK's leading progressive think tank. Along with other tasks, it tries to engage the media and the public in an informed and evidence-based debate on current policy issues. In the 2006 report (described as a 'FactFile') 'Irregular migration in the UK', researchers are at a standstill regarding key questions considering this topic. Recognising the usefulness of the estimate in the Home Office report, this 'FactFile' attempts to do more than estimate numbers and deal with some of the broader policy issues that irregular migration raises. The following questions are at stake:

– Who are irregular migrants? How do they differ from illegal, undocumented and unauthorised migrants?
– Why do irregular migrants come to the UK?
– How many irregular migrants are there?
– How do irregular migrants enter the UK?
– Where do irregular migrants come from?
– Where do irregular migrants work?
– What is the economic impact of irregular migration?
– How long do irregular migrants stay?
– What rights do irregular migrants have?
– What are the policy options?

The report deals with a whole range of different themes and gives a descriptive overview of 'evidence-based facts' that are for the most part of a quantitative nature. After mapping different definitions related to this group of people, push and pull factors for migration are defined. This study goes beyond a mere estimate of the amount of irregular migrants in the country. Nevertheless, all the other questions refer to an evidence-based paradigm. Answers are given in terms of numbers and quantities that are taken as a starting point to get policy making going.

Similar tendencies are apparent at the European level. Clandestino is 'an interdisciplinary project' and is 'a response to the need for supporting policy makers in designing and implementing appropriate policies regarding undocumented migration' (Picum, 2009, p. 2). Twelve EU countries participate in this project that aims at

– providing an inventory of data and estimates on undocumented migration (stocks and flows) in the selected EU countries;
– analysing these data comparatively;
– discussing the ethical and methodological issues involved in the collection of data, the elaboration of estimates and their use;
– proposing a new method for evaluating and classifying data/estimates on undocumented migration in the EU.

Again, the number of undocumented migrants in the different countries is at the heart of the research. The fact that there are large methodological problems

related to estimating irregular migrants has not prevented policy makers from focusing on this area. Paying lip service to unreliable figures is having far-reaching consequences. Before approaching this issue, I will consider, what is said in these reports and the conclusions that have been drawn from the findings.

7.2 Techniques for Estimating Irregular Migration Numbers

A lot of creativity is put into action to estimate numbers of irregular migrants. The list of different techniques that can be used is long and varied. Without going too deeply into the particularities of different techniques and the problems each presents us with, an overview of methods is outlined here. Then attention is given to some examples of the most widespread methods and how they are applied.

In general, there is a consensus on the classification of various techniques. Recent reports of the Clandestino Project (Kraler & Vogel, 2008; Vogel & Kovacheva, 2008; Picum, 2009) and the Home Office (Woodbridge, 2005; Pinkerton et al., 2004) all start from the distinction between '*stocks*' and '*flows*', in the line of earlier research (among others Delauney & Tapinos, 1998; Pinkerton, McLaughlan, & Salt, 2004; Jandl, 2004). This fundamental distinction is made in analogy with data on legal migration and refers to different statistical concepts: *stocks* (e.g. of irregular residents or migrant workers) refer to a number of people that are present in a country *at a particular point in time*, while *flows* (e.g. of illegal entrants or migrants 'overstaying') refer to *movements during a certain period*. Flow estimates thus refer to the number of migrants crossing the border during a period in time. Estimations are mainly based on border apprehension data or entry-exit statistics. Given the volatile nature of migration flows, the scarcity of reliable indicators and the dearth of appropriate methods for estimating flows, most efforts have so far concentrated on estimating stocks rather than flows. Methods for estimating stocks of irregular migrants can be divided into direct and indirect approaches. *Direct measurement* is based on data that 'captures' the subject of research directly. Numbers of irregular migrants taken from administrative statistics based on refusals, infractions or regularisations are multiplied to estimate the total number in the population. *Indirect estimates* do not rely on such data. They mainly start from general data on the population to estimate irregular migrant numbers (e.g. by expecting similar proportions of irregular migrants or calculating residues in official registers). Direct estimation approaches can be further classified into multiplier methods (among which simple multiplier models, capture-recapture models and models using a comparison of administrative registers and random effect mixed modelling), methods of self-identification and snowball sampling methods. Among the indirect approaches, residual methods, demographic methods, subjective estimations or indicator methods, econometric methods on the size and structure of shadow economies, comparisons of immigration and emigration statistics, flow-stock methods and methods based on indirect inferences can be distinguished. It is not the aim here to provide an extensive discussion of all the methods mentioned above (for

a very informative schematic overview, see Appendixes 1 and 2). Three examples are given though, as they are paradigmatic for the problems related to estimating irregular migrants.

Multiplier methods calculate the size of the unknown total from the size of a known subtotal by use of an appropriately estimated multiplier. The use of multipliers to derive the size of a hidden population from the size of a known subtotal of that population is probably the most common way of estimating an unknown population. The problem of giving adequate estimations is then translated in finding the right multiplier (Vogel, 2002). An example of using the multiplier method for estimating irregular migrant numbers is provided by a recent study in Belgium by Van Meeteren, Van San, and Engbergsen (2007). In this study, police data on arrested criminal foreigners is combined with data from in-depth interviews with 120 irregular migrants. The police data on criminal offences committed by foreigners without legal residence is compared with the 'crime rate' among the migrants that were interviewed. Three assumptions are internal to this method, assumptions that modify the quality of the outcomes. First, there is the expectation that the crime rate derived from the interview sample is a good indicator for the actual crime rate among the target population. Second, the estimate is based on the assumption that the reported crime figures are a good indicator for total crime figures. Thirdly, the sampling group is expected to be representative of the total population of irregular migrants. Concerning the first assumption, it can be argued that long in-depth interviews lead to sufficient trust between the interviewer and the respondent so as to retrieve realistic answers. As regards the second assumption, things get slightly more complicated. It is not clear to what extent police data on arrests cover the total amount of criminal offences that are actually committed. It is very likely that the police do not detect all cases of crime and the resulting estimate is therefore unreliable (Kraler & Vogel, 2008). The third assumption implies problems concerning generalisation, as the sampling group is rather small and it is impossible to retrieve a randomised sample if the composition of the total group is largely unknown.

Similar problems arise with the capture-recapture method. This method is used in biology to estimate animal populations in the wild. The basic idea is to develop a multiplier through repeated sampling of the same population. Estimating a fish population goes as follows: first, capture 1,000 fish and mark them before releasing them again. Capture another 1,000 fish and look how many are marked. If for example 100 are marked, the 1,000 fish will statistically make up approximately 10% of the total, i.e. 10,000 fish. In the Netherlands Van der Leun, Engbersen, and van der Heijden (1998) have used this principle to estimate numbers of irregular migrants. Their 'repeated capture method' is based on a dataset that tries to apprehend numbers of illegal immigrants in Amsterdam, Rotterdam, The Hague and Utrecht, (the four big Dutch cities) during 1995. Using the number of persons captured and arrested again, it is argued that the number of people who will show up again follows a probabilistic distribution called the Poisson distribution (λ). On the basis of the available data, the crucial parameter determining the Poisson distribution can be estimated. This

is then used to identify the probability of an individual being caught by the police. Again, a number of problematic assumptions are at work. First, the population has to be homogeneous (with respect to the risk of being caught). This can be met through the use of an appropriate regression method, which accounts for determined features for apprehension, such as age, sex and origin. Second, there has to be a more or less stable chance of getting caught. The supposition that no major policy changes take place during the time span of the research is thus required. But most problematic is the third and final assumption: the population under consideration has to remain constant during the period of research. Such stability is highly unlikely as there are peaks in flows of irregular migration (e.g. seasonal workers). This leads researchers to make the decision to exclude certain groups from the study (Pinkerton et al., 2004).

As a final example of estimation methods, demographic methods start from the idea that rates concerning e.g. birth, mortality and hospitalisation are normally distributed over the total population. Legal and illegal people are thus supposed to experience these events to the same extent. Proportions of irregular migrants' hospitalisations, deaths and births are then extrapolated to estimate total numbers. The practical advantage of these methods lies in the wide availability of the required data, which means that there is no need to acquire new datasets. However, the assumption that these demographic rates are similar for illegal and legal residents can be questioned. In some populations, the birth rate is very high (e.g. Roma population, Ringold, Orenstein, & Wilkens, 2005). Irregular migrants often live in worse conditions than natives and are more susceptible to health problems. A larger risk of disease due to poverty and bad living conditions may thus lead to an over-representation of data. Furthermore, there are registration problems. Irregular migrants are often very mobile and may choose to return due to impending demographic risks (e.g. chronic illness, death). Sometimes, things may be the other way around: people come here to enjoy medical care that is lacking in the country of origin. Finally, differing benefits in the health-care system may lead to serious underestimations. Irregular migrants are not always hospitalised when required, because they cannot rely on the same benefits or they are afraid to profit from them due to their illegal residence status.

7.3 Counting the Uncountable: Some Conceptual Problems

Social scientists generally acknowledge the problems resulting from various estimation methods. They are therefore reluctant to identify estimates, and differ over the degree of reliability that they are willing to attribute to these numbers. Problems of measurement evidently follow from the fact that irregular migration is resistant to registration and statistical description (Tapinos, 1999). Estimating the numbers of illegal residents in a country is almost impossible given the fact that the phenomenon is a grey zone. Nevertheless, a great deal of pressure is put on researchers

to make such estimates. Both journalists and policy makers need numbers to develop a particular argument and make policy. As Clarke (2000) argues, a number that is quoted somewhere in the press has a good chance to develop a life of its own and survive for many years in the media where it is rehearsed as a fact, even if the original estimate was accompanied by disclaimers. Surprisingly, problems pertaining to unreliability do not deter researchers from producing estimates. Scientists seem to be equally attracted to numbers (Kraler & Vogel, 2008), and search for solutions to deal with the observed problems. These problems are redefined in terms of a lack of quality assessment of the methods of estimation. What is thus required is 'a classification scheme that serves as an indicator of the scientific quality of the estimates' (Kraler & Vogel, 2008, p. 6). 'Quality classes' (high/medium/low-quality estimate and low estimate with a plausibility warning) are distinguished, estimates are formulated in terms of a minimum and a maximum and different methods are intertwined ('combined methods') to increase reliability. But still some maintain that 'even low-quality minimum or maximum estimates without ranges can be useful in a national context if they challenge generally shared assumptions about the size or composition of a group' (Vogel & Kovacheva, 2008, p. 9).

The quality of estimation outcomes remains problematic. Moreover, in what follows it will be argued that there are fundamental problems underlying the search for qualitative estimates of a conceptual nature. In most cases these problems are hardly touched upon. They concern differing definitions, data sources and collection methods that are used when comparing data. Moreover, different legislations in the respective countries generate problems of comparability. The problematic character of conceptualisations when estimating irregular migrant numbers will be treated in two different ways. First, the focus is on the relation between the definition of the target group of research and the way the problem is conceptualised. Secondly, the unavoidable tendency towards homogenisation that is intrinsic to reducing the complexity of a large amount of data to a limited number of differences will be problematised.

7.3.1 Different Definitions, Different People

In the literature a number of descriptive strategies are used to denote people without legal residence. Adjectives such as illegal, irregular, unauthorised, undocumented or clandestine are often interchangeably combined with the nouns (im)migrants, aliens or refugees. There is debate over the appropriateness and applicability of these various concepts. Mostly, discussion is on the (often emotive) connotations associated with different terms. Some authors talk about refugees (Hamilton & Moore, 2004), referring to the definition of the 1951 United Nations Geneva Refugee Convention: 'A refugee is a person who owing to a well-founded fear of being persecuted for reasons of race, religion, nationality, membership of a particular social group or political opinion, is outside the country of his nationality and is unable or, owing

to such fear, is unwilling to avail himself of the protection of that country; or who, not having a nationality and being outside the country of his former habitual residence as a result of such events, is unable or, owing to such fear, is unwilling to return to it'. (UNHCR, 1993, p. 6). This term refers to the group of migrants who have been found to 'qualify' for official refugee status (Watts & Bridges, 2005). In many countries, people who do not meet the criteria that are set by the government to be a 'legal' refugee are often called 'undocumented migrants' (or *sans-papiers*), a term preferred by, among others, Paspalanova (2006), although it is unofficial and has no legal force. In scientific literature, the term 'illegal aliens' (Van Dijck, 1996; Ommundsen & Larsen, 1999; Martiniello, 2005) has, to a large extent, been replaced by the term 'irregular migrants' (Jandl, 2007; Broeders & Engbersen, 2007; Laubenthal, 2007). This is due to the fact that 'illegality' is too often connected with criminal behaviour and human rights advocates argue that 'no human being is illegal'.

Without going deeper into the discussion over which concept is most appropriate, it should be noted that the definitions of preference are decisive for the way a problem is conceptualised. Using different terminologies as well as different ways of defining the used concepts can influence the statistical outcomes to a large extent and will finally lead to different conclusions and recommendations. The target group 'irregular migrants' exemplifies the importance of conceptual demarcation: definitions frequently have emotive connotations, they are often unclear and they are used in very different ways. It is thus of crucial importance that research reports clearly indicate the definitions to which they subscribe, who is or is not included in the target group and why this is so. The motives behind conceptual choices should therefore be made explicit due to the potentially unpleasant consequences of not doing so. That is not as evident as it may seem. Definitions of irregular migration always include elements of a negative definition (people *without* papers, who have *no* legal residence status, who *cannot* rely on particular kinds of support). Therefore, policy changes towards regular migration can lead to substantial changes in the irregular migrant population. This is most obviously the case when regularisation programmes take place, but besides this, attribution of, for example, working permits and medical cards to particular groups may temporarily or permanently change their legal status. Definitions are constructed in relation to their practices of reference that may historically and locally differ to a large extent. The impossibility of deriving concepts from empirical data lies at the basis of the incomparability of much of the research data and outcomes. Unfortunately, this does not stop policy makers and journalists making comparisons. The love for numbers comes with a desire to compare. A good example is the Clandestino project that is funded by the European Union. This project tries among other things to estimate the stock of irregular migrants in different countries. The total number of irregular migrants in Europe (Picum, 2009) is estimated to be between 2.8 and 6 million. To complicate matters, there is no implication that the mean estimate is the most likely number. The width of these estimates is huge and their reliability is questionable. Estimates from different countries are made with different (more and less reliable) methods. Different

definitions are used and again it should be stressed that policy differences cannot be overcome. Some figures are even calculated on the basis of extrapolation of the number of irregular migrants that reached their destination countries by looking at the number of migrants arrested at the borders. Based on discussions with border control authorities, a multiplier is defined (for example, multiplying 60,000 border apprehensions in 1993 multiplied by 4–6, cf. Widgren, 1994). Estimations then hardly seem to be more than mere 'guesstimates' as they are to a large extent based on ungrounded assumptions. Moreover, with international comparisons, there are always added difficulties when comparing the different datasets as they measure different things, using different systems, at different times and by different authorities (Clarke, 2000).

This does not keep researchers from drawing remarkable conclusions. It is recognised that 'the review of efforts to estimate the size of irregular migration on a European level has shown that the numbers indicated are based on very rough estimates' because 'often we do not know which groups of irregular migrants are included in a stock estimate, nor do we know whether a flow estimate is meant to measure net inflows or gross inflows' and 'older studies are often quoted in newer studies, so that estimates appear to apply to the present, although they were made some years ago' (Iglicka, in Kraler & Vogel, 2008, p. 17). Still, researchers say that 'approximate comparability is better than no data at all in a situation where a high degree of comparability may never be achieved' (Vogel & Kovacheva, 2008, p. 17) or that 'these estimates will greatly aid policy making' (IPPR, 2006, p. 9). Reference is made to policy: 'In the public sphere, there is a general need to gather reliable information on important social phenomena, to determine whether or not the situation warrants any political action. [. . .] For governments, the perceived size of the phenomenon will have an important bearing on the justification for the expenditure of public resources on alternative uses' (Jandl, 2004, p. 152). The idea that 'approximate comparability is better than no data at all' seems to count here as well.

Given the far-reaching consequences at policy level, glossing over the large and to a great extent insurmountable problems that are related to estimating irregular migrant numbers is highly problematic. Or as Clarke (2000, p. 21) states

> . . . whichever method of assessment is used, estimated numbers of irregular migrants are based on assumptions, many of which are either untested or maybe even untestable. The fact remains that unrecorded and irregular migration is, by its very definition, unquantified and, indeed, unquantifiable. Any figure generated is at best an educated guess. [. . .] And yet there are many pressures from policy makers, the media and other interest groups to come up with a quotable figure. When an estimated figure is 'calculated' no matter what warnings and disclaimers are attached to it, as has been seen a number of times in the last few years of growing interest in trafficking and 'illegal' migration, there is a danger that this figure is picked up and rapidly circulated and, before long, quoted as a fact.

The nature of Clarke's critique is clear, and goes beyond concrete problems related to particular methods. From a scientific perspective, there is a lack of reliability in the estimations obtained. To a certain extent, the phenomenon of irregular

migration indeed *is* unquantifiable and the best that can be expected from these estimations *is* an educated guess. But the weight of these scientific arguments, which strive for objectivity, is limited in the politicised context of policy making. For some this is regrettable. It is true that one may find examples of research that is conducted to provide the rationalisation for a predetermined policy. In these cases, the outcomes of the research are known long before it has even begun. Research is used as an instrument to legitimise political choices. This is unacceptable but is not what is really at stake here. At the very least, the arguments in defence of using estimates should be taken seriously. A requirement for political action is a well-founded insight in the current state of affairs. To be able to deal with problems related to irregular migration, the nature of these problems requires clarification. Part of that job is to get an idea of the scope of the phenomenon, i.e. to 'measure' its 'size'. The 'love for numbers' demonstrated by policy makers may then be an expression of sincere engagement with these problems, rather than a fast solution to legitimise decisions.

The fact that 'something has to be done' by policy makers does not take away the fundamental critique on irregular migrants' estimations. But it does mean that critics have to offer a possible alternative. If not, the argument that from a policy view 'approximate comparability is better than no data at all' is legitimate, although, from a scientific perspective, this is not always so. What is left from the critique now is twofold. First, the critique can serve as a 'reminder' for policy makers. It does so by stressing the importance of conceptual demarcation and underlining that the information provided by estimations is very limited. It is not the estimations that are the problem, but rather the fact that these numbers start to lead a life of their own. Policy makers should be aware of this danger and the limitations of estimations. Second, the critique may be a step-up to develop other research strategies. This refers to the idea that estimates are an *insufficient* condition to gain insight into the problems related to irregular migration. Again, this does not mean that they cannot be helpful; it only means that their contribution has to be put into perspective. Other questions and different kinds of (e.g. qualitative) method(ology)s may form a valuable supplement to the acquired numbers. This may therefore lead to *alternative* ways of approaching irregular migration.

7.3.2 The Homogenisation of Complexity

The question of what can be expected from estimating irregular migrant numbers is currently being explicitly brought to the fore. The group of people involved is predominantly approached from a legal point of view that focuses on residence status. In this case, 'irregularity' serves as the benchmark of categorisation. The ethnicity, nationality, religion, native language, cultural background etc. of the people involved is barely considered, as from a policy perspective what is at stake is their legal status. For asylum seekers, things are slightly different. On arrival, attention is directed at political or personal reasons for taking refuge and nationality is

a very relevant factor for recognition of refugee status. But once people are categorised into the group of 'irregular migrants', they are treated as a monolithic group of people. Contextual elements slide to the background and become rather irrelevant. The residence status is the qualifying benchmark to be included or excluded from all kinds of (citizen) rights, provisions and initiatives. Everything depends on the documents one has, or does not have.

As a consequence, the quest for these documents serves as the binding agent for the people lumped together under the denominator 'irregular migrants'. This experience *is* shared by all of these people. They come across each other e.g. in charity organisations or in shared waiting queues for public services. But at least as important as this guiding similarity are the observed differences between irregular migrants. Various cultural backgrounds, motivations to take refuge and personal histories that lead to irregularity are often incomparable. An Eastern European girl kidnapped and smuggled into Western Europe by human traffickers for prostitution deals with problems that have little to do with those faced by her immigrant worker compatriots who come here to work for a few months on the black market so as to gather as much money as possible to send or take back to their homeland. Similarly the unaccompanied child soldier has little in common with the African soldier of fortune looking for a better life in the rich West, or the South-East-Asian refugee seeking safety after suffering persecution in her homeland. Roma gypsies often feel comfortable with their irregular status, and are often not even interested in regularisation. 'Being irregular' in many cases seems to have become a part of their cultural identity, to which they have adapted. Of course, within these different groups of people there are large differences; homogenising them through stereotypes is not what the preceding examples are intended to do. What I am trying to emphasise here are the divergent perspectives and (previous and current) living conditions of all these people (at a cultural or ethnic but at least as much at a personal level). The government also recognises that incomparable situations refer to different problems that are not necessarily related to residence status. Human trafficking, war traumas of unaccompanied minors, moonlighting, clandestine (drugs) trade, etc. are important problems that cannot be captured under the name of 'irregular migration'. Unfortunately, this is exactly what is being done when numbers of irregular migrants are estimated. None of the studies mentioned above draw attention to these differences. This leads to injustice for the people involved. They are (independent of the particular term used to define irregular migrants) linked to problems that may have nothing to do with their personal conditions. From a policy perspective, it shows that treating these people as a monolithic group takes away the possibility of making desirable distinctions. In statistical terms, this reflects the danger intrinsic to reducing the complexity of a large amount of data to a limited number of differences. Restricting the degrees of freedom to obtain more accurate data inevitably has homogenising consequences. What is left are vague numbers that hardly contain any relevant information. That brings us to the last section of this chapter, in which explore alternative ways of dealing with the aforementioned problems are explored.

7.4 Does It Make Sense To Say. . .

The discussion presented above shows how difficult it is to 'make sense' of the divergent problems related to the group of irregular migrants. In this last part of the chapter we move beyond the methodological matters concerning the adequacy of various estimation methods. Let us now consider the appropriateness of the questions put forward by policy makers and researchers. Why are these questions seen as being so attractive, so apparently integral to the situation presented above? Willingness to 'count the uncountable' is surprisingly strong. The fact that the goal of many irregular migrants is to go into hiding does not diminish the conviction that estimations are helpful when trying to understand the problems at stake. Neither does the finding that it is impossible to even roughly estimate irregular migrants' numbers keep researchers from doing exactly that. Something more profound seems to underlie the apparent 'love for numbers' that to a large extent characterises Western societies.

Blommaert and Verschueren (1998) largely elaborate on their thesis that 'the 'migrant debate' rests on the idea that the ideal society should be as uniform or homogeneous as possible. Homogeneity is not only seen as desirable, but also as the norm, as the most normal manifestation of a human society." (ibid., p. 117) In their linguistic discourse analysis, they put forward an impressive number of examples that support this thesis. '"Homogeneism" is seen as the dominant ideology that directs our thinking about foreigners. Through the abnormalisation of the foreigner, it contains an a priori rejection or problematisation of diversity. [. . .] Foreigners disturb the existing order, they threaten the status quo, their presence alone already turns them into a problem'. (ibid., p. 146). The attempts to homogenise the group of migrants – which seems to hold for irregular migrants too – thus come from a defensive reflex. This attitude expresses a fear towards the unexpected, the new and the unknown i.e. towards whatever can change or damage what has been established.

This critique of homogenisation brings a different perspective to the question of 'how many irregular migrants there are'. If policy makers ask this question so as to bring about 'reliable information' (cf. supra) for dealing with problems related to the presence of irregular migrants, then its contribution seems limited. The repeated search for numbers hardly reveals anything with respect to the content of concrete practices, as by definition it is an expression of the homogenisation tendency mentioned earlier. What is at stake concerns attempts to 'grab' the situation and the changes to that situation. Numbers help to 'get an overview' and in that sense to 'control' what is happening. The apparent 'love for numbers' in Western societies seems to refer to the dominant presence of a thinking that divides the world into 'us' and 'them' and continuously looks for 'ownership'. Numbers become the tools to realise that ownership: quantifying reality becomes synonymous with objectifying reality i.e. gaining control over the world surrounding us. One could even argue that the attraction of numbers refers to what Nietzsche called *Wille zur Macht*: the will to power as the driving force behind human actions.

Recognising the 'will to power' in the current love for numbers gives us sufficient reason to judge the estimation practices in the context of irregular migration. But, as mentioned earlier, that does not solve the problem at hand: what course of action is left for the policy maker? The question that remains concerns whether or not there are other ways to 'understand' the situation that can be distinguished apart from attempts to 'grab' it. A possible alternative may be to develop a problem-oriented policy that works towards finding solutions to current problems. Some irregular migrants would then be liable to prosecution as a result of policy that fights smuggling, human trafficking, poverty, etc. Stretching this line of thought, some decisions might even be directed towards particular cultures, religious groups or ethnicities. One can for example think about discussions on the foundation of Muslim schools in Western European countries or initiatives to promote school participation for Roma people. Irregular migrants are then not approached from the perspective of their (illegal) residence status, but from events they have participated in or been victimised by, or from the ethnical, cultural or religious group they belong to. It is unclear whether or not this would deal with the problem. Even if this approach is adopted, numbers will be sought after to estimate the size of the problem. Whether we are talking about irregular migrants, smuggling, criminality, poverty, Muslims, Roma people or anything else, obtaining an idea of the scope of a social problem seems an indispensable element of 'understanding' that problem and gaining insight into it even if this is only so that policy makers can decide on how the limited means that are available should be distributed.

The earlier critique of attempts to estimate irregular migrants provided evidence of the limited value of these numbers when trying to gain insight into the phenomenon. However, the pragmatics of policy making shows that this limited contribution to understanding irregular migration is better than no contribution at all and that the critique itself has limits too. Estimations make sense, even if their reliability is very low. Although they cannot deliver the profound insight into social problems that is required for a well-founded policy, they can serve as a stepping-stone for further research. The 'will to power' may indeed be present in the attempts to frame social phenomena in numbers, but this does not explain the 'love of numbers' and the attractiveness of quantitative data. Every search for understanding, whether it be quantitative, qualitative or even philosophical, is expressing a will to power i.e. to gain control over a situation.

What makes numbers so attractive then might be their transparency and unambiguous character. That takes us back to the second element of the conceptual critique, namely, the homogenisation of irregular migrants. Here, the proposed policy alternative may deliver added value. This alternative does not start from the dichotomy between regular and irregular people, between people with a legal and an illegal residence. It is important to recognise the diversity of problems irregular migrants are dealing with. Indeed the very concept of an 'irregular migrant' (or any alternative concept referring to this group of people) as an umbrella to homogenise the present diversity should be seen as problematic. This could be a starting point for a more just approach to the complexity of their everyday social practices.

Appendix 1: Methods for Stock Estimates

Approach	Data sources	Method	Estimation technique	Main idea of calculation
Direct approaches	Based on immigration enforcement data	Multiplier methods	Simple Multiplier	Estimation of total with a simple multiplier based on derived or estimated ratio of 'dark field' vs. 'clear field'
			Capture-recapture/ Repeated capture	Estimation based on probabilistic function derived from multiple recaptures of individuals in sample
			Matching of registers	Estimation based on implied non-registration in two or more individually matched registers
			Random effect mixed modelling approach	Estimation using statistical regression model assuming comparable apprehension rates of legal/illegal residents with statistical adjustment for random effects
	Based on administrative statistics	Methods of self-identification	Evidence based on regularisation data	Inferences on the size and composition of irregular migrant stock prior to regularisation from data on applications for and grants of regularisations
			Using data on status adjustments over time	Inferences derived from data on changes in residence status after a period of irregular residence
	Based on surveys	Survey methods	Direct survey methods	Reconstruction of a "random sample" of regular and irregular migrants through a re-weighting of the probability of contacts
			Snowball sampling methods	Estimation using chain referral methods to obtain a sample of persons not registered vs. persons registered

Approach	Data sources	Method	Estimation technique	Main idea of calculation
			Respondent-driven sampling	Recruitment of interviewees through peers and incentive system leads to equilibrium sample of respondents after several recruitment waves which is independent from original sample and can be analysed statistically
Indirect approaches	Based on census/registers	Residual methods	Differences census results – legal immigration data	Indirect estimation based on the calculated difference between census data and data on legal immigrants
			Simple comparison of various registers	Indirect estimation based on a comparison of two or more registers with data on the same target population
		Demographic methods	Use of birth/death rates	Inferences on demographic subgroups based on the comparison of real and expected birth or death rates
	Based on census/registers/ demographic data	Expected population methods	Comparison of census/emigration data and immigration statistics	Indirect estimation of illegal resident population from comparison of emigration estimates with data on legal immigrants at destination
	Based on administrative statistics	Flow-stock methods	Calculating the stock through flow figures	Using estimated inflow and duration of stay indicators to estimate steady-state stock of illegal residents
	Based on complementary data sources and estimates	Indirect inferences	Using information on correlated phenomena as basis of calculation	Making inferences on subgroups of irregular foreign residents on the basis of indirectly related phenomena and estimates such as irregular work, sector-specific demand for irregular services, school attendance or health services (e.g. inference of share and size of irregular foreign workers from econometric estimates on the shadow economy)

Approach	Data sources	Method	Estimation technique	Main idea of calculation
	Based on surveys of 'key informants'	Subjective Estimation/Indicators Methods	Expert surveys	Survey of key informants on their assessments of sizes, ratios and characteristics of target population
			Delphi surveys	Anonymous multiple-round survey of key informants mediated by researcher to attain convergence of opinion
Combined approaches	Based on small-scale surveys	Window/Postal code method	Small-scale study + use of regression analysis	Extrapolation of estimates derived from intense local study with regression analysis
	Based on expert opinions	Localized Delphi	Delphi method + use of regression analysis	Extrapolation of estimates derived from localized Delphi study with regression analysis
	Adjustment to surveys/census data	Non-threatening survey design	Direct survey method + randomized response/3-cards method + residual method	Statistical inferences from employer survey using randomized response method and inferences about share of irregular migrants in sample survey using non-threatening survey questions combined with residual estimation results

Source: Kraler, A. & Vogel, D. (Eds.). (2008). *Clandestino report on methodological issues*. Athens: European Union, p. 22–24.

Appendix 2: Methods for Flow Estimates

Approach	Data sources	Method	Model	
Direct approaches	Based on border apprehension data	Multiplier methods	Simple Multiplier	Estimation of total illegal border crossings by applying estimated or derived multipliers on border apprehension data
Indirect approaches	Based on stock estimates	Differential methods	Net differences in stocks	Deriving estimated annual net increase of irregular migrants through changes in estimated stocks
	Based on entry-exit statistics	Residual method	Double entry card system	Deriving estimated number of overstayers through individual matching of entry-exit records

Source: Kraler, A., & Vogel, D. (Eds.). (2008). *Clandestino report on methodological issues.* Athens: European Union, p. 24.

References

Blommaert, J., & Verschueren, J. (1998). *Debating diversity. Analysing the discourse of tolerance*. London: Routledge.

Broeders, D., & Engbersen, G. (2007). The fight against illegal migration – Identification policies and immigrants' counterstrategies. *American Behavioral Scientist, 50*, 1592–1609.

Clarke, J. (2000). The problems of evaluating numbers of illegal migrants in the European Union. In P. De Bruycker (Ed.), *Regularisations of illegal immigrants in the European Union* (pp. 13–22). Brussels: Bruylant.

Delauney, D., & Tapinos, G. (1998). *La mesure de la migration clandestine en Europe*. Population and social conditions working paper 3/1998/E/nr.7. Luxembourg: Eurostat.

Hamilton, R., & Moore, R. (2004). *Educational interventions for refugee children. Theoretical perspectives and implementing best practice*. London: Routledge.

IPPR. (2006). *Irregular migration in the UK*. London: Institute for public policy research. Retrieved September 1, 2009, from http://www.ippr.org/members/download.asp?f=%2Fecomm%2Ffiles%2Firregular%5Fmigration%2Epdf

Jandl, M. (2004). The estimation of illegal migration in Europe. *Studi Emigrazione/Migration Studies, 41*, 141–155.

Jandl, M. (2007). Irregular migration, human smuggling, and the eastern enlargement of the European Union. *International Migration Review, 41*, 291–315.

Kraler, A., & Vogel, D. (Eds.). (2008). *Undocumented migration: Counting the uncountable. Data and trends across Europe. Report on methodological issues*. Athens: Clandestino. Retrieved September 1, 2009, from http://clandestino.eliamep.gr/wp-content/uploads/2009/02/clandestino_report-methodological-issues_final1.pdf

Laubenthal, B. (2007). The emergence of pro-regularization movements in western Europe. *International Migration, 45*, 101–133.

Martiniello, M. (2005). Alien policy in Belgium, 1840–1940. The creation of guest workers, refugees and illegal aliens. *International Migration Review, 39*, 973–974.

Ommundsen, R., & Larsen, K. S. (1999). Attitudes toward illegal immigration in Scandinavia and United States. *Psychological Reports, 84*, 1331–1338.

Paspalanova, M. (2006). *Undocumented and legal Eastern European immigrants in Brussels*. Leuven: Katholieke Universiteit Leuven.

Picum. (2009). *Report on an International Workshop organized by PICUM, MRN and COMPAS on Understanding Irregular Migration in Northern Europe* (London, UK, March 27, 2009). Brussels: Picum.

Pinkerton, C., McLaughlan, G., & Salt, J. (2004). *Sizing the illegally resident population in the UK* (2nd ed., Home office online report 58/04). London: Migration Research Unit, University College London.

Ringold, D., Orenstein, M. A., & Wilkens, E. (2005). *Roma in an expanding Europe: Breaking the poverty circle*. Washington, DC: The World Bank.

Tapinos, G. (1999). Clandestine immigration: Economic and political issues. In Organization for Economic Co-operation and Development (Ed.), *Trends in international migration* (pp. 229–251). Paris: OECD.

United Nations High Commission for Refugees. (1993). *The 1951 refugee convention: Questions and answers*. Geneva: UNHCR.

Van der Leun, J., Engbersen, G., & Van der Heijden, P. (1998). *Illegaliteit en criminaliteit: schattingen, aanhoudingen en uitzettingen*. Rotterdam: Erasmus Universiteit Rotterdam.

Van Dijck, T. A. (1996). Illegal aliens. *Discourse & Society, 7*, 291–292.

Van Meeteren, M., Van San, M., & Engbergsen, G. (2007). *Irreguliere migranten in België. Inbedding, uitsluiting en criminaliteit*. Rotterdam: Erasmus Universiteit Rotterdam.

Vogel, D. (2002). Ausländer ohne Aufenthaltsstatus in Deutschland. Methoden zur Schätzung ihrer Zahl. In E. Minthe (Ed.), *Illegale Migration und Schleusungskrinimalität.* (pp. 65–78). Wiesbaden: Eigenverlag der Kriminologischen Zentralstelle.

Vogel, D., & Kovacheva, V. (2008). *Clandestino classification report. Quality assessment of estimates on stocks of irregular migrants*. Athens: Clandestino.

Watts, M., & Bridges, D. (Eds.). (2005). *Higher education opportunities for refugees, asylum seekers and migrant workers in the East of England*. Cambridge: Association of Universities in the East of England.

Widgren, J. (1994). *Multilateral Co-operation to Combat Trafficking in Migrants and the Role of International Organisations*. Paper submitted to the 11th international organization for migration seminar on migration 26–28 October 1994, Geneva: International Response to Trafficking in Migrants and the Safeguarding of Migrant Rights, Paper number 6, Geneva: IOM.

Woodbridge, J. (2005). *Sizing the unauthorised (illegal) migrant population in the United Kingdom in 2001* (Home office online report 29/05). London: Migration Research Unit, University College London.

Chapter 8
European Citizenship and Evidence-Based Happiness

Naomi Hodgson

8.1 Introduction

Michel Foucault's analysis of the development of modern forms of government in the eighteenth and nineteenth centuries and his account of governmentality opened up a perspective through which to analyse the operation of power (Foucault, 1991a, 1991b, 2005). The emergence of statistics – "the science of the state" (Foucault, 1991a, p. 99) – and the objects of statistical knowledge were identified as part of the complex of techniques for the production of a particular form of normalised individuality. More recent analyses from this perspective have shown how the object of government has shifted in advanced liberal modes of governmentality (e.g. Barry, Osborne, & Rose, 1996) and thus how the citizen has been reconceptualised in the context of the decline of the welfare state and the increasing privatisation of expertise, "implanting in citizens the aspiration to pursue their own civility, wellbeing and advancement" (Barry et al., 1996, p. 40; Biesta, 2009, p. 150). In this context, education or learning has become central to the production of a particular form of citizenship. Learning is no longer understood as taking place in particular institutions or designated periods in our lives (Rose, 1999) but as being a lifelong and life-wide process of self-improvement, requiring an attitude of willingness to adapt to the constantly changing demands of the globalised, competitive, knowledge economy (Simons & Masschelein, 2008). The role of statistics has also shifted: while "statistics might once have been a governmental activity, since the middle of the twentieth century it has become a business" (Rose, 2000, p. 230).

From this perspective, I draw attention here to the objects of statistical analysis that provide a current focus for measurement and policy-making in Europe and thereby constitute key indicators according to which states, institutions and individuals become measurable, comparable and governable. I focus, in particular, here on happiness and well-being and how this relates to the construction of 'active citizenship' (Hoskins et al., 2006).

N. Hodgson (✉)
Institute of Education, University of London, London, UK

P. Smeyers, M. Depaepe (eds.), *Educational Research: The Ethics and Aesthetics of Statistics*, Educational Research 5, DOI 10.1007/978-90-481-9873-3_8,

The objects of statistical analysis, the presentation of data and the way in which citizens are addressed in terms of questions of happiness and well-being are seen to be constitutive of not only the European citizen but also Europe itself. The creation of European citizenship as a formal legal category by the Maastricht Treaty in 1992 has since required measures to create an allegiance to Europe among this citizenry. The creation of symbols of Europe, such as the flag and the currency, is particularly a visible example. Educational and cultural measures, in particular, have also been used to interpellate individuals as European citizens, for example, through the promotion of a shared European heritage (see Shore, 2000) and through educational exchange programmes between member states (e.g. SOCRATES and ERASMUS). More recently, developments have taken place at the local political and individual level as Europe has sought to standardise practices of measurement, presentation and accountability within and across the member states, for example, through the Bologna Process to create a European Higher Education Area (http://www.ond.vlaanderen.be/hogeronderwijs/Bologna/) and the Lisbon Treaty, which came into effect in 2009 (http://europa.eu/lisbon_treaty/index_en.htm). This process of what has been termed the "Europeanisation" of Europe (Shore, 2000, p. 221) requires the categorisation of phenomena as European. Concepts such as "European citizen", "common European values", "European culture" and "European public opinion" have become part of our vocabulary that renders them less open to question. They "become part of the fabric of our subjectivity" (Shore, 2000, p. 29).

I will first provide some recent background to the development of happiness and well-being as objects of political concern. This is not to suggest that these have not been a matter of governmental concern previously but that they appear in a particular way in the current context. I will then provide examples of the way in which this academic and professional knowledge has been taken up at different levels of governance (e.g. OECD, Europe, national government) and formalised to become an accepted way of understanding what such organisations should seek to measure, and how this relates to the wider discourses prevalent in policy-making today of transparency, democracy, sustainability and social justice. I will then turn to discuss how the discourse of happiness and well-being operates in a specific educational context.

8.2 The New Science of Happiness

In recent years, GDP (Gross Domestic Product), the well-established measure of a country's progress and wealth, has been criticised for its inability to fully account for the complexity of societies and their populations. The EU now speaks of moving "beyond GDP". According to one European Commissioner for Economic and Monetary Policy:

> GDP was never intended to be anything but an indicator of economic performance. It cannot distinguish between activities that have a negative or a positive impact on wellbeing. In fact, war and even natural disasters may register as an increase in GDP. . . Of course, economic growth can bring about an improvement in quality of life, but only up to a point. Indeed,

> many studies of affluent countries do not register an increase in happiness in line with wealth. (Almunia, 2007, p. 2)

GDP cannot take account of gaps between rich and poor or of levels of sustainability, for example. Measurement now needs to account for quality rather than only quantity. This move beyond GDP is formalised in an International Initiative comprising the European Commission, the European Parliament, the OECD, the Club of Rome and the World Wildlife Fund (see http://www.beyond-gdp.eu/ and the European Commission's 2009 report "GDP and beyond: measuring progress in a changing world").

The statement by Almunia cited above derives from the influential findings of the economist Richard Layard in the emergent field of "happiness economics". Layard contends that although wealth has increased significantly in Western countries in the last 50 years (both reported and measured) levels of happiness have not. As such, it is no longer considered appropriate to measure the success of societies solely in economic terms. On the basis of this new economic theory, defined in his text *Happiness: Lessons from a New Science* (2005), a plethora of policy statements and research initiatives have emerged. In a 2003 lecture Layard stated:

> GDP is a hopeless measure of welfare. For since the War that measure has shot up by leaps and bounds, while the happiness of the population has stagnated. To understand how the economy actually affects our well-being, we have to use psychology as well as economics. Fortunately psychology is now moving rapidly in the right direction and I hope economics will follow. (Layard, 2003a, p. 4)

The field of positive psychology in particular illustrates the approach that Layard praises here. One of its leading proponents is Martin Goleman, author of *Emotional Intelligence* (1995) and various subsequent texts that seek to apply the psychological profiling and techniques that he and his colleagues have developed to the professional (*Working with Emotional Intelligence*, 1998) and personal aspects of our lives (*Social Intelligence: the New Science of Social Relationships*, 2006). Another leading figure in the field of positive psychology is Martin Seligman, who describes the field thus:

> The aim of positive psychology is to catalyze a change in psychology from a preoccupation only with repairing the worst things in life to also rebuilding the best qualities in life. . .The field of positive psychology at the subjective level is about positive subjective experience: well-being and satisfaction (past); flow, joy, the sensual pleasures, and happiness (present); and constructive cognitions about the future – optimism, hope and faith. At the individual level it is about positive personal traits – the capacity for love and vocation, courage, interpersonal skill, aesthetic sensibility, perseverance, forgiveness, originality, future-mindedness, high talent and wisdom. At the group level it is about the civic virtues and the institutions that move individuals towards better citizenship: responsibility, nurturance, altruism, civility, moderation, tolerance and work ethic. . .. (Seligman in Snyder & Lopez, 2002, p. 3)

The approaches advocated by Goleman, Seligman and others seek to make psychological knowledge and techniques available to all, for the betterment of quality of life at the level of the individual, the interpersonal and the social. This illustrates what Nikolas Rose has termed the "generosity of expertise"; the lending

of professional "vocabularies of explanation, procedures of judgment and techniques of remediation 'freely' to others. . .on the condition that [we] act a bit like experts. . .the expertise of subjectivity has proliferated at a 'molecular' level" (Rose, 1999, p. 92). Layard's findings have led to proposals to increase the availability of mental health services such as counselling and therapy in the UK, and there is a move across the EU to reduce the incidence of mental health problems, but the need for psychological intervention is no longer predominantly understood as relating to the diagnosis of a defect or pathology in the individual, but is made available to us all.

In the sections that follow I will illustrate how this academic knowledge has been adopted in to evidence-based policy-making at the global, European and national levels and how the discourses and technologies to which these ideas are related contribute to the construction of "active citizenship", and thereby a particular form of subjectivity.

8.3 Beyond GDP

The evidence presented by the new science of happiness has quickly been adopted as a new truth about how societies should best be governed in the face of new challenges (for example, environmental sustainability) and persistent problems (for example, unequal distribution of wealth). This is evident in the formation of the multi-agency International Initiative comprising the European Commission, the European Parliament, the OECD, the Club of Rome and the World Wildlife Fund, entitled "Measuring progress, true wealth and the well-being of nations". In its most recent publication it is stated that:

> The overall aim is to develop more inclusive indicators that provide a more reliable knowledge base for better public debate and policy-making. The Commission intends to cooperate with stakeholders and partners to develop indicators that are internationally recognised and implemented. (CEC, 2009, p. 2)

As the lead agency the European Commission is central to efforts to develop such indicators, the rationale for which references the discourse of inclusion and of public (stakeholder) engagement and accountability as a key aim of evidence-based policy-making. In the presentations made to the conferences held by the International Initiative there is a common refrain in the papers; they each echo the address cited above given by Commissioner Almunia that GDP is no longer enough.

The European Commission uses various tools with which to gather data, for example, the Eurobarometer Surveys are frequent polls of public opinion; the European Social Survey and the European Quality of Life Survey both provide a regular overview of trends, and the Eurostat office collates and standardises data from the national statistical organisations of the member states. The second (and most recent) European Quality of Life Survey (2009), published by the European Foundation for the Improvement of Living and Working Conditions (Eurofound),

illustrates how the concern for happiness and well-being appears. In the introduction to the publication of results, explicit reference is made to the work of Richard Layard. It reads:

> The European Quality of Life Survey (EQLS) has developed as a tool to inform the social policy debate in Europe. There is clearly a need to establish the main social challenges facing Europeans in Member States and to better understand the situation of different socioeconomic groups. However, it is also evident that data on living conditions should be complemented by information from another perspective, "where people's feelings are treated as paramount" (Layard, 2006). This viewpoint is critical to assessing the European social model, which emphasises values of trust and tolerance, solidarity and justice. (Eurofound, 2009, p. 1)

The concern for happiness and well-being is not only evident in the practice of measuring these aspects, but seen to be served in the very gathering and presentation of statistics themselves. The gathering of data on happiness and well-being is used as evidence of Europe's commitment to the core value of social justice of which its commitment to democracy is part. As such, the communication of this data is presented as a means to build this trust between Europe and its citizens. The discourse of social justice, transparency and accountability is also evident in the language of Eurostat. In the foreword to the 2009 "pocketbook", the current Director-General of Eurostat, Walter Radermacher, describes the production of statistics as "a public good and a basis for the smooth functioning of democracy" (Eurostat, 2009, p. 7); he also makes a direct link between the production of statistics and Europe's "prosperity and commitment to solidarity and social justice" (Eurostat, 2009, p. 7). The Eurostat pocketbook is one of a number of publications this office produces and is intended for a more general audience, as a way of engaging citizens in knowledge about Europe, and demonstrating Europe's commitment to them. It is not only in the gathering of data on which Europe evidences its commitment to its core values but through their presentation to the public. The questions asked illustrate how citizens are addressed in terms of their happiness, well-being and the associated life-satisfaction and how these aspects correlate to other concerns.

The European Quality of Life Survey begins with some factual questions about the household and employment: How many others live in the household with you, how old are they and what do they do? What do or did you do if anything, in what sector and for how many hours a week? The questions then shift slightly to require answers based on opinion or perception: How likely do you think it is that you might lose your job in the next 6 months? Or questions on working conditions or what might be termed work–life balance to be answered according to a scale between "Strongly Agree" through "Strongly Disagree" to "Don't Know" or between "Several Times a Week" and "Never". The questions then return to the factual: What is the occupation of the main wage earner in the household? How many rooms are there in the accommodation in which you live? Throughout the survey, questions shift from the factual, as just shown, to those relating to opinions or perceptions. Here are some examples:

Q. 23 Generally speaking, would you say that most people can be trusted, or that you can't be too careful in dealing with people? Please tell me on a scale of 1 to 10, where 1 means that you can't be too careful and 10 means that most people can be trusted.

Q. 25 In all countries there sometimes exists tension between social groups. In your opinion, how much tension is there between each of the following groups in this country?
Poor and rich people
Management and workers
Men and women
Old people and young people
Different racial and ethnic groups
Different religious groups

Other questions relate more specifically to the feelings or perceptions of the individual:

Q. 28 Please tell me whether you strongly agree, agree, neither agree or disagree, disagree or strongly disagree with each statement:

I am optimistic about the future
On the whole my life is close to how I would like it to be
In order to get ahead nowadays you are forced to do things that are not correct
I feel left out of society
Life has become so complicated today that I almost can't find my way
I don't feel the value of what I do is recognised by others
Some people look down on me because of my job situation or income

Q. 29 All things considered, how satisfied would you say you are with your life these days? Please tell me on a scale of 1 to 10, where 1 means very dissatisfied and 10 means very satisfied.

Questions pertaining to happiness, well-being and life satisfaction refer not only then to the material circumstances of one's life but also to the individual's assessment of their ability to deal with these circumstances. There are similarities and some cross-over between the questions asked in the European Quality of Life Survey and the European Social Survey (http://www.europeansocialsurvey.org/index.php?option=com_content&view=article&id=63&Itemid=356). The latter, however, begins with questions about the media. Of television, radio and newspapers, respondents are asked how much time per week they spend using each. This is followed in each case by a question asking how much of this time is spent watching/listening to/reading about news, politics and current affairs (A1–A6). Then there is a switch, and the question of trust appears again: "Generally speaking, would you say that most people can be trusted, or that you can't be too careful in dealing with people? Please tell me on a score of 0 to 10, where 0 means you can't be too careful and 10 means that most people can be trusted." (A8) This is then echoed by similar questions: Would most people try to take advantage of you if they got the chance, or would they try to be fair? (A9) Most of the time would people try to be helpful, or are they mostly looking out for themselves? (A10) The questions return again to politics, to interest, understanding, and ease of decision-making, and to trust in political and legal institutions at the national and European levels (B1–10). Questions then

turn to satisfaction, with your life as a whole, with how the national government is doing its job, the way democracy works, and the state of education and health provision (B24–29). Some later questions relate clearly to policy issues: for example, in asking the respondent to locate their position on the scale between "European unification has gone too far" and "Unification should go further" (B34). Or asking for their opinion, between "Strongly agree" through "Strongly disagree" to "Don't know", on questions such as:

> *The government should take measures to reduce differences in income levels.*
> *Gay men and lesbians should be free to live their own life as they wish.*
> *Political parties that wish to overthrow democracy should be banned.*
> *Modern science can be relied on to solve our environmental problems.* (B30–33)

Questions are then asked regarding immigration, for example, its economic and cultural impact, before turning to "questions about you and your life". The section begins:

> *Taking all things together, how happy would you say you are?* (C1)

These questions continue by asking how often you socialise (C2) and how this compares to others (C4), before turning to questions about crime. After asking whether the respondent has been a victim of burglary or assault in the last 5 years, the questions are concerned with how the respondent feels about crime: How safe do you or would you feel about walking alone in this area after dark? Do you worry about your home being burgled? And does this worry have an affect on your quality of life? (C5–C8).

The selection of questions I have provided here suggests that the gathering and use of European public opinion operates in a particular way in relation to the concern for happiness and well-being. For example, levels of crime, the likelihood of becoming unemployed or the lack of trust are not necessarily the issues of greatest concern. Instead, it is whether there is fear of or perception of these, and further whether, on the whole, such a perception impedes the individual's perceived quality of life or happiness. The aim in the governmental concern with happiness and well-being is not to achieve an ultimate peak of happiness for its own sake. "Happiness" and "well-being" are composite indicators, operationalised concepts constituted by a number of quantified factors.

The concern overall is how these indicators of societal progress can be improved, and education and training are seen as central to this. As the EU's Strategic Framework for Education and Training 2020 states,

> In the period up to 2020, the primary goal of European cooperation should be to support the further development of education and training systems in the Member States which are aimed at ensuring:
>
> (a) the personal, social and professional fulfilment of all citizens;
> (b) sustainable economic prosperity and employability, whilst promoting democratic values, social cohesion, active citizenship, and intercultural dialogue. (OJEU, 2009, p. 2)

As part of the evidence for happiness and well-being, "active citizenship", as stated in (b) above, requires measures and benchmarks to verify its existence. Such measures have been developed by the Centre for Research on Lifelong Learning (CRELL) at the Joint Research Centre of the European Commission. Active citizenship relates to a particular concern with participation as a central facet of democratic citizenship. Data from the European Social Survey discussed above have been used to develop the Active Citizenship Composite Indicator. The CRELL report on the development of the indicators reads:

> The research project on "Active Citizenship for Democracy," coordinated by the European Commission's Centre for Research on Lifelong Learning (CRELL), has produced the following definition of "Active Citizenship for Democracy". (Hoskins, 2006)
>
> *Participation in civil society, community and/or political life, characterised by mutual respect and non-violence and in accordance with human rights and democracy.*
>
> Active citizenship is partially overlapping with the concept of social values concentrating its interest mostly at meso- and micro-level. Thus, active citizenship is understood in the very broadest sense of the word "participation" and is not restricted to the political dimension. It ranges from cultural and political to environmental activities, on local, regional, national, European and international levels. It includes new and less conventional forms of active citizenship, such as one-off issue politics and responsible consumption, as well as the more traditional forms of voting and membership in parties and NGOs. The limits of active citizenship are set by ethical boundaries. People's activities should support the community and should not contravene principles of human rights and the rule of law. Participation in extremist groups that promote intolerance and violence should therefore not be included in this definition of active citizenship. (Hoskins et al., 2006, p. 11)

The definition developed is derived from social capital theory (p. 9) but in order to build the composite active indicator it needed to be operationalised, that is, made measurable.

> Towards this end we identified measurable and distinctive elements in the definition of active citizenship, which we designated "dimensions of active citizenship." The dimensions are: participation in Political Life, Civil Society, Community Life and the Values needed for active citizenship (recognition of the importance of human rights, democracy and intercultural understanding). (p. 11)

"Political life" refers to involvement with political parties in the form of participation, membership, donation, volunteering or working, and to voting in national or European parliamentary elections. "Civil society" refers to political non-governmental participation, with sub-indicators relating to "*protest*, *human rights organisations*, *environmental organisations* and *trade union organisations*... *Protest* includes activities such as signing a petition, taking part in a demonstration, boycotting products and ethical consumption" (p. 12). "Community life" refers to "activities that are less overtly political and more orientated towards the community – 'community-minded' or 'community-spirited' activities" (p. 12). As with the previous two categories, degrees of participation are gauged in relation to "questions of participation, volunteering, membership and donating money" (p. 12).

The development of composite indicators for active citizenship illustrates the ways in which we are asked to account for ourselves in terms of our citizenship, the

data according to which governments and institutions evidence their accountability to the citizen-consumer, and the means through which universally applicable measures are sought according to which comparisons can be made and benchmarks set at a global level. Despite the possibility of global indicators for active citizenship and well-being, the ways in which statistics are used by various agencies create distinct comparable entities within that. The presentation of statistics on Europe through reports and graphics makes Europe tangible to its citizens as well as its allies and competitors, and further refines the ways in which Europe will seek to make itself accountable and visible to its citizens in order to encourage active citizenship, within the prescribed parameters.

The understanding of the political presented in the rationale for the indicators of active citizenship indicates a broad understanding of the ways in which individuals engage with politics today, referring to volunteering and signing petitions, for example, as well as party membership and voting. Biesta (2009) argues, however, that this presents a depoliticised form of citizenship as it is closely allied with an understanding of citizenship in terms of competencies and individual benefit rather than collective concern or action. I will turn now to consider how the discourses of happiness and well-being are becoming apparent in the school curriculum, which illustrates how such competencies for active citizenship are encouraged. The example given below of an aspect of the UK school curriculum derived from the ideas of Daniel Goleman illustrates how the concern for well-being and self-improvement in terms of which the adult as a privatised citizen is addressed is applied to the children and staff in the school. In light of the way in which "active citizenship" is operationalised and measured the ways in which citizens (children and adult staff alike) are asked to express themselves can be seen to produce a particular form of subjectivity through a particular use of language and forms of conduct that prescribe ways of relating to ourselves and to others.

8.4 "Learning to Be Together"; "Keep on Learning"; "Learning About Me"

The three statements above are three "themes" taken from the recently introduced Social and Emotional Aspects of Learning curriculum in England and Wales. The introduction of the SEAL programme is explained in relation to the weight of new research evidence, and the work of Daniel Goleman is directly cited as part of this. It reads:

> Goleman (1996) popularised the term "emotional intelligence" in his book of the same name and made the case that emotional and social abilities are more influential than conventional intelligence for all kinds of personal, career and school success. Since this book appeared in the mid-1990s, work has developed at an extraordinary pace, in psychology, neuroscience, education and other disciplines. This work has demonstrated, from a variety of perspectives, that social and emotional skills are at the heart of positive human development, effective social groups and societies, and effective education. (DCSF; http://nationalstrategies.standards.dcsf.gov.uk/node/66405?uc=force_uj)

In line with the understanding of positive psychology cited earlier, the principles of SEAL are applied to "all pupils and staff, not just those pupils with identified social, emotional and behavioural difficulties" (ibid.) for the benefit of learning, standards, inclusion, positive behaviour and stress management. The emphasis on learning and managing in particular is evident in the summary headed "What role does SEAL play in learning?":

Social and emotional skills underpin effective learning by helping all pupils to do the following.

- Learn to manage their impulses, helping them settle quickly, concentrate and not disrupt others.
- Build warm relationships, which help them to care what others (e.g. staff and peers) think and to respond positively to them.
- Manage strong and uncomfortable emotions such as anger and frustration, and become more resilient, which helps them rise to the challenges of the learning process and stick at it if things get tough.
- Learn to feel good about themselves, which reduces the likelihood of disruptive behaviour and increases capacity for independent learning.
- Manage anxiety and stress, including around tests and examinations.
- Learn to empathise, for example, with other pupils' desire to learn, which helps them contribute to a positive learning environment.
- Reflect on longer term goals, which help them see the point of learning, raise their aspirations and become more able to resist negative pressure from others.
- Feel optimistic about themselves and their ability to learn, which improves their motivation to work hard and attend regularly.

(DCSF: http://nationalstrategies.standards.dcsf.gov.uk/node/66414?uc=force_uj)

The outline shows us who the individual (child/adult/citizen) is asked to be: this refers not only to there being desirable qualities for individuals to have but that the development of these becomes an explicit educational practice and object of individual concern and responsibility. Not only do I need to feel empathy, feel good about myself and feel optimistic, I must adopt a vocabulary in which I can explicitly demonstrate that this is the case. We can see this as part of the responsibilising of the child/adult/citizen as an autonomous individual, to instil the required attitudinal orientation for citizenship as an adaptable lifelong learner. The vocabularies of expertise to which Nikolas Rose referred are offered to enable us to label our feelings and emotions, to give voice to the qualitative aspects of our lives, and therefore to be able to chart, measure, compare and improve them alongside other variables.

The SEAL programme also provides an example of and extends the understanding of the school as an environment (see also Simons & Masschelein, 2008). Learning is no longer restricted to the school but is to be pursued across all aspects of our lives; the school provides one such environment, a term that invokes the need for sensitivity to change and the possibility of adaptation characteristic of the discourse of entrepreneurialism in the knowledge economy. The school environment is described within the SEAL programme as follows:

> A useful way to think about a school's culture and environment is to divide it into three aspects – the learning climate, the social climate and the physical climate. The ongoing challenge for schools will be to establish, develop and sustain the development of social and emotional skills work systematically across the whole school and create and maintain a supportive environment, culture and ethos to underpin the work. (DCSF; http://nationalstrategies.standards.dcsf.gov.uk/node/66349?uc=force_uj)

The idea of the climate suggests something subject to fluctuation, something to which we must be able to adapt, and can be seen in relation to recent wider social and political discourses of environmental endangerment and ethical consumerism: the school, society and the environment must be adapted to, but are also something that we must protect, that we are each individually responsible for as citizens.

8.5 Conclusion

I have sought to illustrate here interrelated examples of a current object of governmental concern as a way of exploring the way in which the European citizen is understood and addressed. The use of statistics in the construction of European citizenship is seen here in the context of a shift from statistics as a matter of governmental concern to an interrelationship between the governmental, the academic and the private as sources of knowledge, and providers of expertise made available to the citizen.

The examples I have provided suggest a particular ethics of citizenship in this context of which statistics, data and harmonised numerical indicators form a part. Numbers include us all; they create Europe while overcoming the historical, political, cultural and linguistic barriers that thwart a particular idea of a United Europe. The role of data in constituting this ethics is evident at a number of levels. For example, in addition to the presentation and making accessible of statistical data on Europe by the European Commission, such documents are accompanied by the expression of the relationship between data and democracy, solidarity and social justice. Further, the ethic of citizenship in relation to one's self – autonomy, self-determination, learning – is expressed in the desire to use tools and indicators, personally or within a formal educational or professional environment, with which to gauge one's self in relation to a norm and to measure one's progress. It is not sufficient to be concerned with these aspects of one's life however; one must seek to make this explicit.

This requires the adoption of a particular vocabulary (for example, as in the SEAL programme, learning to label one's emotions appropriately and relating these to the physical feelings that indicate them). Whether one is, for example, responding to a survey on happiness and well-being, telling a classmate, friend, partner or colleague, "When you said... you made me feel...", or reporting feelings and plans for self-improvement in a journal or blog (an increasingly formal part of the requirements for the reflective practitioner as well as a personal pursuit) we are asked to give voice to our emotions and opinions, to make them evident and explicit. Such competencies are often referred to as forms of literacy: emotional literacy

and political literacy are versions of or constitutive of "emotional intelligence" and "active citizenship". But as Nikolas Rose has indicated, a form of numeracy in relation to ourselves is also a central component for the mode of citizenship required today: "Democracy… depends on the delicate composition of relations of number and numeracy enabling a calculated and calculating government to be exercised over persons and events to be governed" (Rose, 2000, p. 232). The examples of the focus and means of gathering statistics concerned with citizenship, happiness and well-being illustrate the way in which statistics operate to enable a permanently calculable environment or system at all and between levels: our bodies, our minds, our relationships to others, our homes, our consumption activities, our political participation. In an analysis of democratic citizenship, then, it is necessary not to be concerned with the definition of citizenship or of happiness that is being used but to the ways in which we are asked to calculate and account for ourselves in terms of particular renderings of them. It is the "mundane practices of pedagogy, of accounting, of information and polling, and to the mundane knowledges and 'grey sciences' that support them" (Rose, 2000, p. 232) that are more instructive for analysing how we are made subjects today.

References

Almunia, J. (2007). *Measuring progress, true wealth and wellbeing.* Speech to Beyond GDP Conference, Brussels, 19 November 2007. Retrieved October 22, 2009, from http://www.beyond-gdp.eu/download/almunia_speech.pdf

Barry, A., Osborne, T., & Rose, N. (Eds.). (1996). *Foucault and political reason: liberalism, neo-liberalism and rationalities of government*. Chicago, IL: University of Chicago Press.

Biesta, G. (2009). What kind of citizenship for European higher education? Beyond the competent active citizen. *European Educational Research Journal, 8*(2), 146–158.

Commission of the European Communities. (2009). *GDP and beyond: Measuring progress in a changing world*. Brussels: Commission of the European Communities. Retrieved October 22, 2009, from http://www.growthintransition.eu/wp-content/uploads/beyond_gdp.pdf; http://ec.europa.eu/education/lifelong-learning-policy/doc1120_en.htm

DCSF (UK Government Department for Children, Schools and Families). (2009). *Why is there an interest in SEAL?* Retrieved October 29, 2009, from http://nationalstrategies.standards.dcsf.gov.uk/node/66405?uc=force_uj

DCSF (UK government Department for Children, Schools and Families). (2009). *What role does SEAL play in learning?* Retrieved October 29, 2009, from http://nationalstrategies.standards.dcsf.gov.uk/node/66414?uc=force_uj

DCSF (UK government Department for Children, Schools and Families). (2009). Retrieved October 29, 2009, from http://nationalstrategies.standards.dcsf.gov.uk/node/66349?uc=force_uj

Eurostat. (2009). *Key figures on Europe 2009* (Edition, Eurostat Pocketbooks). Retrieved October 15, 2009, from http://epp.eurostat.ec.europa.eu/portal/page/portal/product_details/publication?p_product_code=KS-EI-08-001

Eurofound. (2009). *Second European quality of life survey overview*. Luxembourg: Office of official publications of the European communities. Retrieved October 15, 2009, from http://www.eurofound.europa.eu/

Foucault, M. (1991a). Governmentality. In G. Burchell, C. Gordon, & P. Miller (Eds.), *The Foucault effect: Studies in governmentality* (pp. 87–104). Chicago, IL: University of Chicago Press.

Foucault, M. (1991b). *Discipline and punish*. London: Penguin.
Foucault, M. (2005). *Society must be defended*. London: Penguin.
Goleman, D. (1996). *Emotional intelligence*. New York: Bantam.
Hoskins, B., Jesinghaus, J., Mascherini, M., Munda, G., Nardo, M., Saisana, et al. (2006). *Measuring active citizenship in Europe*. Luxembourg: Office for official publications of the European communities. Retrieved February 3, 2010, from http://crell.jrc.ec.europa.eu/
Layard, R. (2003). *Lionel Robbins memorial lectures. Happiness? Has social science a clue. Lecture 1: What is happiness? Are we getting happier?* London school of economics, 3 March 2003. Retrieved October 15, 2009, from http://cep.lse.ac.uk/events/lectures/layard/RL030303.pdf
Layard, R. (2006). *Happiness: Lessons from a new science*. London: Penguin.
OJEU. (2009). *Council conclusions of 12 May 2009 on a strategic framework for European cooperation in education and training* ("ET 2020"). Retrieved January 13, 2010, from http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:C:2009:119:0002:0010:EN:PDF
Rose, N. (2000). *Powers of freedom: Reframing political thought*. Cambridge, IL: Cambridge University Press.
Seligman, M. (2002). Positive psychology, positive prevention, positive therapy. In C. Snyder & S. Lopez (Eds.), *Handbook of Positive Psychology* (pp. 3–12). Oxford: Oxford University Press.
Shore, C. (2000). *Building Europe: The cultural politics of European integration*. Abingdon: Routledge.
Simons, M., & Masschelein, J. (2008). The governmentalization of learning and the assemblage of a learning apparatus. *Educational Theory, 58*, 391–415.

Chapter 9
The Persuasive Power of Figures and the Aesthetics of the Dirty Backyards of Statistics in Educational Research

Ulrike Stadler-Altmann and Edwin Keiner

This chapter focuses upon the varying contexts empirical educational research knowledge is embedded in and how this knowledge becomes subject to processes of de- and re-contextualisations according to expectations of the respective social groups it refer to. On the one side, there is an aesthetics of representation of educational research knowledge and of the rhetorically persuasive power of figures and graphic accounts, an aesthetics of providing answers. On the other side, there is an aesthetics of deconstructive scepticism and of epistemological relativism, an aesthetics of raising questions. The combination of both aesthetic forms and the provocation of both giving answers and raising questions in a formalised educational research presentation are related to two further aspects of aesthetics the chapter will focus upon. We will analyse a more rhetorical and a more epistemological focusing on two areas: (a) (oral) presentations of educational research projects (from a more microanalytical perspective) and (b) (written) publications addressing the knowledge society or educational research outcomes from large-scale assessment studies (from a more macroanalytical perspective).

It is important to note that the chapter will not address methodology or theory formation of educational research. These methodological debates were suffused with normative ideas, convictions and options. We do not start from such a perspective, but try to compare aesthetic forms of literary and research presentations. The chapter focuses upon (oral and written) presentations of educational research and the texture and function figures and statistics play within the rhetorical form of an exposition. Thus, it contributes to finding rules of a grammar of presentation within educational research.

U. Stadler-Altmann (✉)
Friedrich-Alexander-University Erlangen-Nürnberg, Erlangen-Nürnberg, Germany

P. Smeyers, M. Depaepe (eds.), *Educational Research: The Ethics and Aesthetics of Statistics*, Educational Research 5, DOI 10.1007/978-90-481-9873-3_9,

9.1 Oral Presentation of Empirical Studies – Contextual Framework

From a detached observer's point of view (oral) presentations of educational empirical research findings can sometimes be perceived as a kind of 'lyric poetry'. The formal structures of such presentations often resemble surreal or dadaist poems showing a perseverative form: problem – theory – research design – method used – findings – discussion – conclusion. Like a poem, there is the structure of intro, recall, and conclusion in an oral presentation of educational empirical research. Like the use of language in a dadaist poem, we find that the same words and sentences reoccur during oral presentations of educational empirical research. Let us first look at the literary aspect of the comparison. New and surprising combinations of words and sentences are characteristic features of poetry. Poets use everyday language to show our reality in another light. When they collect everyday words and put them into unexpected contexts the audience suddenly sees what is extraordinary about ordinary life. This is a well-known technique of constructing poems which can surprise us and introduce new insight.

The dadaist poems 'Anna Blume' and the 'Ur-Sonate', both written by the German poet Kurt Schwitters, illustrate this approach.

Kurt Schwitters: Eve Blossom (original 1919)
(Kurt Schwitters' own translation of 'An Anna Blume')

Oh thou, beloved of my twenty-seven senses, I love thine! Thou thee thee thine, I thine,
thou mine, we?
That (by the way) is beside the point!
Who art thou, uncounted woman, Thou art, art thou?
People say, thou werst,
Let them say, they don't know what they are talking about.
Thou wearest thine hat on thy feet, and wanderest on thine hands,
On thine hands thou wanderest
Hallo, thy red dress, sawn into white folds,
Red I love Eve Blossom, red I love thine,
Thou thee thee thine, I thine, thou mine, we?
That (by the way) belongs to the cold glow!
Eve Blossom, red Eve Blossom what do people say?
PRIZE QUESTION: 1. Eve Blossom is red,
2. Eve Blossom has wheels
3. what colour are the wheels?
Blue is the colour of your yellow hair
Red is the whirl of your green wheels,
Thou simple maiden in everyday dress,
Thou small green animal,
I love thine!
Thou thee thee thine, I thine, thou mine, we?
That (by the way) belongs to the glowing brazier!
Eve Blossom,eve,
E - V - E,
E easy, V victory, E easy,
I trickle your name.
Your name drops like soft tallow.

Do you know it, Eve?
Do you already know it?
One can also read you from the back
And you, you most glorious of all,
You are from the back as from the front,
E-V-E.
Easy victory.
Tallow trickles to stroke over my back
Eve Blossom,
Thou drippy animal,
I
Love
Thine!
I love you!!!!

When we look at a scholarly presentation of educational research results, the new combination of data in educational research can produce new meanings that derive from theoretical considerations. Even new, surprising and sometimes magic knowledge emerges from this combination of data, theories and the ordinary (research) life as a practice of its own right. Hence, in this sense it may be possible to compare structures of poems with structures of oral presentations of empirical studies seeing them as different forms of aesthetics delivering different meanings in respect to different contexts. Both the poet and the author of research presentations try to make sure that the scholarly audience has a clear understanding of their meanings. However, they use everyday language in different ways. Statistical understanding is both a linguistic and a conceptual matter (see Vergnaud, 1998, explanations for mathematics). Therefore, we have to accept, that even oral presentations of empirical social studies serve two points of reference: in order to understand statistics, one not only has to be able to identify relationships between statistical or mathematical symbols, but also to identify these symbols' relationships to natural everyday language (see Stadler, 2004). Finally, one has to connect this relationship to the relationships between the empirical study and questions of 'real life'.

The symbolic language of mathematical and statistical concepts is based on a formal definition that is independent from spoken language. Since these mathematical and statistical languages are difficult to teach and to understand within their symbolic format, spoken language is used to informally define statistical concepts. Within these informal definitions, statistical symbols are replaced by proper words from the spoken language. Following this model, statistics has become independent from language, yet its informal definitions are also an integral part of this language. For example, understanding of statistics is based on the distinction between samples and populations. On the one hand, some statistical concepts clearly refer to population quantities (such as expected value). On the other hand, however, some of these concepts (such as standard deviation) need a specified reference (population or sample) to distinguish the precise statistical definition from the ambiguous everyday language use of such concepts. As an example the following table presents some symbols used in statistics and in poems (which also could be transformed in everyday language).

Table 9.1 Analogies between statistical symbols and the poem 'Anna Blume'

Empirical study	Poem (English)	Poem (German)
R	Oh, thou	O
* ** ***	!!!!	- -
A		! -
B	E easy	dich dir
X^2 (df ..)	V victory	Ich dir
NFI	E easy	du mir
CFI	E – V – E	'a – n – n – a'
RMSEA	E-V-E	a-n-n-a

If the scholarly audience listens to an educational research presentation and looks at symbols like those in the first column of Table 9.1, they feel that they understand the relevance of these symbols. It is not necessary for the presenter to provide a more detailed explanation. For example, * (one asterisk) means 'significant', ** (two asterisks) 'highly significant'. If we look at the literary dimension we also find some symbols in poetic (as well as in the everyday) language that do not require an explanation. For example, '!!!!' means 'very important', and 'E-V-E' has at least a double, but clear meaning: it accentuates the name 'Eve' or it serves as an acronym for 'Easy Victory Easy'. In this respect one could in fact assume a comparability of the symbols' meanings in statistical presentations and in poems. They share the form, but they differ regarding the meaning that derives from respective contexts.

However, are the sense and meaning of these symbols always the same regardless of what the author wants to convey?

Researchers have explored various areas regarding the understanding of statistical concepts and symbols (Falk, 1986; Haas, Pattuelli, & Brown, 2003) in order to see how technology can help to understand, integrate, and apply fundamental statistical concepts (Chance Garfield, & delMas, 2000). Pimm (1987) relates understanding of symbols to the use of everyday words with particular mathematical meaning. This relationship is also relevant in the case of statistics. He also considers that analogies and metaphors, that house rather different meanings in everyday language, are very important for the construction of precise mathematical and statistical meaning.

9.1.1 Aesthetics in Oral Presentation of Empirical Studies

The analysis of a scholarly oral presentation could be done in a similar way as an analysis of a poem. Using the standards of literary criticism, we can interpret the different parts of the presentation as stanzas. In a scholarly oral presentation particular importance is usually attached to graphs, figures and numbers indicating research results. The presumed evidence, significance and validity of the figures take centre stage in order to be suggestive of objectivity and clarity. Like the author of a poem the author of an empirical study uses elements of enchantment to show how new results disenchanting the former vision of the world. The 'magic' in empirical

studies is conjured through the use of statistics. Even if the scholarly audience knows a lot about the statistical background the author constructs him-/herself as an expert.

The difficulties of carrying out and understanding statistical presentations are due to the complex structure of statistics. According to Sfard (1991), mathematical entities and concepts have a dual nature. Referring to the differences between structural and object conception Sfard says,

> Whereas the structural conception is static, instantaneous and integrative, the operational is dynamic, sequential and detailed . . . The former [object] is more abstract, more integrated and less detailed than the latter. (p. 4)

The same could be said about a sophisticated, structural poem. A structural poem is static, instantaneous, integrative, more abstract and less detailed. The sense of a poem is sometimes locked into a few words just as the sense of a statistical finding is also locked into a few words or symbols. Just like the 'Medium is the Message' (McLuhan) the structure is the meaning. Therefore, every poem can be reconstructed as regards both its structural form and the operational processes of structuring and re-configuring words. Similarly, every empirical study can be read in regards to both its epistemological structure and an operational process of trial and error pertaining to the juggling of words and symbols. In this respect an oral research presentation as well as a poem can be perceived from structural and operational points of view. Every word in a poem could have different meanings and different modes of ignorance, and every element of data in an empirical study could have different answers for different questions. The particular meaning emerges in situ while words and symbols are used and played with. Of course, this assumes that the general meaning hovers in the background as a commonly shared structure.

Every educational empirical oral presentation usually starts with a little intro: name of the author(s), goals of the research and theme of the presentation. Let us return to the poem 'Eve Blossom' ['Anna Blume'] where Kurt Schwitters provides the title within which reside the goal and theme of his subject. Then he repeats his insight or result four times: 'I love thine' ['ich liebe dir' (3, 12, 21, 32)]. Similarly, in a presentation of educational research, the goals and the results are repeated at least four times: in the intro, at the end of the theory-chapter, in the result-chapter and in the conclusion (Table 9.2).

If we look only at the construction of presentations and poems we see some parallels: An (oral) presentation of educational statistical research is based on the

Table 9.2 Comparison between elements of poems and empirical studies

	Poem	Empirical study
1	Stanza	Chapters
2	Words	Data
3	Symbols	Statistical symbols
4	Magic	Statistics
5	Recite	Oral presentation
6	Understanding	Judgement
7	New view	New questions

collected data and theories that provide a pool of information and meaning. This serves as the structural background for the presentation, which to some extent can be perceived as a scholarly performance to satisfy the audience. Poems are based on everyday language and a pool of – in our example German – words with ambiguous and different meanings and practices, which can be perceived as a literary composition to enchant the reader. The researcher, looking over the results and imagining the possible audience, selects those words and graphs which illustrate his/her research questions and answers. Like the researcher, the poet selects those words and symbols that she/he assumes would help the audience to understand her/his meaning.

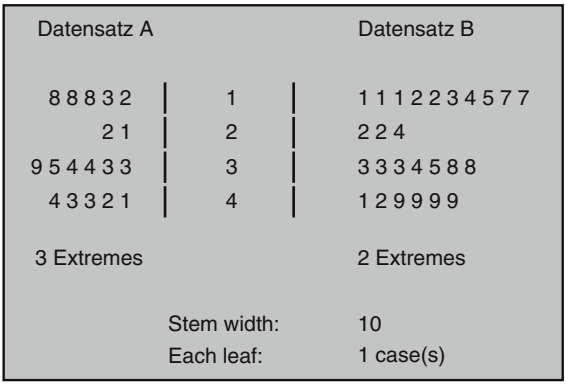
Kurt Schwitters uses two types of symbols: words and graphical presentation (in his recitation of 'Anna Blume', which was recorded since 1922, he also uses different emphases). In an oral presentation of educational statistical research we often use mathematical symbols, which immediately become transferred into everyday language and take on different meanings. And, similar to the expectation of Kurt Schwitters, we presuppose that the scholarly audience knows the particular meaning in our particular context. However, both, the poet and the author of research presentations often do not chart the reaction of the audience. Therefore, they can never be sure about the degree and mode of understanding.

The magic in the poem 'Eve Blossom' ['Anna Blume'] is the arrangement of the words, the emphasis and the breaks between the words or the stanzas. Even the new and surprising combination of words like the grammatically incorrect phrase 'I love thine' ['ich liebe dir' (3)] or the illogical phrase 'Blue is the colour of your yellow hair' ['Blau ist die Farbe deines gelben Haares' (19)] make us imagine the magic. The magic in educational research sometimes issues from statistics. New combinations and new methods seem to serve as 'magic' for the scholarly audience. In this case the author produces her-/himself as an expert: an expert similar to the author of a poem. Nobody would ask Kurt Schwitters to use grammar in a correct way and, during an oral presentation few people would query the use of the statistical methods (though they might do this after the presentation). Knowledge about statistical methods is taken for granted in the scientific community. Most questions, therefore, addressed from experts to the presenter, will focus on contents and not on basic mathematical techniques and statistical procedures.

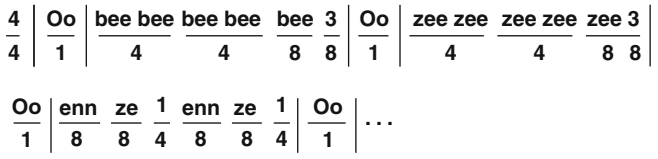
When describing the understanding of the scholarly audience of the research presentation, which is similar to the understanding of a poem, we should also focus upon the poems' performance in a live setting. Some people in the audience have an emotional response, others hear only the spoken words, and others primarily see figures and concepts. There are people who reify empirical concepts viewing them as real objects that exist outside the human brain. Indeed, most empirical or statistical presentations discuss characteristics of functions and groups just as scientists discuss the structure of physical materials. It can be said that the statistical and even mathematicians' perception of mathematical concepts is structural perception. However, in mathematical textbooks we can find definitions that uncover a different kind of perception: process-based perception. The latter is based on a dynamic imagery of mathematical concepts. Sfard (1991) says that structural perception precedes object perception. He describes the development from an operational to a

structural conception as a process of reification. In order to internalise the concept meaning, the student becomes acquainted with it by computing in single steps. These steps connect to each other in the next phase of condensation. Reification is the ability to see the concept as a whole. It is a static state where 'the concept becomes semantically unified by this abstract and purely imaginary construct' (Sfard, 1991, p. 20).

It is not only possible to compare the composition of a poem and an oral research presentation, the graphical format also reveals analogous shapes. You could read, sing or dance an empirical data set, and read, interpret or present a poem as an empirical result. This is shown in the figures below.



(Source: http://marktforschung.wikia.com/wiki/Stem-and-Leaf)



Source: www.kurt-schwitters.org/n,750017,2050046,1.html

Is this a part of statistical output? Like a stem-leaf diagram? No, it is also poetry; the latter example is Kurt Schwitters' 'Ursonate in Urlauten'.

Here we see that poets use words in new combinations and graphical ways to express and illustrate their impressions. Empirical researchers take combinations of words and figures to express and illustrate their understanding of our world and social relationships. However, to make analogy is not to say that these two things are the same, which raises the question: and the truth?

9.1.2 And the Truth?

Scientific (and thus, statistical) thinking seeks for ways in which existing understanding might be wrong, because such exceptions could illuminate the path towards a better understanding of the truth.

Some disciplines see statistics in this way. Writers within these disciplines regard statistical analysis as a gatekeeper to sound empirical research. Statistical significance is the first requirement for publication in many areas of social science. For example, social scientists discuss 'evidence-based education' which means selecting treatments based on the best scientific (usually statistically based) evidence to date. Educational statisticians often have to resist the popular impression that learning how to use statistics is synonymous with learning how to lie with data. Those who incorrectly believe that statistics is solely a branch of mathematics (and thus algorithmic), often see the use of judgement in statistics as evidence that statisticians do indeed manipulate their results. In their push to teach formulas and definitions, they may fail to emphasise the important role played by judgement. The principle guiding these judgements should be the honest search for 'truth' about the world, and the principle of seeking such 'truth' should have a central place in discussions about statistics and even in presentations. There is also a seeking for truth in poems and in the discussion of poems as a way to understand our world. But the literate audience do not assume that the poet lies, it knows something about the 'magic' of the poetic speech as a specific way to see the truth. Even statistics in social and educational research is one way to look at our world and see the truth.

Of course, one can wield the tools of statistics to mislead. But even those who repeat the quotation do not believe that the purpose of statistics is to mislead, or that there is something fundamentally dishonest about statisticians. Statistics does not lack respect because people think that statisticians are crooks. One way of showing truth in oral statistical presentations is to fix the structure of a presentation. Then, the scholarly audience finds similar structures in different research presentations and feels safe and secure – the author has used the right methods. That helps to understand the data and helps to accept the honesty of statistical use and the conclusions drawn from it.

One of the major goals of statistics is to facilitate the discovery, understanding, quantification, modelling, and communication of facts about the world. This is also a major goal in poetry: to discover the world in a new way and to communicate the experience of the author.

We probably could better understand some oral presentations of empirical studies, if we see the presentation of quantitative results as a lyric presentation. In qualitative research we could find a similar mode of presentation: Holley and Colyar (2009) demonstrate the narrative structure of qualitative research and they use a theory of narrative to deconstruct qualitative research texts. So, in our example we could use the theory of poems and lyrics to try to deconstruct the oral presentation of quantitative research presentations. Although research texts are a distinct genre as compared to fiction, the basics components of lyric structure are sometimes similar to those of research texts. In both areas the use of these components function

as play, play with symbols, which creates meaning in the process of structuring and creates structures in the process of constructing meaning. When we listen to an oral education research presentation, looking at the PowerPoint Presentation of statistical research results, we could enjoy the performance without having to comprehend anything. But, if we want to understand the research, its methodological presuppositions and its' meaning we should sometimes work as literary critics in order to find the sense behind words and symbols.

9.2 The Knowledge Society and Its Statistical Knowledge Basements

The rhetoric about the knowledge society as well as social science research on education to meet this goal presents a persuasive aesthetics and aims at both convincing the audience and presenting the authors and/or organisations as experts through 'impression management'. However, there are often deep gaps between the rhetorically claimed and visualised informative value and the complex types of knowledge the mass of information is framed and contextualised by. Therefore, the second part of this chapter, that takes a macro-level perspective, looks at OECD statistics, especially at the relationship between the figures displayed and the appendices. This combination of factors serves to inform its audience about the limits of validity (based on the strong desire for comparison). Just like in a scholarly oral speech the figures presented are deeply attractive. They are used to augment power, competition and governance not only in research contexts, but also in economics, policy and administration. Less attention, however, is paid to the appendix, the condition of possibility to form a justified opinion and a sound interpretation of the figures, in order to not distract the focus from the main goal we are challenged to meet in the future: the knowledge society. Enchantment and disenchantment, magic and enlightenment serve as two sides of the same coin.

The information or knowledge society has not, as yet, become a reality. At present we find a mainly rhetorical focus directed to this future, a future that depends on lifelong learning, competition and cooperation, constructions of the expected self, etc. Against this rhetoric we would claim that the terms: 'Knowledge' society and 'information' society – are (self-) descriptions of a society by social factions within that society. 'Knowledge' is their 'modern' ideology, a universal expectation addressed to individuals to learn, to cooperate, to compete and, finally, to feel responsible for constructing their vita without having the markers of a successful life at their disposal. Educational institutions serve as career regimes, even regimes of life courses attributing markers to their clients and collecting masses of standardised data to assure their (institutions and clients) success. This combination of universal expectations and organisational power of assessing, ascribing and labelling the level and degree to which people meet these expectations is based on large-scale administrative and research machinery. This machinery is put to the service of negotiating and agreeing upon commonalities, of defining categories and

collecting, exploiting and displaying data – a gargantuan enterprise to both crown enlightenment and finalise the pursuit of happiness and common welfare. 'Good decisions are informed decisions' is one of the great mantras of modern rationality. Social sciences and educational research are expected to significantly contribute to its realisation in the field of education, training and lifelong learning by providing reliable information, which, in close connection with educational planners, can be used for political decisions.

The intended tight linkage between educational research and educational management and policy, however, is usually based on the implicit assumption that – reversing the phrase – 'informed decisions are good decisions'. Furthermore, the assumption also implies that the quality of decisions increases with the amount of good information available.

Statistics are one of the most important means to govern this direction and to convince people about the appropriateness of means and direction. The (written) presentation and use of statistics follows a pattern that combines a more rhetorical and a more epistemological aesthetics thus intertwining policy and research (Ozga 2007; Ozga, Seddon, & Popkewitz, 2006). However, the political or scientific reference, which is necessary to judge the validity and reliability of the statistical information presented, usually remains rather implicit and unclear.

Economic, social and educational advancement as normative points of reference drive the demand of standardised and comparable research designs and so-called evidence-based educational research knowledge. At first glance, 'knowledge' seems to be used as an inclusive term covering all kinds of knowledge. However, if one takes a closer look at what is required knowledge is narrowly defined as 'useful knowledge'. Institutions and organisations that govern career and life courses, define what counts as licensed (or not licensed), as useful (or less useful) knowledge.

9.2.1 Information, Knowledge and Social Contexts

A core problem is the commonly homonymic use of the terms knowledge and information. To distinguish these terms, however, is very important, because it relativises and thus provides the means to criticise what is 'given'. In addition, it makes it possible to qualify and to assign the aesthetic epistemological and rhetorical format of a given knowledge as dependent from the social contexts it is related to.

Information, derived from the Latin verb '*informare*' in the sense of 'to give form to the mind', lacks a precisely definable content. It is, one might say, rather decontextualised knowledge. Information is a phenomenon that does not undergo scrutiny, more a word than a term, more a cipher than a number. On the other hand, knowledge can be seen as information loaded up with particular meaning. This depends on the particular context it is embedded in.

A good example of uploading and unloading, of transitions from knowledge to information and back again, is the book *Education at a Glance*, published every second year by the OECD. The very detailed appendix of this book shows the methodical conditions under which the figures presented are valid. This is not only

a question of the ISCED classification, but also the categories used, the weights added to the figures and the conversions made. The appendix shows that *Education at a Glance* is made according to the standards of social science research, at least regarding its methodological (not necessarily to its theoretical) frame. Although the OECD as well as the European Commission try to make statistics more convincing by using well-designed tables, charts and cover pages, which aestheticise the messages based on figures, the methodological appendix constrains their coverage. This shows that the explanatory range and the degree of comparability of the data are very, very limited. It raises more questions, than answers.

A short story: Some years ago I gave a seminar on methodology in educational research, and, what I did first was to read the methodological appendix of *Education at a Glance* with my university students. Following this, the students came to regard *Education at a Glance* as a worthless, meaningless collection of incomparable data. I therefore had to defend *Education at a Glance* as a weak, but nevertheless the best and most comprehensive collection of data that is currently available. In the second part of the seminar I read an article with my students about research on the degree of financial investment contributing to educational development in underdeveloped countries (indicated by the relationship between financial investment and degree of literacy). The article stated that roughly the cost of three American Pershing rockets (or the annual budget of Berlin!) would be enough to double the degree of literacy in one such country within 15 years. This article provoked a very intense and political debate on the necessity of development aid and bringing more investment to the educational systems of developing countries. All students agreed upon this goal. At the end of the session I asked the students to look at the statistical basis of the article. The statistics were mainly derived from *Education at a Glance*. It became clear that the validity of the information given depends not only on the methodological frame but also on the political intentions behind the use of information.

What did the students do? First, they subtracted, separated, unloaded and peeled off the data (information) from the methodological knowledge frame that constituted their scientific validity. They then decontextualised research knowledge, bringing it down to the level of data. Second, they uploaded and added an alternative frame to the data – e.g. social desirability, ethical principles, political usefulness or economical viability – which constituted a new subject, new knowledge. The students recontextualised the data, thereby elevating it to the level of knowledge. This shows that the pattern of data and categories the data are embedded in serve as a kind of a puzzle picture, which displays similar, but slightly different shapes depending on the frame. The frame attributes the particular meaning to the data and categories converting them to structured and meaningful knowledge. From this point of view, truth and adequacy are not necessarily justified by scientific validity. This only counts as one truth inter alia. Truth turns out to be a question of power. Those who hold power are able to define the appropriate frame and merge it with statistical data so as to sell the construct as the truth.

The variety of possible different contexts disappears and becomes superseded by a system's principles of functioning, the authority or power of a social group, a speaker, an institution, political promises or the 'Zeitgeist'. The aesthetics of

statistics then is connected to the variability and possibility of being used as puzzle picture of knowledges, and the ethics of statistics depends on the legitimacy of power, clever rhetoric and the legal possibilities of disagreement.

Against this background the distinction between information and knowledge serves as an instrument of epistemological and sociological criticism. It necessarily considers the context that constitutes knowledge as particular, specific knowledge.

Knowledge and information, therefore, can be distinguished not only epistemologically as different concepts, but also sociologically as concepts preferred by particular social groups. For example, one group is more concerned with technology and management. The other focuses upon key concepts, smart interpretations and visions based on a theoretically informed, literate background. Research on the use of knowledge shows that the former group (usually made up of middle managers), tends to prefer statistics. The latter group is primarily composed of senior managers and business consultants (Beck & Bonss, 1994).

These particular social groups attract information and produce, distribute and repeat rather undetermined, uncertain, conflicting notions of 'knowledge'. They reproduce their own educational background, their worldviews and their social positions. Their middle-class and highbrow intellectual background allows them to govern the discourse on the 'knowledge society' and to define licensed knowledge as knowledge of high and low value. Bourdieu's (1984) structuralist theory, especially elaborated in his *Distinction: A Social Critique of the Judgment of Taste*, could describe and explain this social structure whereby knowledge producers expect others to produce knowledge and keep the (hidden) criteria for quality and value in their hands. It is important to note that an ambiguous, uncertain and concealed concept of knowledge is useful to gain and exert power, because – according to organisational and management theory (Crozier & Erhard, 1979) – power can be defined as the ability to control uncertainty.

9.2.2 Education and Education Research

In certain ways educational research contributes to this pattern of defining adequate knowledge, first, by its tendency to respond to external expectations. Educational research rather breaks and redefines these expectations into the logic of methodologically distant research knowledge (Bourdieu, 1998, p. 19); it shows high resonance ability for public and political themes (Stross & Thiel, 1998). Taking up these themes in relation to the mode of traditional empirical research or in the mode of reformative reflection involves looking at two sides of the same coin. The former tries to convince via methodologically warranted truth, whereas the latter tries to convince by assuming trust in a scholarly paternalism. Second, educational research incorporates a broad range of normative orientations (goals and means, reform, tasks and challenges or the endless theory–practice relationship debate (Keiner, 2002)). It tends to educationalise social problems (Smeyers & Depaepe, 2008) even in the mode of educational research (Keiner, 2006) and in doing so participates parasitically in the imagination of political power.

This struggle concerning politics, education and power (as regards the 'right' knowledge) is also embedded in and directed towards organisational structures such as the school or university, which, according to organisational complexity theory, can be described as a loosely coupled system. In organisational sociology, Weick, March and Olsen apply this concept to education systems. They maintain that loose coupling characterises these systems.

Theory of coupling, used in computing, states very clearly, 'Coupling refers to the degree of direct knowledge that one class has of another. ... Strong coupling occurs when a dependent class contains a pointer directly to a concrete class which provides the required behavior. Loose coupling occurs when the dependent class contains a pointer only to an interface, which can then be implemented by one or many concrete classes' (Loose Coupling, Wikipedia 2010).

Seen against this background we find loose coupling in education between input and output, promises and redemption, the ceremonial level of self-description and the operative level of daily work. There are contradictions and paradoxes everywhere you look. Loose coupling shows a picture of an efficient, fluent, organic homogeneity (sometimes close to romantic images). It offers flexible, deregulated provisions for transforming various, rather arbitrary information into respective kinds of knowledge. With regard to the social dimension: selectivity and accountability of bureaucratic-hierarchical structures and standards are replaced by social mechanisms, collectively ascribed roles and invisible criteria of evaluation and assessment. Loose coupling prefers open structures that can turn into a greenhouse of the future or a 'cold' structured organisation of qualification.

Statistics, then, function as an interface to which a dependent class is able to point to with the expectation that one or many other classes will react by implementing the expected behaviour. What is at stake relates to who constructs and controls the interface and whether or not they have the ability to define the direction and structure of transforming information into licensed/accepted knowledge. There is no actor, no devil and no conspiracy. It is the socialised and socialising habitus (Bourdieu) that exerts a high degree of influence on this interface.

In the middle ages, only priests and monks were able to read and understand the subtle symbols and holy words within a differentiated system of structured meaning. People got the myths, the canon and the subordination. Nowadays the MANY are flooded by information and set free to use it, but the rules and tribes, techniques and methods to combine, assemble, select and show the licensed knowledge is available to the FEW, the modern secularised priests who promise the paradise of the knowledge society and light the path to its entrance. Their descendents get to jump the queue. The education system is a core element in the control of access to the road to paradise and the reproduction of social structures. Statistics as an interface serves to show 'facts' and to hide the function thereby creating transparent intransparency. Commonly it is said – even by large-scale assessment specialists – that input, structure and output/outcome of recent new education initiatives are hardly measurable when using statistics. So-called qualitative research is requested for this kind of pedagogy. A good indicator to investigate the relationship between statistics and power and the function of statistics as interface could be the degree to which new education

initiatives become subject to quantitative education and research. I assume that this degree is rather low.

The efficiency of the rhetoric of the knowledge society and the sweeping application of categories, numbers and figures to education systems could also indicate the high persistence of reproduction of the social strata. It could explain the boom of visions on 'new education' as a pathway for middle-class children to take the opportunity to ascend the social ladder. This trend would also allow upper-class children to escape from being statistically measured and qualitatively modelled. They would be elevated from standardised comparability to incomparable uniqueness.

9.2.3 The Dirty Backyards?

Against this background it is necessary that we do not simply defend the statistics. Rather we should focus on the relationship between statistics as a technique and the research context (theory and method). We should look not only at numbers and figures, but also at the methodological conditions of their meaningful construction and at the theoretical frames of their interpretation. And we should use the only means we are experts at: sound (so-called quantitative and qualitative) research and stringent criticism. By using the metaphor of foreground and background or of camouflage and disclosure (perhaps even enlightenment) we can direct attention to our own (and others') 'dirty backyards' – the methodological conditions of significance and validity of education research findings usually concealed in footnotes or appendices. From an epistemological point of view these left-over elements could indicate an aesthetics of the residual, the fringe, the halo, the ground noise, the no man's land of the outlaws in the outback. This aesthetics of the DIRTY represents an alternative vision to the future marked out as the knowledge society. It is close to pedagogical textures of thinking and reasoning. It takes the present world as a deficient mode of a better one, the seed of which could be found in the outback, the 'excluded' areas, and the hopeless hopes. Against this background, the DIRTY should not be taken as the remainder to be avoided, but as characteristic of good empirical research, which produces new challenging questions instead of non-credible answers. What remains is the aesthetics of construction as deconstruction and deconstruction of construction. It is the aesthetics of critique.

Annex

Kurt Schwitters

1 *An Anna Blume*
O du, Geliebte meiner siebenundzwanzig Sinne, ich
liebe dir! - Du deiner dich dir, ich dir, du mir.
- Wir?

Das gehört (beiläufig) nicht hierher.
Wer bist du, ungezähltes Frauenzimmer? Du bist
- - bist du? - Die Leute sagen, du wärest, - laß
sie sagen, sie wissen nicht, wie der Kirchturm steht.
Du trägst den Hut auf deinen Füßen und wanderst auf
die Hände, auf den Händen wanderst du.
Hallo, deine roten Kleider, in weiße Falten zersägt.
Rot liebe ich Anna Blume, rot liebe ich dir! - Du
deiner dich dir, ich dir, du mir. - Wir?
Das gehört (beiläufig) in die kalte Glut.
Rote Blume, rote Anna Blume, sie sagen die Leute?
Preisfrage: 1. Anna Blume hat ein Vogel.
2. Anna Blume ist rot.
3. Welche Farbe hat der Vogel?

Blau ist die Farbe deines gelben Haares.
Rot ist das Girren deines grünen Vogels.
Du schlichtes Mädchen im Alltagskleid, du liebes grünes
Tier, ich liebe dir! - Du deiner dich dir, ich dir, du
mir, - Wir?
Das gehört (beiläufig) in die Glutenkiste.
Anna Blume! Anna, a-n-n-a, ich träufle deinen
Namen. Dein Name tropft wie weicher Rindertalg.
Weißt du es Anna, weißt du es schon?
Man kann dich auch von hinten lesen, und du, du
Herrlichste von allen, du bist von hinten wie von vorne:
"a - n - n - a".
Rindertalg träufelt streicheln über meinen Rücken.
Anna Blume, du tropfes Tier, ich liebe dir!

References

Beck, U., & Bonss, W. (Eds.). (1994). *Weder Sozialtechnologie noch Aufklärung?* Germany: Frankfurt/M: Suhrkamp.

Bourdieu, P. (1984). *Distinction: A social critique of the judgment of taste*. London: Routledge.

Bourdieu, P. (1998). *Vom Gebrauch der Wissenschaft. Für eine klinische Soziologie des wissenschaftlichen Feldes.* Konstanz: Universitätsverlag Konstanz.

Chance, B., Garfield, J., & delMas, R. (2000, December 11–14). Developing simulation activities to improve students' statistical reasoning. *Proceedings of the International Conference on Technology in Mathematics Education (PICTME), Auckland, NZ.*

Crozier, M., & Erhard, F. (1979). *Die Zwänge kollektiven Handelns. Über Macht und Organisation*. Königstein im Ts.: Äthenäum.

Falk, R. (1986). Misconceptions of statistical significance. *Journal of Structural Learning, 9*, 83–96.

Haas, S. W., Pattuelli, M. C., & Brown, R. T. (2003). Understanding statistical concepts and terms in context: The GovStat ontology and the statistical interactive glossary. *Proceedings of the American Society for Information Sciences and Technology (ASIST) Annual Meeting, 40* (pp. 193–199). Long Beach, CA.

Holley, K. A., & Colyar, J. (2009). Rethinking texts: Narrative and the construction of qualitative research. *Educational Researcher, 38/9*, 680–687.

Keiner, E. (2002). Education between academic discipline and profession in Germany after World War II. *European Educational Research Journal, 1*, 83–98.

Keiner, E. (2006). The science of education – Disciplinary knowledge on non-knowledge/ignorance? In P. Smeyers & M. Depaepe, M. (Eds.), *Educational research: Why 'what works' doesn't work* (pp. 171–186). Dordrecht: Springer.

Loose Coupling (2010). Retreived March 13, 2010, from http://en.wikipedia.org/wiki/Loose_coupling

Ozga, J. (2007). Knowledge and policy: Research and knowledge transfer. *Critical Studies in Education, 48*(1), 63–78.

Ozga, J., Seddon, T., & Popkewitz, T. S. (Eds.). (2006). *Education research and policy: Steering the knowledge-based economy* (World Yearbook of Education 2006). New York: Routledge.

Pimm, D. (1987). *Speaking mathematically*. New York: Routledge and Kegan Paul.

Sfard, A. (1991). On the dual nature of mathematical conceptions: reflections on processes and objects as different sides of the same coin. *Educational Studies in Mathematics, 22*, 1–36.

Smeyers, P., & Depaepe, M. (Eds.). (2008). *Educational research: The educationalization of social problems*. Dordrecht: Springer.

Stadler, E. (2004). *Language and understanding of mathematical concepts*. Paper presented at the Nordic pre-conference to ICME 10 at Växjö University, pp. 1–6.

Stross, A. M., & Thiel, F. (1998). Themenkonjunkturen und Disziplinentwicklung. Eine Untersuchung erziehungswissenschaftlicher Zeitschriftenbeiträge 1987–1994. In: A. M. Stross & F. Thiel. (Eds.), *Erziehungswissenschaft, Nachbardisziplinen und Öffentlichkeit. Themenfelder und Themenrezeption der allgemeinen Pädagogik in den achtziger und neunziger Jahren* (pp. 9–32). Weinheim: Beltz.

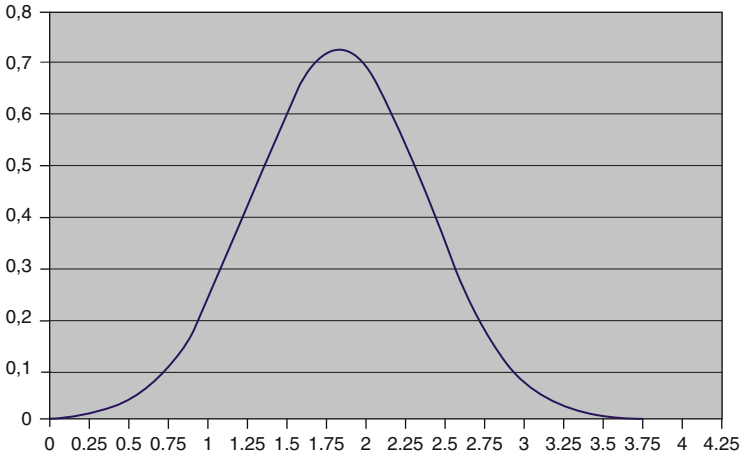
Vergnaud, G. (1998). Torwards a cognitive theory of practice. In A. Sierpinska & J. Kilpatrick, J. (Eds.), *Mathematics education as a research domain: A search for identity* (pp. 227–240). Dordrecht: Kluwer.

Chapter 10
The Good, the Beautiful and the Literate: Making Statistics Accessible for Action

Jean Paul Van Bendegem, Karen François, and Kathleen Coessens

10.1 Introduction

Compare this figure



with this formula:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

'Working' mathematicians will claim without any hesitation that the above figure and formula represent the same thing, namely a normal distribution or Gauss bell-curve. The label 'working' is apposite here, as some more philosophically minded mathematicians will insist that the formula is the Gauss curve, whereas the drawing is a representation of the curve. Since the authors of this chapter are themselves

J.P. Van Bendegem (✉)
Vrije Universiteit Brussel, Brussel, Belgium

P. Smeyers, M. Depaepe (eds.), *Educational Research: The Ethics and Aesthetics of Statistics*, Educational Research 5, DOI 10.1007/978-90-481-9873-3_10,

quite philosophically minded, we will also take this issue seriously. Formula and figure are not the same kind of thing. Here, we are saying something quite similar to the distinction that is made between (mathematical) proofs and diagrams, a topic that will be developed during this chapter. Does not the figure contain additional elements that have nothing to do with the formula, such as the thickness of the curve or its colour? And, at the same time, does the figure not erase crucial information? After all, from the drawing one cannot conclude that the elusive number π plays a part in it. We do not want to focus on ontological matters here. Rather, our interest is in the ethics and aesthetics of the matter (along with epistemology concerns when they become relevant). However, this is a complex matter. A first-order approximation whereby statistics is seen as either a part of pure mathematics or a part of applied mathematics[1] and where 'only' the ethical and aesthetical dimension are considered, leads to a first-order four-case classification, illustrating the complexities of the issue as follows:

	Ethics	Aesthetics
'Pure' statistics	What are the (often hidden) assumptions internal to thinking 'in a statistical way'?	What are the aesthetic values that mathematicians foster in relation to statistics?
'Applied' statistics	What are the conditions and requirements that allow for the use of statistics in a given setting?	How is the information represented to those who need it?

Some comments relating to each of these possibilities are as follows:

- Pure statistics and ethics: it is clear that (formalised) statistics cannot exist without an underlying theory or conception of probability. Without such a theory, it is, for example, impossible to *prove* the central limit theorem – given a set of independent random measurements of a population with any distribution, the sum of these measurements will be distributed according to the normal distribution – and similar theorems that form the basis of deductive, statistical thinking. Probabilities, however, presuppose a set of conditions. What are the cases about which a probability statement is made (individual cases, samples or populations)? How are these cases weighted amongst one another (think of the concept of an 'honest' die)? How are probabilities calculated? The mere fact that, mathematically speaking, the cases could be virtually anything carries ethical implications. What I mean by this is that, in terms of applications, hardly any cases are ruled out. It is precisely that generality, or dare we say universality, that mathematicians so famously and continuously search for that strengthens the ethical dimension.
- Pure statistics and aesthetics: this in itself is really interesting territory to explore. Mathematicians do have aesthetic ideals that their practice aspires to, but these relate mainly, though not exclusively, to proofs. Now, it is often argued that statistics does not 'favour' proofs and is more 'empirically' focused. Hence, it does not appeal to the same kind of aesthetics. In fact, many mathematicians

regard statistics as the ugly sibling of the mathematical family. But, as mentioned above, there is a clear connection between statistics and deductive, formalised mathematics. Even after Kolmogorov really changed the subject through his axiomatic-algebraic approach (which made it possible to calculate probabilities in cases that were previously excluded) statistics does not seem to share in the aesthetic standards of 'pure' mathematics.[2]

- Applied statistics and ethics: Undoubtedly, is the area where most of us have immediate ideas about what the problems are. Surely each time we apply statistical methods to some real-life problem, the first question must be whether the application is, in some sense, acceptable. Probability theory and statistics really took off in the seventeenth century once international sea trade bloomed (e.g. Hacking, 1975). Insurances became acceptable up to and including not merely a ship's load but its crew as well. From that moment onwards, bets could be made with human life as its literal stake. In that sense it is obvious that ethics enter into the discussion.
- Applied statistics and aesthetics: at first sight, this combination might seem rather odd. It would seem that we are talking here about applications and what could be the aesthetic value of an application? However, we take a different tack. When applied, the tool that is being used has to become very concrete. It cannot remain on an abstract plane, but must be 'translated' or 'transformed' into something concrete. Anything concrete has a specific form and where specific forms are present, aesthetic considerations must come into play. A comparison with architecture (or perhaps music as well?) seems apt in this context. On the one hand, the plans surely have an aesthetic value of their own. On the other hand, once the plan has been implemented through the construction of a concrete building, the building has aesthetic qualities of its own that are lacking in the plans. Think of the aesthetic value a building has for those either living in it or looking at it from the outside.

In addition, it must be clear that the four cases are not independent from one another. Thus, even if it is our aim to focus on the third and fourth case, it is obvious that the two other cases will intervene time and again to provide important insights or to support the claims made. To avoid dividing this chapter up into two separate sections on ethics and aesthetics, we propose to introduce a common element that will allow us to keep both dimensions in view. More specifically, in order to make connections with the philosophy of educational theory, we will focus on the representation of statistics or, if you like, its visualisation. We believe that visualisation is directly connected with an important concept, in terms of education and educational views, namely, accessibility or, as it is sometimes expressed, literacy. When statistical data are presented to a particular audience, how many of its members have 'true' access to the data presented? By this we mean how many effectively understand what is presented to them and can thus be considered literate about the subject? Note that, once again, different levels can (and should) be distinguished: problems of access can occur in the collecting of data to support a (quantitatively

expressed) educational theory, but it can also occur in a situation where a scientist is defending his educational theory in front of non-specialists, say, politicians or related decision-makers. We will treat the question of accessibility or literacy in as broad terms as possible. However, this breadth will be compensated by the presentation of a specific case study concerning visualisation, namely Otto Neurath's ISOTYPE.

10.2 Some Notes and Thoughts About and Related to Otto Neurath's Isotype

One of the most innovative approaches to the representation of statistics in such a way that the greatest accessibility can be guaranteed is the ISOTYPE (International System Of TYpographic Picture Education), developed by Otto Neurath (see Lehrer & Marek, 1997; Nemeth & Stadler, 1996, for the main sources). Much has been written about the *Wiener Kreis*, an organisation that Neurath belonged to. The book series *The Vienna Circle Collection* is an overwhelming source: the writings of Ernst Mach, Hans Reichenbach, Karl Menger, Moritz Schlick, Otto Neurath, Hans Hahn, Friedrich Waismann, Felix Kaufmann, Victor Kraft and others are all available there. It is therefore possible to sketch a sufficiently detailed picture of this group to see the similarities, but also (less obviously) the differences amongst its members. As so often happens, a group is identified with its manifesto if and when they happen to have one. As it so happens, the text *Wissenschaftliche Weltauffassung – Der Wiener Kreis*, published in 1929, is just such a manifesto. It is quite a curious text, as curious as its genesis. It is dedicated to Moritz Schlick, but was written in his absence by Rudolf Carnap, Otto Neurath and Hans Hahn. This is presumably because Schlick would not agree with its content.[3] The most striking feature is no doubt its ethical–political–social commitment. Often logical empiricism is depicted as a purely scientific undertaking and, more specifically, as primarily focused on logic, mathematics and physics. There was a tendency to reduce all other sciences to physics. This encompassed the view that there is nothing wrong with limiting your model to physics, as physics will be the only true or fundamental science. As a consequence, the philosophers defending this view are not really interested in societal matters, let alone ethical matters. The manifesto does not exemplify such thinking. Let us consider two passages from the manifesto where it becomes immediately clear that its writers had an ethical–political agenda:

> The representatives of the scientific world-conception resolutely stand on the ground of simple human experience. They confidently approach the task of removing metaphysical and theological debris of millennia. Or, as some have it: returning, after a metaphysical interlude, to a unified picture of this world which had, in a sense, been at the basis of magical beliefs, free from theology, in the earliest times. (Neurath & Cohen, 1973, p. 317)

and

> We witness the spirit of the scientific world-conception penetrating in growing measure the forms of personal and public life, in education, upbringing, architecture, and the shaping of

> economic and social life according to rational principles. The scientific world-conception serves life, and life receives it. (Neurath & Cohen, 1973, pp. 317–318)

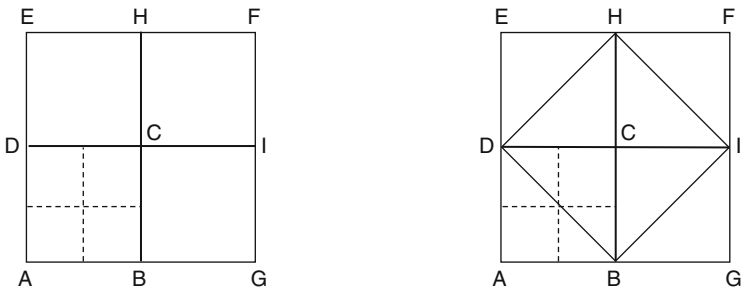
It is quite striking that reference is made to 'earliest times'; we hear the echo of a long lost past where man and the world surrounding him were united. It is the challenge for philosopher and scientist alike to restore this situation to its original form, 'according to rational principles'. Notably, Otto Neurath, one of the three authors, tried extremely hard to realise this plan through the development of a universal pictorial language.

It is a strange phenomenon that the graphical language developed by Otto Neurath is hardly known.[4] The number of papers and books on the topic has to be considered marginal given the number of publications on the *Wiener Kreis*. There is a possible explanation for this: examples of ISOTYPE were known, but no (full) text of Neurath (providing an explanation of the whole enterprise) seems to exist. What was known, however, is that Neurath had written on the subject.[5] This was confirmed with the appearance of three manuscripts. Nemeth and Stadler (1996) include one of these manuscripts, entitled *Visual Education – Humanisation versus Popularisation* in their book.

At first sight there seems to be nothing special about ISOTYPE. The idea that a graphical-visual language might better succeed in transmitting ideas, concepts, theories and the like, is an old and venerable idea (whether correct or not). However, an important element in Neurath's view is that one should avoid the temptation to look for an isomorphy between ordinary language and the visual language one if searching for. The easiest way to make this distinction clear is to contrast the following two images.[6]

The meaning of the image on the left is clear: someone is walking through a door.[7] We see the elements of the sentence in the picture – there is a door, there is a figure representing a person, etc. However, the picture on the right shows what happens if the structure of the sentence 'A person walks through the door' is isomorphically translated (up to and including a curious sign for 'a'). In this case it is as if we have been given a riddle to solve. The puzzling aspect of this derives from the isomorphy the sameness of form, proving how difficult it is to see the same (abstract) form in two different representations. In contrast, the picture on the left almost 'shows' its' meaning. As such, the diagram offers the possibility of seeing everything 'at a glance', allowing possible perspectives and transformation processes which are not necessarily linear or successive, but which can be seen in something resembling an 'aha' experience.

This important observation that a visual and a written representation of the same data are not (and should not be) isomorphic raises equally important questions for the use of diagrams and drawings in various forms of reasoning, including mathematical reasoning. As it happens, this is not a new idea. One of the oldest examples of this idea features in Plato's Meno dialogue. This example concerns knowledge acquisition (Hoffmann 2007).[8] Although Plato is not really interested in diagrammatic reasoning but rather in processes of 'recollection' of ideas, interestingly enough, the example offers evidence of the diagram functioning as an external 'scaffold'. Let us briefly recapitulate the story. Socrates asks a young slave boy to perform a geometrical task. If the side of a square is 2 ft, we know that the surface of the square equals 4 ft^2. What then is the length of the side of a square of double surface – a surface of 8 ft^2? Obviously, the slave's first answer is a length of 4 ft. Socrates explains the slave's error by drawing the first square and then transforming it following the slave's answer (see figure below on the left). By looking at the visual scaffold of the diagram, the slave can immediately grasp his error if, that is, he can interpret the basics of Euclidian geometry. He sees that, by doubling the length of the side, the square becomes four times larger than the original square. Then Socrates goes further, realising transformations on the squares, drawing diagonals on the first square and reaching the agreement of the slave that two triangles together have the same surface as the original square. By further transformations in the second square, he obtains four triangles, forming a square with the right surface (see figure below on the right).[9]



As Hoffmann indicates, the transformation of thought by creating the diagonals and finding the answer by looking at these diagonals is a creative, both visual and cognitive act, involving a leap of and in thought.

Both the Neurath and the Meno examples refer to the importance of diagrammatic reasoning in cognitive processes. By using, creating and reading diagrams, humans solve problems, understand knowledge and make cognitive leaps in complex situations. The diagrams do not merely facilitate understanding, but also offer an external scaffold to help and broaden internal cognitive capacities. Moreover, as external representations, diagrams not only have a cognitive function, but also a communicative one and here we can connect back to Otto Neurath.

Matters become quite interesting if one realises that ISOTYPE was mainly to be used for societal matters. It was Neurath's hope that well-informed citizens would reach 'wiser' decisions that would benefit the community. They need not know about logic, mathematics or physics reach this goal, at least not directly. First of all they need to be informed about societal, political and economical matters. Taking a scientific outlook, this implies that part of the research done in order to gather the relevant data will inevitably rely on statistics. Now it is notoriously difficult to explain statistics to the non-mathematician, that is, most of us. Therefore, Neurath sought and found a way round this problem by representing the statistical data graphically or pictorially *without loss of any of the essential characteristics*. I emphasise the word 'essential' because a great deal of attention is given in ISOTYPE to the elimination of irrelevant components or details that tend to blur the picture rather than to highlight the essentials. The two pictures that follow illustrate this point nicely. Above you can see a diagram as 'ordinarily' presented, and beneath the result after translation in ISOTYPE.[10]

Figure 3: The infographics "Paths of Opium and Cocain to the Buyers" present masses of information instead of quality in terms of information. (*Focus* 3 1993, p.59)

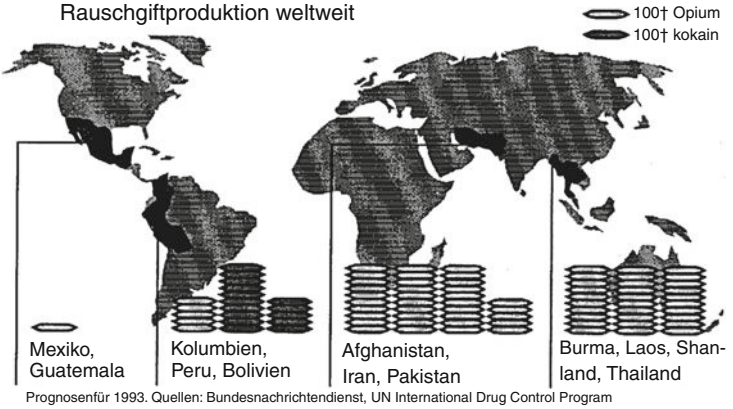


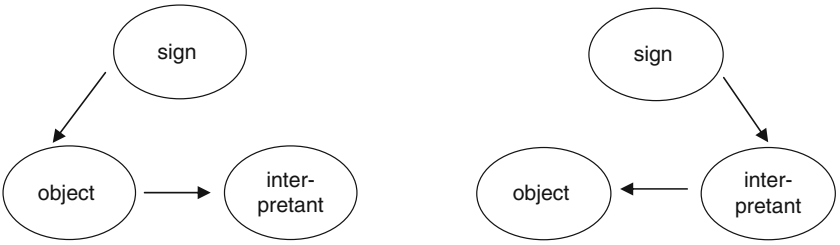
Figure 4: Redesign of fig. 3 in the ISOTYPE style: "Worldwide Drug Production" (Angela Jansen, 1994).

A good indication of the seriousness with which Neurath tried to give concrete form to this project of a universal language is to be found in his attempts to start a museum. His first attempt was the *Museum für Siedlung und Städtebau*[11] (*Museum for City Planning*). This was followed by the creation of the *Gesellschaft- und Wirtschaftsmuseum* (*Museum for Economy and Society*). As must be clear, these were not meant to be museums in the sense of an archive of selected parts of cultural memory, such as paintings, sculptures, art products in general, but rather as places for education – and here we have a first connection to the challenging problem of statistical literacy – of the general public, in such a way as to allow full-blown participation of all citizens in the societal process.

From a more general perspective, the most intriguing aspect of Neurath's work is that it is best characterised not so much as a logico-mathematical way of approaching the problem, but rather as a *semiotic* approach. Perhaps it would not be a bad idea to describe Neurath's view as *semiotic* empiricism, rather than logical empiricism. Seen from this perspective, Neurath's project has deeper connections with the work of Charles Morris[12] and Charles S. Peirce than with his fellow philosophers of the Wiener Kreis. Indeed, in Neurath's intentions, all ingredients for a triadic semiotic interpretation are present. For Peirce, all semiotic relations are triadic relations, involving a sign (or representation), an object and the interpretant. Only combinations of these elements offer and produce meaning. In one sense, they offer a representational insight: a sign represents something for somebody – this is Peirce's (1998) sense.

> I define a sign as anything which is so determined by something else, called its Object, and so determines an effect upon a person, which effect I call its interpretant, that the latter is thereby mediately determined by the former. (p. 478)

But let us turn the semiotic triangle, as in diagrammatic reasoning, the other way round.



The triangle then offers even more: a sign, only by being interpreted and as part of this (specific) realm of interpretation, signifies something. As such it implies socially embedded knowledge: a sign is embedded in a representational (cultural/conventional) system allowing the interpretation of something. It follows from this that sign systems and their tools are not just 'out there'; they are 'human' elements of creation and reasoning. As Hoffmann (2007) writes,

> In this way, representational means are private as well as public, external as well as internal, they are, at the same time, the means of a culture that we have at our disposal and means that live only in our thinking and acting. (p. 9)

Moreover, 'a sign signifies something for somebody' and 'a sign by being part of a dynamic representational and interpretational system shared by a community or society, signifies something'. Once we see things in this way, we engage in pragmatic, practice-based semiotic interpretation. We could look at this not only as a quite radical interpretation of Morris' pragmatic development of Peirce's triadic semiotics but also as an interesting semiotic interpretation of Neurath's pragmatic endeavour. However, it is now time to move on. Even in the best of all possible worlds, where all participants are guaranteed optimal accessibility, where everyone is statistically literate, coming to terms with the *use* of this form of knowledge is still a major problem. More specifically, how is one to *act* on the basis of statistical data, or, in present day terms, how is one to reach a decision and be able to justify it afterwards? How does one deal with the uncertainties, the margins, the exceptional cases and the different sorts of averages or variance of statistical data, regardless of how well they are visually represented? The next section treats this matter first by an analysis of an elementary example, taken from daily life, and then extends the discussion to the actual problem of statistical literacy in our societies today.

10.3 What Is to Be Done? From Knowledge to Action

Let us start with an example taken from everyday life. One of the authors of this chapter was and still is amazed by the weather forecast in his or her newspaper. S(he) is especially intrigued by the enigmatic statement that today there is a 20% chance of rain. What, no matter how it is represented, can this statement possibly mean? This is particularly pertinent as regards the sorts of decisions we make in response to it – should I or should I not take my umbrella along? As it happens, the matter is indeed complex and, in terms of actions to be undertaken, of little value. For, as is shown in Joslyn, Nadav-Greenberg, and Nichols (2009), the meaning of 20% chance of rain is not (a) that it will rain that day for 20% of the time (in which case you should take your umbrella along), (b) that 20% of the surface of your country will get rain that day (if you do not know where exactly this will happen, you should take your umbrella along, unless you like to gamble). Rather it means (c) that in 20 days out of 100 similar days to this day, it will rain a minimal amount. In case (c), what should you do? Note that only a minimal amount is referred to, so it is perfectly possible that on those 20 days it really pours like hell.

There is a simple argument to show how useless this piece of information is, no matter how well it is represented. The argument relies on decision theory itself, the very instrument developed to be able to deal with probabilities and statistics. Consider the following table:

	Day with rain (0.2)	Day without rain (0.8)
Umbrella	Good	Silly
No umbrella	Possible disaster	Good

The first thing we must do is to order the three possible outcomes. Most plausibly, the order should be: good > silly > possibly disaster. So let us assume that value (good) $= a >$ value (silly) $= b >$ value (possibly disaster) $= c$. Then the value of the first alternative to take the umbrella is $(0.2)a+(0.8)b$, whereas the second alternative has value $(0.2)c+(0.8)a$. If we compare the two expressions, it is perfectly possible that in some cases $(0.2)a+(0.8)b < (0.2)c+(0.8)a$. Suppose that $a = 2, b = 1$ and $c = -1$, then $(0.4) + (0.8) = 1.2 < (-0.2) + (1.6) = 1.4$. So, depending on your choice of the values a, b and c, different outcomes will result. To a certain extent this seems odd, for a figure such as 20% presumably carries with it a 'certificate' of objectivity, a result of the painstaking process of statistical data gathering? But then it only seems to work if a sufficient number of subjective elements are brought into play. Actually one should even further refine the calculation to include that, within the 0.2 probability of a day with rain, some days will produce little or no rain, whereas others will be disastrous. Whether that will help is very questionable, because one has no idea whatsoever of the distribution of rainfall within the set of rainy days. This is precisely why one of the authors who is concerned about this problem, always carries his or her umbrella around, using the basic heuristic rule that has saved many a life, 'you never know'.

The point of this example is, of course, to show that, once we want to make the transition from information to action, the ethical–aesthetic dimensions come together and create a set of problems of their own. At first sight, the '20% chance of rain' seems a quite accessible figure, so objective and so aesthetically pleasing. However, in terms of use-in-action, it is rather worthless or too individualised (depending on a set of subjective assessments). At the very least it is not generalisable and hence it is ethically disputable. In short, the problem seems to be that a high degree of aesthetics in the visualisation need not imply an equally high degree of ethics in the action. This bold statement transcends the lessons to be learnt from this single example. Let us therefore have a closer look at a more global problem, namely, the issue of statistical literacy.

To sketch briefly the background of this problem, it is important to realise that the question about the accessibility of statistical information became a central topic within the philosophy of educational theory and, more specifically, the topic became meaningful as part of the discussion on literacy. The concept of literacy, which was originally connected to a competence with written language, both in reading and writing, has been broadened out to mathematics and, more particularly, to statistics. In the case of mathematics, the concept of numeracy is also used, but apart from that notion, the concept of mathematical literacy is a generally used notion in the field of educational theory. The understanding that statistics is not just mathematics

has given rise to a somewhat independent conception of *statistical* literacy and, accordingly, to a new field of study which is called statistics education. This field of study has emerged as an important and sufficiently independent discipline – with its own conferences and journals – that supports the teaching and learning of statistics. Statistics education, as a specific area within educational theory, can thus be seen as a product of the convergence of statistics, on the one hand, and mathematics education, on the other hand – and is currently establishing itself as a unique field of study (Garfield & Ben-Zvi, 2008).

The growing interest in the accessibility of statistics goes hand in hand with the fact that statistical information is, as argued in the Wiener Kreis manifesto, 'penetrating in growing measure the forms of personal and public life' (Neurath & Cohen, 1973, p. 317). It was part of the ethical–political–societal agenda of Otto Neurath, as referred to above, to bring the scientific world-conception closer to human beings' ordinary lives. We can see the same ethical–political–societal programme within educational theory where the questions of literacy became a matter of ethical considerations based on the general idea of emancipation, which can be found in UNESCO's view on education. Moreover, we see in UNESCO's mission an interest not only in socio-economic development, but also the ambition to realise an enduring and peaceful world, while respecting diversity and maintaining human rights. Above all, UNESCO believes that education is the key to social and economic development. We work for a sustainable world with just societies that value knowledge, promote a culture of peace, celebrate diversity and defend human rights. This can be achieved by providing education for all.

The translation of this mission of the UNESCO Education Sector is to

- Provide international leadership for creating learning societies with educational opportunities for all populations;
- Provide expertise and foster partnerships to strengthen national educational leadership and the capacity of countries to offer quality education for all (UNESCO, 1948).

The development of a universal pictorial language for statistical data, such as the one conceived by Otto Neurath was (one of) the earliest 'didactical' innovations to realise a generalised accessibility. However, as we noted earlier, visualising (statistical) information is just one aspect of a more complex matter. The meaning of the concept 'statistical literacy' has experienced an evolution since it came into existence. It became more strongly associated with ethics (Wallman, 1993). In the following section we will elaborate on this important shift in the meaning of statistical literacy.

Initially the term 'statistical literacy' was used to describe the knowledge that people need in order to technically understand and make decisions based on statistical analysis. Haack (1979) states that in order to interpret statistics people need to consider and scrutinise certain aspects of them, which include the source, the type of data, definition and measurement problems and certain considerations concerning the survey sample. As with most authors who began to develop the concept of

statistical literacy, Haack emphasises elements which are basically related to the technical dimension of knowledge on statistics. It seems that this perspective on statistical literacy is based on accepted academic uses of statistics.

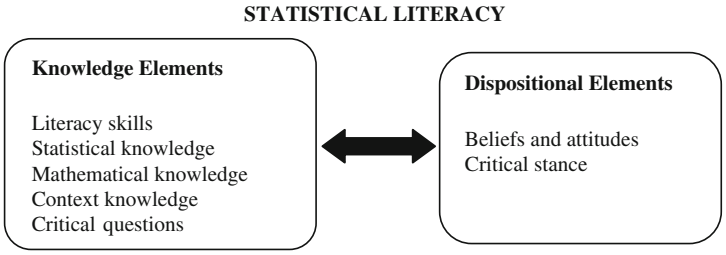
Different authors introduced wider perspectives on statistical literacy related to the kinds of statistical skills which people need in everyday life (e.g. Evans, 1992). Wallman (1993) states in her presidential address to the American Statistical Association, 'statistical literacy is the ability to understand and critically evaluate statistical results that permeate our daily lives – coupled with the ability to appreciate the contributions that statistical thinking can make in public and private, professional and personal decisions' (p. 1). At this point, the ethical–political dimension starts to emerge. The comprehension of statistical information is no longer reserved to the technical and/or professional sphere. It is widened to the sphere of personal life and becomes connected to the notion of 'critical thinking'.

If we compare this wider perspective on statistical literacy, given by Wallman (1993), with the description of *mathematical literacy* given by the 'Programme for International Student Assessment (PISA) 2003 Establishment', we can see a high correspondence.

> Mathematical literacy is an individual's capacity to identify and understand the role that mathematics plays in the world, to make well-founded judgments and to use and engage with mathematics in ways that meet the needs of that individual's life as a constructive, concerned and reflective citizen. (OECD, 2004, p. 37)

In line with this definition, Gal (2002; 2004) emphasises the need for statistical literacy for all citizens who interpret statistics in various everyday situations. Furthermore, Gal suggests that when people read statistics in the media they have to make inferences, quite often in the presence of irrelevant or distracting information, and they may also have to apply mathematical operations to data contained in graphs. The figure below illustrates Gal's perspective on statistical literacy (adapted from Gal, 2002).

This statistical literacy model represents two ranges of elements which, when combined, can enable readers to understand statistical messages. On one side of the diagram there are knowledge elements, which involve cognitive components of statistical literacy (e.g. rational understanding of the data such as knowing how to decode and make calculations about and with it). On the other side are dispositional elements, which comprise a range of 'non-cognitive' aspects (e.g. a person who interprets a graph can have knowledge, experiences and beliefs which might differentiate his/her interpretation of the graph). According to Gal, statistical literacy is based on the interaction of components that comprise each range of elements. Gal's statistical literacy model underlines the fact that the academic or formal schooling background is not the only determinant of the use of statistical skills. To develop statistical literacy, it may be necessary to work with learners in ways that go beyond instructional methods currently in use. To implement all knowledge bases supporting statistical literacy, it is inevitable that topics and skills that are normally not stressed at school need to be addressed (Gal, 2004, p. 73).



The increasing attention paid to statistical literacy gives rise to a number of critical observations. We will restrict ourselves to one important issue that is related to the role of statistical literacy. Here we consider its association with the ethically loaded conception of the development of active and critical citizens who can read and interpret statistics thereby making connections to different areas, reading the world and understanding its complexity. Therefore, statistical literacy should enable people to do more than just read the data but should allow them to criticise and propose alternative interpretations of a given set of data. School systems have a crucial role in developing statistical literacy. They should enable students to understand why and how statistics present a way of describing the world (Frankenstein, 1998; Moreira, 2002). Garfield and Ben-Zvi (2008) distinguish between statistical literacy, statistical reasoning and statistical thinking where statistical literacy provides the foundation for reasoning and thinking. They prefer a definition of statistical literacy that states the following:

> *Statistical literacy* is a key ability expected of citizens in information-laden societies, and [it] is often touted as an expected outcome of schooling and as a necessary component of adults' numeracy and literacy. Statistical literacy involves understanding and using the basic language and tools of statistics: knowing what basic statistical terms mean, understanding the use of simple statistical symbols, and recognizing and being able to interpret different representations of data. (p. 34)

This basic knowledge makes it possible to reason using statistical ideas and to make sense of statistical information. At this stage, students must be able to connect one concept to another and to combine ideas about data and chance. This is called statistical reasoning. The final stage of statistical thinking includes a deep understanding of the theories underlying statistical processes and methods. It also includes the critical competence of understanding the constraints and limitations of statistics and statistical inferences. That is why this stage of statistical thinking is called 'the normative use of statistical models' by Garfield and Ben-Zvi (2008).

In summary, the connection between the concept of the accessibility of (statistical) information, known as the competence of statistical literacy, became explicitly connected to the ethical dimension from the early 1990s on (Wallman, 1993) and it is still a central topic in Garfield and Ben-Zvi's (2008) notion of the final stage of statistical thinking which includes a deep understanding of the theories underlying statistical processes and methods. Thus, making a full circle as we end up with the theoretical issues we started with.

10.4 Conclusion

In this chapter we have merely initiated a train of thoughts about statistics, its ethics and aesthetics, and its literacy. We still do not know whether we should take our umbrella or not, but realising that we do not know seems to have some value in and of itself. It indicates the borderlines where quantitative measures cease to have a societal relevance, where the quantitative needs to be translated into the qualitative, in this case, into the visual, and where access means relevant access.

Notes

1. This in itself is quite a tricky matter: some will claim that statistics can only be seen as applied and that thus the phrase 'statistics as a part of pure mathematics' does not make sense. From that perspective the combination of ethics and statistics as a branch of pure mathematics could refer to the empty set. However, there is no unanimity on this question and so, as long as no definite proof of emptiness has been produced, we consider the combination as meaningful. An additional argument in favour of non-emptiness is the fact that statistics is part of the classification scheme (section 62) of the American Mathematical Society and, in fact, the first subsection is Foundational and Philosophical Topics (see http://www.ams.org/).
2. Unless, of course, one now sees probability theory as part of algebra. Talking about sigma-algebras and functions defined over such structures, satisfying certain conditions, often in the forms of integrals, is quite a stretch away from statements such as 'the probability to get a 3 with a fair dice is 1/6'. More specifically, Kolmogorov's approach made it a lot easier to deal with infinite cases. If one wonders whether this notion makes any sense, just think of a rotating wheel and any question involving the probability of the angle of rotation, which is a continuous, hence infinitely fine-grained variable.
3. So it is claimed by Heinrich Neider in his memories about Neurath. See Neurath and Cohen (1973, p. 49).
4. It is, for example, quite curious that in the extensive and authoritative publications of Tufte on visual representations of information (see Tufte, 1983, 1990, 1997) only in the third part ISOTYPE is briefly mentioned and the name of Otto Neurath is lacking altogether.
5. In Neurath and Cohen (1973), Marie Neurath quotes from one of Otto Neurath's papers on ISOTYPE (pp. 224–248). So at least in 1973 it was known that some texts had to exist.
6. The example is taken from the text of Neurath himself (see Nemeth & Stadler, 1996, p. 330).
7. Here, we should allow room for discussion. It could be a drawing of a person standing still in a framework of some kind (a scanner at an airport?), or of a person leaving or entering a room, or it could be a still from a film. In that sense the figure is definitely ambiguous and open to multiple interpretations.
8. Hoffmann (2007) in his article 'Cognitive Conditions of Diagrammatic Reasoning' offers a very deep analysis of Plato's case and the cognitive importance of diagrammatic reasoning.
9. These two figures have been taken from Hoffmann's article (2007, pp. 15 and 20).
10. These two figures have been taken from the contribution of Angela Jansen in Nemeth and Stadler (1996, pp. 147 and 148).
11. The name of the museum betrays Neurath's interest in architecture. As it happens, Neurath not only lectured at the Bauhaus and convinced Rudolf Carnap to do likewise, but he also participated as non-architect – an adaptation of the statutes was required to allow for his presence – in the CIAM (Congrès International d'Architecture Moderne), one of its founders being Josef Frank, who was the brother of Philipp Frank, a member of the Wiener Kreis.
12. It is therefore in my view not surprising that one of the contributions to the 'International Encyclopaedia of Unified Science' – or, to be more specific, the launching of the

Encyclopaedia – was written by Charles Morris, one of the core figures in semiotics in the twentieth century, entitled 'Foundations of the Theory of Signs' (Chicago: University of Chicago Press, 1938). Interestingly enough, Morris sees semiotics both as a subdiscipline of the human sciences and as a general science of all sciences since all sciences need signs to express themselves. To make the picture complete, also consider that Morris was heavily inspired both by logical empiricism and by C. S. Peirce. The semiotic connection then becomes inescapable.

References

Evans, J. (1992). Mathematics for adults: Community research and 'barefoot statisticians'. In M. Nickison & S. Lerman (Eds.), *The social context of mathematics education: Theory and practice* (pp. 202–216). London: Southbank Press.

Frankenstein, M. (1998). Reading the world with maths: Goals for a critical mathematical literacy curriculum. *Proceeding of the 1st International Mathematics Education and Society Conference*, Nottingham [online report]. Retrieved January 29, from http://www.nottingham.ac.uk/csme/meas/papers/frankenstein.html

Gal, I. (2002). Adult statistical literacy: Meanings, components, responsibilities. *International Statistical Review, 70*(1), 1–25.

Gal, I. (2004). Statistical literacy. Meanings, components, responsibilities. In D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning and thinking* (pp. 47–78). Dordrecht: Kluwer.

Garfield, J. B., & Ben-Zvi, D. (2008). *Developing students' statistical reasoning: Connecting research and teaching practice*. New York: Springer.

Haack, D. (1979). *Statistical literacy: A guide to interpretation*. North Scituate: Duxbury Press.

Hacking, I. (1975). *The emergence of probability*. Cambridge: Cambridge University Press.

Hoffmann, M. H. G. (2007). Cognitive conditions of diagrammatic reasoning. [online working paper] Retrieved February 13, from http://www.spp.gatech.edu/faculty/workingpapers/wp24.pdf (Forthcoming 2010 in J. Queiroz & F. Stjernfelt (Ed.), Peircean diagrammatical logic [Special issue]. *Semiotica*.)

Joslyn, S., adav-Greenberg, L., & Nichols, R. M. (2009). Probability of precipitation. *American Meteorological Society, (February issue)*, *90*, 185–193.

Lehrer, K., & Marek, J. C. (Eds.). (1997). *Austrian philosophy past and present. Essays in honor of Rudolf Haller* (Vol. 190). Boston studies in the philosophy of science. Dordrecht: Kluwer .

Moreira, L. (2002). Mathematics education and critical consciousness. *Proceedings of the 26th Conference of the International Group for the Psychology of Mathematics Education, UK, 3,* 369–376.

Morris, C. (1938). *Foundations of the theory of signs*. Chicago, IL: University of Chicago Press.

Nemeth, E., & Stadler, F. (Eds.). (1996). *Encyclopedia and utopia: The life and work of Otto Neurath (1882–1945)* (Vol. 4, Vienna Circle Institute Yearbook). Dordrecht: Kluwer.

Neurath, M., & Cohen, R. S. (Eds.). (1973). *Otto neurath: Empiricism and sociology: Vol. 1. Vienna circle collection*. Dordrecht: Reidel.

Organisation for Economic Co-operation and Development. (2004). *Learning for tomorrow's world: First results from PISA 2003*. Paris: OECD.

Peirce, C. S. (1998). *The essential peirce: Vol. 2. Peirce edition Project*. Bloomington, IN: Indiana University Press.

Tufte, E. R. (1983). *The visual display of quantitative information*. Cheshire: Graphics Press.

Tufte, E. R. (1990). *Envisioning information*. Cheshire: Graphics Press.

Tufte, E. R. (1997). *Visual explanation images and quantities: Evidence and narrative*. Cheshire: Graphics Press.

United Nations Educational, Scientific and Cultural Organization (1948). *Universal declaration* of *human rights* [online report]. Retrieved January 29, from http://portal.unesco.org/education

Van Bendegem, J. P. (2004). Why do so many search so desperately for a universal language (and fortunately fail to find it)? In F. Brisard, S. D'hondt, & T. Mortelmans (Eds.), *Language and revolution. Language and time.* (Vol. 106. Antwerp Papers in Linguistics; pp. 93–113). Antwerpen: Universiteit Antwerpen.

Wallman, K. (1993). Enhancing statistical literacy: Enriching our society. *Journal of the American Statistical Association, 88*(421), 1–8.

Chapter 11
Statistics and the Inference to the Best Explanation: Living Without Complexity?

Paul Smeyers

11.1 The Scope of the Envisaged Problem: Statistics and 'Making Sense of Them'

Statistics are everywhere. In their descriptive mode they often indicate the level of satisfaction of clients/customers of post offices and hospitals, of one's use of an email system (e.g. Eudora), of the number of downloads (provided by publishers of journals), and are gathered for various purposes (including to evaluate scholarly conferences such as AERA). This may be innocent as far as it goes, but it carries with it a number of presuppositions that should not necessarily be taken for granted. One of these is that it is possible to represent reality (and have a grip on it), shifting often to the simple equation that this *is* the reality we belong to. This either implies that (1) there is nothing else to know about a subject or (2) if there is anything else to know, this is still the best way to deal with reality, the one that is the least harmful and most objective. Another presupposition (or should one say promise) that descriptive statistics carry with them is the possibility of control, particularly when an explanation is offered (by this we mean the control of the evolution of dependent variables in terms of the independent variables which have led to the distributions that are offered or the correlations that may be observed). Thus a paradigm of (quasi-)causality enters that goes together with estimation and management of risk (this applies in the case of experiments as it does for randomized field trials). At the same time, the attraction of statistics lies in its simplicity (the reduction of the variability of the data to as few 'factors' as possible, as few homogeneous classes as the data permit, even for $N=1$ studies where a simple formula is strived for or if more variables are dealt with, where the interaction between them is estimated/calculated), as well as in the accompanying belief that it is possible to characterize reality and the way it shapes our lives. Though what has been outlined so far may well be recognized (or deplored) by many, there is a more complex story

P. Smeyers (✉)
Ghent University, Ghent, Belgium

P. Smeyers, M. Depaepe (eds.), *Educational Research: The Ethics and Aesthetics of Statistics*, Educational Research 5, DOI 10.1007/978-90-481-9873-3_11,

to be told. Moreover, though we may be familiar with the possible objections and take these 'warnings' into account, this does not really change the fact that we often unthinkingly rely on what is presented to us. Why is that the case, or in other words, what do statistics do for us that make them so irresistible?

Consider for instance the following abstract which was sent around recently to the staff of the Faculty of Psychology and Educational Sciences, K.U.Leuven. The abstract was for a talk to be given by Peter Wilhelm (University of Fribourg, Switzerland) titled 'How to Assess Synchronization in Family Members' Daily Life?'

> How family members affect each other's feelings and behaviours is a crucial question in family and social psychology. To answer this question, ambulatory assessment techniques that capture family members' experiences and behaviours in their usual daily life over days or weeks are especially useful. To date various diary studies have been conducted to investigate transmission of affects in couples and in parent-child dyads. However, diary studies to investigate the synchronization of current affects in entire families are rare. The aim of our research was to investigate the extent to which family members' current moods synchronize in their usual daily life and how this synchronization can be explained. Using a computer-assisted diary approach, six times a day over the course of a week we asked 192 parents and 122 adolescent children from 96 Swiss families how they were feeling, where they were, with whom they were, and whether they had experienced conflicts with each other. To answer our research questions we propose a multilevel approach in which the similarity between family members' average mood (trait component) and the similarity between their current mood (state component) is reflected in the covariances of the random part of the model. Using this approach we were able to show that average mood was moderately correlated among family members. As expected, family members' current mood (state component) was more similar when they were together, than when they were apart. This indicates that family members' mood synchronizes when they come together. By extending the analyses we could explain the amount of synchronization by shared setting factors that have an impact on mood, such as location, activity, and time of the day. However, conflicts between the family members produced the largest decreases in the covariances and therefore seem to contribute most to the synchronization of current mood in family members' daily life.

It may be fun to speculate about the practical consequences of the insights this study comes up with, i.e. the family members' mood synchronizes when they come together especially when they have conflicts. Has it thus been established that one should test a relation to its destruction or has a more general argument been developed in favour of quality time? A positive reading may indicate that the latter is obviously the case – we need to spend time together if we desire synchronization of current affects in families. Though this seems rather obvious, this research can substantiate it. Granted, this does not go very far, but at least it provides insight into how an issue in family and social psychology is dealt with. It provides answers and takes us beyond mere speculation in regard to something we may be (really) interested in (how things are). The following abstract of a talk by Jeff Rouder (University of Missouri-Columbia) for another seminar at the same Faculty, titled 'Using Bayesian hierarchical models to search for structure in recognition memory', is much more straightforward when indicating what, according to its author, is at stake:

> Science has conventionally progressed by identifying regularities or invariances; for instance, planets orbit the sun as ellipses. These invariances then serve as the phenomenon to be explained by theory; for instance, elliptical orbits may be explained by the Newtonian theory of mechanics. In this mode, the search for invariances is a necessary precursor to theoretical development. Though this approach has been fruitful, psychologists, who tend to focus on verbal explanations of differences idiosyncrasies across conditions and populations, have not adopted it. The search for invariances is complicated in many domains by the presence of unintended or nuisance variability. For instance, in memory research, there is variability from the selection of participants and items in addition to that from the mnemonic processes at hand. The conventional approach is to aggregate across these nuisance sources, but this aggregation may distort the structure and mask important invariances in the data. To address this problem, my colleagues and I have developed a collection of Bayesian hierarchical models for modelling nuisance variability so that structure in psychological processes may be explored. I apply these models to recognition memory, where there is a vigorous debate whether memory is supported by a single mnemonic process or by multiple distinct processes. Model analysis reveals that ROC [receiver operating characteristic] curves across people and conditions form an orderly field, much like the order in gravitational of magnetic fields. As a consequence, the structure of recognition memory seems to be one-dimensional and may be accounted for with a single parameter of mnemonic strength.

A single parameter of mnemonic strength is the result of a search for invariances in this area. It is a 'structural quality' (a reduction of the complexity of the data here as invariance, in other cases for instance a pattern), which should now be taken up by theory and explained (or 'given a place'). We might ask what exactly the relationship is between this result and the particular method that has been used (in this case, a collection of Bayesian hierarchical models). Furthermore, we might consider whether it is in some sense useful. Clearly this is all about reduction, about 'seeing' something in a simpler form. For this reason (maybe for this reason alone) it is something to be strived for.

In educational research as everywhere else in science, academia or even ordinary life, we are for various reasons interested in trying to understand (some would say, 'make sense of') the phenomena we are surrounded by or subjected to. This understanding is supposed to give us a grip on what is happening, bring some order to the chaos. It is not in the least bit concerned with how we see ourselves, or what we aspire to achieve. Some explanations are more attractive to us than others, some are more popular in particular periods, and some focus on groups. Others centre on the individual. Crucial in this endeavour are issues about truth (A) but at the same time a lot will depend on what one is particularly interested in. (B) The latter may or may not be intertwined with one or other sort of good that is envisaged either by a person or a particular group and this opens up the sphere of manipulation (C) but also of responsibility and responsiveness. (D) Thus matters of the particular concept that is used when studying a problem come to the forefront (E) as crucial factors concerning the characterization of that phenomenon. This will have implications for where one can arrive at in terms of solutions for dealing with it.

Let me begin by arguing that whether something is *really* explained, or whether 'reality' here is merely the opposite of fiction, should not necessarily invoke a correspondence theory of truth where sense data are the exclusive building blocks. Instead, as Peter Winch (1958) rightly argued, it is always about 'what is real for

us'. It goes without saying that answering a research question in terms of causes and effects will not generate an answer in terms of the understanding of those involved. But this kind of circularity is not to be regretted. It is characteristic of all explanation. Scientific, and for that matter any other kind of explanation, will always take the data which are to be interpreted to a higher step of abstraction. This will engender a particular theoretical construction that makes sense. This is a circular process in which each level is taken to account for, to derive from, or to elaborate on the other. Thus instances are explained by patterns and patterns by instances. For Winch too, there is more at stake than just the practitioner's understanding or the concepts of those involved (i.e. raw data as interpreted phenomena), but those of the 'student of society' as well. Clearly, here it is not prediction that may exclusively provide us with a point of reference, nor is the method of the natural sciences the only way to come to valid conclusions. But if the possibility of prediction is what one is interested in, even then a meaningful background must be present. That sense data and observations will play an important role does not jeopardize the claim that concepts are always involved. Possible exceptions occur at a very basic level. For example, light of a particular intensity is painful. In such cases one may argue that particular phenomena have 'meaning in themselves' and do not presuppose a shared meaningful background. With the example of the light, what is 'shared' here is the human body and its particular physiology. But these cases/examples are rare, and moreover, once we start speaking about them, a language in which we can determine 'what makes sense for us' is implied. Again this has to be distinguished from a context in which reasons can be given. It is therefore not correct to argue that phenomena which have a 'meaning in themselves' play no role whatsoever in our understanding of human behaviour. But it would also not be correct to ignore that more is involved if they have a place in our lives. In that case they also presuppose some kind of shared meaning that will include elements over and above the physical, chemical, physiological or biochemical level. I therefore conclude that when an explanation is offered *(1) concepts necessarily refer to 'what makes sense for us'*.

In this chapter I will not focus on the complex issue that deals with what has to be taken into account to determine which concept will be appropriate. Although it may not be possible to ignore this issue completely it is not the main focus of this chapter. By the same token, I will not be focusing on what is usually addressed in qualitative or interpretative research, i.e. the reasons (F) which people hold (referring to practices or the 'form of life') as opposed to explanations which use variables (independent and dependent) in a (quasi-) experimental design (or, if you like, a randomized field trial). Again this does not mean that issues such as freedom and responsibility are not relevant here (the threat of the disappearance of the ethical, thus inviting us to live 'beyond freedom and dignity'), but they will not be my main focus. Thus, bracketing (to some extent) (A), (B), (C), (D), (E), and (F) allows me to address something much simpler. What I want to address is *why we are so eager to turn to one or other kind of statistics when trying to understand and deal with particular social practices*. The strength but also the weakness of statistics lies in reducing the complexity of reality. Can this complexity be reduced? It seems to be something we yearn for but why is this so? I will therefore give some attention to

whether we are simply unable to live with complexity. But before turning to this, let us consider some more general explanatory points about statistics. Though I will try to be as brief as I possibly can, it is paramount to foreground a number of central issues that set the scene when we say we 'explain' something.

11.2 Explanation and 'Causality'

When trying to understand natural phenomena one may want to distinguish a unified style of explanation from explanations of a mechanical kind. The former explains in terms of basic comprehensive principles (for instance in biology: selection, mutation, heritability of traits). In other words, a unified style of explanation situates phenomena into an overall scheme. The latter style of explanation answers questions of how things work (sometimes understood as what they are made of). It therefore comes as no surprise that Wesley Salmon, a philosopher of science, argues that the constant efforts of the sciences to explain involves 'revealing the mechanisms at work in the world' (1989, p. 156). In terms of understanding a society one may be interested in how it functions at large, what and why people do what they do, and 'who' they are. However, in many cases in this area, the concept of 'causality' seems to pervade our thinking about ourselves and others, about our environment, even concerning the entire universe we live in. It is involved in the use of technology (where we attempt to achieve particular effects while avoiding undesirable ones) and in our everyday practical planning and dealings. For a lot of authors, to explain an event is to identify its antecedents, i.e. its causes.

Though clearly in many cases the use of statistics is aimed at unravelling causal mechanisms, this is not always the case. For example, one may be interested to know how many cases there are with this or that characteristic and this can be useful for a particular purpose. These descriptions (a distribution of the actual cases concerning a particular variable) may be vital for policy issues. Let us take bullying as an example. How many cases of bullying are registered in primary schools? This can be specified further for particular subgroups such as boys and girls, according to age, ethnicity, various living conditions, and so on and so forth. A model of causality looms up behind these descriptions. Now it goes without saying that to have an informed estimate of the frequency of the occurrence of a particular problem (as detailed as this can be) is quite essential in educational contexts. Policy needs to take this into account, as it can be an element in the process of determining how relevant the problem is and what should be done about it. It should, however be noted that when studies use *descriptive statistics (2)*, which evidently have a place in their own right, they take a step further. Then statistics become part of a much broader (sometimes 'causal') concern, whether they are used to explain what is going on, or to be made use of (i.e., used to tackle the envisaged problem). I refer to this as *the explanatory use of statistics (3)*. I will not deal with the area of problems the 'use of' statistics confronts us with (these are typically identified as 'contextualization', 'generalizability', or even 'eco-validity'). Instead, I will focus on what lies

at the basis of the explanatory approach. The open door descriptive statistics offers lures us into an area that is beset with serious problems. Such problems are often put aside ('forgotten') in the interest of explanation.

In quantitative research, one typically looks for a distribution of variables (how many cases are there with this or that characteristic) and uses this for explanations. Such explanations can be of a deductive-nomological kind (incorporating universal laws) or be of an inductive nature. Explanations of an inductive nature are supposed to contribute to what Peter Lipman calls '... Inference to the Best Explanation' as they provide 'a partial solution to the problem of description by giving an illuminating account of the black box mechanism that governs our inductive practices' (Lipman, 2004, p. 164). Clearly, bringing something under a set of laws can also offer an explanation not in terms of an argument (a logical structure with premises and conclusions governed by some rule of acceptance), but as a presentation of the conditions relevant to the occurrence of the event and a statement of the degree of probability of the event given these conditions.[1] Thus the problem of determinism versus indeterminism is invoked. For the determinist, the fact that we are unable to make perfect predictions in all cases is the result of human ignorance and other limitations. This is not because nature is lacking in precise determination; clearly, accuracy of prediction is irrelevant to whether determinism is true in principle. This framework is challenged by frameworks of indeterminism, for instance by quantum theory. It is not simply that quantum mechanics is *prima facie* non-deterministic, but that under plausible constraints no deterministic completion of the quantum theory is possible. In view of this it seems inadvisable to accept determinism as an *a priori* principle – and of course the truth or falsity of quantum mechanics is a matter of physical fact. Doubts about the possibility of finding causes for everything, either on the basis of logical or empirical considerations or on the basis of relativity theory, press the case for a move to indeterminism as the more rational choice for the overarching framework. However, as far as I can see, this sort of metaphysical discussion is not terribly pertinent to the argument presented in this chapter. Indeed, as argued above, laws can be seen as offering an explanation by presenting the conditions relevant to the occurrence of the event and a statement of the degree of probability of the event given these conditions. I therefore conclude that the discussion of determinism or indeterminism does not damage the use of causes and laws. Moreover, because of this it will not diminish the interest in causes and causal explanation. We explain facts (general or particular) by exhibiting the physical processes and interactions that bring them about, but such mechanisms need not be deterministic to have explanatory force. In the following I will talk about laws and causes with this in mind.

The form philosophical discussions of causality take is usually as follows: there are two facts (or types of) or two events (or types of) between which there is a relation R. Sometimes the logical structure of the relation is discussed in terms of necessary or sufficient conditions or a combination of both, which given the interaction of several conditions leads to complex schemes for understanding particular occurrences. To put the issue again more generally, when two types of events, A and B, are positively related to each other, we hunt for a common cause C that is

statistically relevant. The statistical-relevance relations must be explained in terms of two causal processes in which C is causally relevant to A and C is causally relevant to B. This is the heart of matters where it is claimed that a statistical explanation is based on causality. Now the question is, why should we prefer, for explanatory purposes, the relevance of C to A and C to B over the relevance of A to B, which we had in the first place? The answer is that we can trace a spatiotemporally continuous causal connection from C to A and from C to B (while the relation between A and B cannot be accounted for by any such direct continuous causal relation). Recall that, according to Hume, causal explanations present us with a problem. As deductive logic cannot provide the answer (that explains why ball number 2 is set in motion after being hit by ball number 1), Hume turns to empirical investigations. On the basis of his observations he concludes that in situations where we believe that there is a causal relation, there is a temporal priority of the cause to the effect. There is furthermore a spatiotemporal contiguity of the cause to the effect and finally, on every occasion on which the cause occurs, the effect follows – there is constant conjunction. As there is, in his opinion, no physical connection between the cause and the effect (the connection does not exist outside of our own minds), the relation between cause and effect is to be found in custom and habit. We may want to qualify the latter and argue for a more robust interpretation of causality (see for instance, Salmon, 1998), but for our purposes little depends on this. It does not affect either the use of cause or our interest in causes. As indicated, it is important to realize that when statistics are used to explain the occurrence of events, a model that uses causes is operative. Hempel's position, often referred to in this context, makes this clear.

In an explanation one may cite specific conditions obtaining prior to the event (initial conditions) and invoke general laws. It is held that the occurrence of the event to be explained follows logically from those premises, i.e. those initial conditions and laws. One can distinguish between deductive explanations that incorporate universal laws (which hold without exceptions) and inductive explanations, which employ statistical laws (which hold for most or many cases). According to Hempel (1965) scientific explanation consists in deductive or inductive subsumption of that which is to be explained under one or more laws of nature. This is referred to as the deductive-nomological model (D–N). For Hempel, however, inductive-statistical explanations are essentially relativized to knowledge situations – he suggested the requirement of total evidence that took the form of the requirement of maximal specificity, where all possibly relevant knowledge is available. If there were an inductive-statistical explanation whose law-like statistical premise involved a genuinely homogenous reference class then we would have an instance of an inductive-statistical explanation *simpliciter*, not merely an inductive-statistical explanation relative to a specific knowledge situation. However, as there are, according to Hempel, no inductive-statistical explanations *simpliciter*, ideally inductive-statistical explanation would have no place in his account. There is a striking similarity between this kind of explanation and Laplace's formulation of determinism. In view of this close relationship it is tempting to conclude that events that are causally determined can be explained, and those that can be explained are causally determined. We have set aside above the problems of determinism versus

indeterminism, as well as those concerning the formulation of laws (among others under-determination; see also Note 1). What is important is to focus on the fact that in many cases we do not have enough facts to be able to construct a full explanation and we can never be sure that a new condition might not turn up: one can never exclude the possibility that a further relevant subdivision of a reference class might be necessary on the basis of additional knowledge. Moreover, an explanation requires a sufficient condition that is based on empirical evidence that something actually happened, and inference on the other hand refers to something in the future. To infer something that lies in the future not only presupposes that everything relevant has been taken into account but requires that the future replicate the past. Notwithstanding these difficulties, it seems that for a statistical-relevance model it is the amount of relevant information that counts; it consists of a probability distribution over a (maximum) homogeneous partition of an initial reference class (and thus is all about the gain in information it provides).

So one answer to the question 'why are we attracted to statistics?' comes down to the fact that it helps us to find what is hidden (either a law, or a law-like generalization). We think that this hidden law may assist us in dealing with future problems. Therefore, though we are aware of the problems explanations in terms of causality are beset with, and we are knowledgeable about the limits of the use of statistics (and take these into account), there is no more trustworthy method to rely on when we want to know how things work (either in the natural or the social world). Let me remind you that I have bracketed of course (A), (B), (C), (D), (E), and (F). Of course, you and I may agree that what has been bracketed cannot in fact be bracketed. Therefore, the idea that there is no other more trustworthy method to rely on if we want to 'know how things work' may not get us very far. But let me also remind you that I am talking about attractiveness of statistics when it comes to 'explaining' things. This is notwithstanding what I have highlighted concerning the indebtedness of concepts to 'what makes sense for us' nor the different use that is sometimes made of statistics such as the descriptive function, also taking on board the problems associated with (E), i.e. the delineation of the concepts themselves. It is time to pause for a moment and consider the connotations that 'investigation' and 'detection' present us with. After all, researchers investigate, do they not?

11.3 On Crime Stories and Looking for the Truth

There are many kinds of detective fiction, a format that is well suited to dramatic presentations and this results in the popularity of many television and film detectives. Particularly focusing on the latter one may want to distinguish (cf. also Wikepida) between the amateur detective (Marple, Fletcher), the private investigator (Holmes, Poirot), the police detective (Dalgliesh, Morse) and the forensic specialists (CSI). It is hardly ever a matter of either/or, but it may suffice for our purposes here to distinguish those who primarily deals with a search for evidence when trying to find out what happened (such as CSI), and those who focus much more on understanding

the reasons for the crime and thus on the interplay between the individual characters (part of a particular social environment and social practices) such as 'Morse'. One could argue that in the first case the programme maker is drawing the viewer's attention to physical evidence and in the second case to the clues which are hidden in what people do (or did) and how understanding that will resolve the issue (who committed the crime). Suspense is created in both cases until we find out how everything hangs together. Yet, in some sense, this plays a subordinate role in series such as 'Morse' or 'Waking the Dead', which is about so-called cold cases. Here the discipline of experts is combined in the field of psychological profiling, forensic science and police detection. Detective Superintendent Boyd plays the driven and dynamic leader of a crack squad of investigators which includes psychological profiler Dr Grace Foley, forensic scientist Dr Frankie Wharton, DC Mel Silver, and DS Spencer Jordan, Boyd's second-in-command. Pushing it to the limit, the enjoyment of watching these two kinds of series lies (with the first category) in knowing who committed the crime; in the second, though that still plays a part, it seems that even if one knows who committed the crime, it is still possible to watch it a second time and enjoy the complex plot and various clues that trigger other ways of seeing things. It would be understandable if after all, it turns out that someone else had committed the murder.

I will entertain for a moment the idea that what we do in research (explain how things work) has some similarities with what is at stake in these crime stories (the search for the truth). There are evidently huge differences: a crime story starts from the fact that a crime has been committed and that someone has committed it (it is about the past). This entails that it is at least in theory possible to unravel the past and thus to identify unambiguously the person who committed the crime. The researcher who is interested in how things work has a serious disadvantage. As he relies on concepts and theories to explain what happened, he will never be in a position to be as firm as the detective. Furthermore, once the mystery is solved, that is the end of the story. Research does not 'end' in this way. There is always more to be taken into account and the situation may change (it is thus not only about the past but also about the future). But leaving this aside, there are interesting parallels. For example, the methods that are used to identify the criminal bear some resemblance to various statistical methods – in some cases statistics are used. Looking for DNA (sometimes conclusive but in cases of more distant relatives often not), for membership of particular groups, work environments and criminal records. Any of these methods may help to identify the perpetrator. While none of them are conclusive, the combination of some of them may generate elements which put the burden on the accused to explain, for example, why she was at a particular place. It is of course possible that there is a rational explanation for having been at a particular place (alternatively there are also cases when someone is framed), but this will have to be given before the detective is convinced. Further differentiations can be made when the two types of crime stories are considered together. Somehow the investigator of the CSI-type is blinded by the range of the kind of evidence that she is looking for – is there a logic of discovery operative here? There are methods to be followed and precautions to be taken so as not to contaminate the evidence; there

are steps (a plan as in a script), there are procedures to be followed (to ascertain that the obtained evidence will be acceptable in a court of law which necessitates search warrants) and evidently, if the investigator is in one way or another herself related or otherwise involved, she cannot be part of the investigating team (it may jeopardize her objectivity or even corrupt the investigation).

Of course that does not mean that the above is not relevant if a different approach is taken, but crime stories which focus on historical circumstances (the past) and characters approach the problems to be resolved somewhat differently. Physical evidence remains important. But the way to gather it and the ways in which sense is made of what is found are somewhat different. For instance, by employing a profiler, 'Waking the Dead' holds in some sense a middle ground between the rather impressionistic approach of Morse (who often gets it wrong) and the almost exclusive attention given to physical evidence by Horatio Caine in CSI. An example of this middle ground can be found for instance in 'Deathwatch' episode 2 of the second season. When trying to identify a person investigators combine different sorts of evidence. The individual suffered from emphysema (asbestosis) and had been in the military. These factors, taken together with other elements, meant that investigators decided to search for someone who has worked in confined spaces, perhaps an engineer or a submariner. In the same episode a clue is given to the audience when the executioner (Truelove) says 'I always thought of the police as my colleagues and now what do you make me think of them'. The audience are alerted to police corruption and that they should look there to understand what happened in the past (the hanging of an innocent man) and a more recent murder.[2]

Statistics I argued are attractive because they inform us about how things are and how they work. Crime stories of the first kind create suspense by postponing the answer to the question 'who committed the crime'; crime stories of the second type are doing much more, i.e. they open up the complexity that is involved, yet they too have a plot and a 'solution'. They would not be (so) attractive if these things were left out. Indeed, if they were left out, we might say that the show belonged to a different genre. Anyway stories of the second type are not only concerned with 'why' but also 'how'. Would we criticize someone for being interested in 'who did it'? Would this be legitimate? If so, why not make an analogy with the enthusiasm for statistics? The same element presses itself forward. Some people are only interested, so it would seem, in simple answers (to complex things). Maybe they do not want to live with complexity. Do we really have an argument against seeing things in this way, or is it really only a matter of taste (defined as preferring cola to lemonade everything else being equal)?

Incidentally I am not trying to draw a parallel with the conceptual element present in a theoretical research approach (in the sense of a logic of discovery) that is supposed to be operative in educational research. The parallel I am interested in is about 'how something works' and 'who committed the crime'. In comparing the two types of crime stories I am trying to make the point that the approach of the CSI-kind is limited and that the Morse-type is much broader. By the same token, in some cases physical evidence can be decisive. Studying causal interactions may explain what is happening in a particular case and reasons beyond a more minimal conceptual level

will not play an important part. Let us go back to the issue we ended with in the previous paragraph: is there a case for more complexity? Again I will draw some evidence for this from another crime story (of the Morse-type).

What I have in mind can be found in an episode of the French series 'Maigret' (another example of the Morse-type). In 'Maigret et le clochard'[3] the story is told of an attempted murder of a tramp: While he was sleeping under the Pont Marie, someone had hit 'Doc' over the head and tossed him into the Seine to drown, but a couple of Belgian bargemen nearby had fished him out in time to save him. His identity card revealed that he is François Keller, from Mulhouse, where, coincidentally, Mme Maigret's sister lives. Thus Maigret learns that he had been a doctor there, left his wife and daughter and gone off to Gabon, in Africa, to work 'like Dr. Schweitzer'. But things hadn't gone the way he had hoped, and for the past 15 years or more, he had been a *clochard* in Paris, sleeping under the bridges, working as a sandwich man to buy his bottle of wine. Maigret interviews his daughter and mother, but there seems to be no motive or contact. He goes down the Seine to find the bargeman again, and learns that his wife's father had owned the barge until he had drowned in the Seine 2 years before, near the bridge at the Quai de la Rapée, where Doc had slept at that time (though not since). The former owner of the barge used to beat his daughter, and when he found out that she was having a relationship with the present bargeman, Jef van Houtte, the situation became unbearable. Though the drowning of this man was labelled 'accidental death', Maigret is convinced that the bargeman Jef van Houtte is guilty. Van Houtte denies all, and Doc is no help – he refuses to say anything. Maigret is finally forced to let van Houtte go, but a few months later a chance comment by Doc that 'it's impossible to pass judgment' convinces Maigret that he was right, though Doc maintains that he had said nothing. Clochards have their own idea of justice and do not want to get involved in other people's affairs. Here we have a case where the kind of investigation that Maigret conducts is stopped, because no witnesses (who are willing to speak) come forward. In a way he gives up due to the daughter's situation (before her father 'drowned') feeling that the father has received 'justice' and that possible 'injustice' would result from further efforts to clear things up completely. Maigret seems to be satisfied by his investigation. He is willing to accept no for an answer in view of the greater good. He did not continue his search for more physical evidence, nor did he put more pressure on the witness. He left things as they were, because understanding more (either about 'how it worked' or 'why') would put him in a difficult position. Thus he has to live with this and bear the burden of not intervening further and seeking to do justice 'for its own sake'. Here one is tempted to say that following the simple logic of crime and punishment can create more injustice; at the same time one could argue that leaving things as they are is the 'simpler' answer. Who lives under the bridges is invisible, but sees everything. They and others should be left alone. To live with this (evidently a case of a different kind of complexity) may be more than many are inclined to do. It foregrounds what 'justice' is about and moreover that simple facts are important (which is as a consequence of complexity often forgotten). One may want to compare this more generally with discussions about social justice where a lot is made of inclusion and exclusion but remains oblivious

to the simple fact that if we do nothing people will die. There is a real danger that knowing too much may push other aspects out of the way. It is not only that reductionism looms, but that it becomes quite difficult to get the balance right. Too much particularity thus turns out to be potentially dangerous. It is to the issue of whether statistics can do justice to reality, to one or other kind of complexity, to which I now turn.

11.4 In the Structure of the Language: Living with Complexity and Statistics

In the *Philosophical Investigations* Ludwig Wittgenstein argues against essences and metaphysics, or more precisely against our metaphysical disposition. For Wittgenstein, this is exemplified in the general propositional form of a sentence. Against the Augustinian picture of language (all propositions are or contain a description), he argues that 'language plays us entirely new tricks' (Wittgenstein, 1966, *Aesthetics*, I, # 3). And in section 115 of the *Philosophical Investigations* he writes 'A *picture* held us captive. And we could not get outside it, for it lay in our language and language seemed to repeat it to us inexorably' (Wittgenstein, 1953, *I*). The reason for this is that in a proposition, something of the subject of that sentence is always said. His move is in the opposite direction, towards particularity, and thus philosophy offers reminders for a particular purpose. In this sense (and only in this sense, because Wittgenstein is wary of theory generally and of looking for hypotheses or 'what is hidden from us'), statistics may not be that different at all from our 'normal' way of speaking. The antidote against the latter may give us some clues to how we might develop an antidote against the former. We might also ask what the place of policy in such a framework could be. However, clearly retreating into 'a new way of looking at things' as one or other kind of aesthetic existence may not be enough to convince the sceptic. Though much can be said for particularity at some point we may want to transgress such a stance because resources are limited and we are urged to make decisions on a larger scale. There is much more to be said about this, but I should return now to the central focus of this chapter, namely, the attraction of statistical explanation.

Interestingly, the aesthetic dimension we touched upon has also been addressed more generally in regard to research methodologies in the social sciences. David Fenner (2006) for instance argues that quantitative and qualitative methodologies could be seen as exemplifying respectively a disinterested and an engaged approach. He maintains that there exists a strong parallel between current research methodologies in the social sciences and the two most central and popular approaches to aesthetics over the last 4 centuries. Fenner approvingly quotes Eddy Zemach who argues that within the stance of scientific realism, 'Scientists apply criteria to theories, such that those theories that best meet those criteria are those theories judged to most closely approximate the truth' and that '[a]t least in part, these criteria are

aesthetic: simplicity, elegance, unity, etc.' (Fenner, 2006, p. 321). Somewhat further he endorses the position that scientists do not celebrate aesthetic properties like ambiguity or 'sublime disorder'. In the conclusion he claims that '[m]otivations to employ one educational research methodology are quite similar to motivations to favor one sort of approach to aesthetic experience (or judgment) over the other' (Fenner, 2006, p. 327). This far-reaching conclusion reiterates the point that was made earlier concerning a taste for the simple or the complex. That the simple can be dangerous is evident. Though simplicity may often be a good thing, it avoids the hazards of the over-sophisticated, the baroque or even the decadent (as John Dupré argues in 'The lure of the simplistic', 2002, p. 284 concerning the theory of evolution), this attractiveness can make us blind to what is naïve, unsubtle, or just misguided – he speaks of the 'unifying and simplifying urge, reductionism', Dupré, (2002, p. 291). So is statistics only about the simple in its pejorative sense?

There have been attempts to rescue analytical statistics in a way that is acceptable to skeptical critical realists. In an interesting paper 'A critical epistemology of analytical statistics', Wendy Olsen and Jamie Morgan (2005) argue that the significance of a method can be transformed by a reconstruction of the methodology. In their words, 'It is not implausible to suggest that methods predicated on closure and regularity can contribute to an open methodology that in turn has something valuable to say about an open social reality' (p. 262). Furthermore, they argue that '. . . a post-positivist social science would want to reject empiricism but not necessarily all the techniques associated with empiricism' (Olsen & Morgan, 2005, p. 265). There could be a case for optimism regarding the possibilities for operationalizing aspects of social reality. Furthermore, realists need not be empiricist in their interpretations. They argue that eight claims or criteria can be set up to provide for a multifaceted defence of analytical statistics.[4] This allows us to avoid throwing out the baby with the bathwater. Here are Olsen and Morgan:

1. Methodological closure need not presume a closure in reality for explanatory significance. Non-identity between the two means that a closed method may still contribute to a realist account of an open system where the degree of openness is known in advance.
2. A manipulation that constructs regularity need not imply that the basis of that relative regularity is arbitrary or unrepresentative of the aspects of the world under scrutiny. Synthetic epiphenomena highly determined by method itself can occur, but do not necessarily occur.
3. Regularity-seeking analytical statistics are capable of highlighting non-regularity and the breakdown in relative regularity. Analytical statistics can accommodate complexity and contingency.
4. The interpretation of analytical statistics allows non-atomistic inferences about relations.
5. As a consequence of 1–4, the results obtained through analytical statistics can be counter-phenomenal or unexpected. As such they are able to contribute to a qualitative understanding.

6. The role of the analyst in the initial choice of method(s) and in the subsequent development of the particular research application is highly significant in realizing the possibility of 1–5.
7. As a consequence of 1–6, manipulations can contribute to retroduction [which explains what conditions in reality may have or could have led to these observations] to causal mechanisms rather than hypostatising [acting as if, or assuming that, the relations between things are fixed] variables as chains of events through interpolation.
8. 1–7 imply that an analytical statistical method may be appropriate as part of a methodologically pluralist research project. (Olsen & Morgan, 2005, pp. 269–270)

Olsen and Morgan claim that one need not believe that data *represent* the world for them to be useful in constructing arguments (Olsen & Morgan, 2005, p. 277). Moreover, it may be extremely difficult to disagree with their conclusions (Olsen & Morgan, 2005, p. 278):

> Counterfactual work in analytical statistics requires comparative statistics, and can usefully bring out findings about minority groups vs. the majority, or different populations which have common causal mechanisms. The class structure of different countries, and different policy regimes which together generate different employment outcomes, illustrates the possibilities for comparative statistics. To throw out statistics as a method would involve discarding all possibilities for such comparative work, yet such work has been very illuminating in many disciplines.

Sylvia Walby (who is interested in a more theoretical stance) develops a parallel argument (2007) where the implications of complexity theory for the analysis of multiple intersecting social inequalities are investigated and applied. Similarly Helga Nowotny (2005) draws attention to the fact that we seem 'to be engaged in describing and interpreting complexities with the desire to understand and to engage in both – in building higher-level complexities because they are a more efficient way of doing things and to reduce complexity in order to minimize undesirable effects and to be able to cope with the increasing levels of complexity around us' (Nowotny, 2005, p. 19). Both, she claims, are indicative of a dynamic which points to the ongoing co-evolutionary processes between science and society.

Evidently, there is a real danger here of committing a logical fallacy, in which one ignores the original topic of an argument and subtly changes the subject, but still claims that the conclusion concerning the original subject is reached even though the argument has little to do with the conclusion. This so-called 'red herring fallacy' looms around the corner for statistics. It is committed when the arguer diverts the attention of the reader or listener by addressing a number of extraneous issues and ends by presuming that some conclusion has been established. This confuses scientific with faith-based belief. What can we conclude from this? It seems to me that when one does not hypostasize (by this I mean formally treat or represent something abstract as a concrete reality) the results of statistical analyses, they have the potential to engender better understanding. Of course, statistics may reveal what we are interested in or point us to something that we were unaware of (for example

that in the Flemish-speaking part of Belgium compared to the French-speaking part more drugs are used to treat ADHD and less for depression). Yet at the same time they may conceal what we should (also) be concerned with. But to simply regard statistics as dangerous (or redundant) does not do justice to the complexity of what is involved in understanding.

The philosophical problems about the structure of language, namely the particular metaphysical enframing (and longing) that unavoidably seems to take place (this is overwhelmingly clear in Plato's work and those positions that were developed in this vein) and which was foreshadowed in pre-Socratic stances (such as the fascination for numbers) and present in later developments for instance by Bacon and Descartes, haunts us – to use a Wittgensteinian expression. As a result, the ideals of objectivity (bracketing the performative embeddedness) and rationality that, since the Enlightenment, have characterized our understanding of reality may be seen as indicative of an unwillingness to live with complexity. Humans not only long for knowledge (to know how things are, for instance 'Is this a tenor characterizing a particular voice?' or 'How long will I be in the traffic jam – incidentally, knowledge of that does not shorten the time?'), they also seem to have an insatiable need to gain control over the world. That we always use concepts that invoke something general, that there is no alternative to this even when we take it fully into account, is a profound Wittgensteinian insight. When looking at crime stories we saw that statistics can take us down the right road. Their attraction lies in the fact that they make things simpler and answer a 'human all too human need' to get some kind of grip on reality. This is bound up with the force of rhetoric, a force that may be easier to exploit now than ever before given the availability of super computers and web-based dissemination of what has been found to be the case.

Notes

1. Incidentally, it is interesting to notice that examples from the atomic and subatomic world show us that there is a limit to the joint precision with which two so-called complementary parameters can be known: there is an inescapable uncertainty if one attempts to ascertain the values of both the position and momentum of a particle. The same thing goes for energy and time. Ascertaining the position of an electron with great precision makes us unable to ascertain its momentum exactly, and vice-versa. More precisely, when a photon strikes an electron, the direction in which the electron will go is not determined. There is a probability distribution over all possible directions. Furthermore, in this collision the amount by which the frequency of the photon will change is not determined. A probability distribution over all possible amounts exists. Because of the conservation of energy and of momentum, there is a perfect correlation between the direction of the electron and the change in frequency of the photon. The pair of values is however not determined. Incidentally, it is important in this context to refer to problems with our instruments of measurement. The click that results from a genuine photon detection is utterly indistinguishable from the click that results from a spurious count. And finally, there is of course the presumption that conditions surrounding this particular occurrence can be specified in enough detail to establish the existence of a unique necessary and sufficient cause (this example is discussed in Salmon, 1998). To offer an explanation here is something different: it comes down to the assembling of a total set of relevant conditions against which the event can be explained, and to the citing of the probability of that event in the presence of these

conditions. It is in this case not an argument (a logical structure with premises and conclusions governed by some rule of acceptance), but rather a presentation of the conditions relevant to the occurrence of the event, and a statement of the degree of probability of the event given these conditions. Evidently, a persistent statistical correlation – a genuine statistical-relevance relation – is strongly indicative of a causal relation of some sort, but one should not confuse statistical correlation with genuine causation. This would be to conflate symptoms with causes.

2. Fea, V. (Producer), & Phillips, M. (Director). (2002). *Waking the dead. Deathwatch* (Season 2, Episode 2). London: BBC.
3. Vercel, E., & Lassa, C. (Producer), & Heynemann, L. (Director). (2003). *Maigret et le clochard* [*Maigret and the dosser*]. Paris: Dune France 2. (Based on a novel by G. Simenon)
4. They use what they call a widely accepted notion of analytical statistics as the mathematical process of manipulating survey data in an attempt to reach 'well-founded' conclusions which generalize across the region and time-period from whence the data came. The analytical methods can be seen as including regression as well as some exploratory methods such as factor analysis (cf. Olsen, & Morgan, 2005, p. 280, footnote 1).

References

Dupré, J. (2002). The lure of the simplistic. *Philosophy of Science* (Suppl.)*, 69*(3), 284–293.

Fenner, D. E. W. (2006). The aesthetics of research methodologies in the social sciences. *International Journal of Applied Philosophy, 20,* 311–330.

Hempel, C. (1965). *Aspects of scientific explanation.* New York: Free Press.

Lipman, P. (2004). *Inference to the best explanation* (2nd ed.). London: Routledge.

Nowotny, H. (2005). The increase of complexity and its reduction. Emergent interfaces between the natural sciences, humanities and social sciences. *Theory, Culture and Society, 22,* 15–31.

Olsen, W., & Morgan, J. (2005). A critical epistemology of analytical statistics: Addressing the sceptical realist. *Journal for the Theory of Social Behaviour, 35,* 255–284.

Salmon, W. C. (1989). *Four decades of scientific explanation.* Pittsburgh, PA: University of Pittsburgh Press.

Salmon, W. C. (1998). *Causality and explanation*. New York: Oxford University Press.

Walby, S. (2007). Complexity theory, systems theory, and multiple intersecting social inequalities. *Philosophy of the Social Sciences, 37,* 449–470.

Winch, P. (1958). *The idea of a social science and its relation to philosophy.* London: Routledge.

Wittgenstein, L. (1953). *Philosophical investigations/Philosophische Untersuchungen* (G. E. M. Anscombe, Trans.). Oxford: Basil Blackwell.

Wittgenstein, L. (1966). In C. Barrett (Ed.), *Lectures and conversations on aesthetics, psychology and religious belief*. Oxford: Basil Blackwell.

Chapter 12
Performativity, Statistics and Bloody Words

Ian Munday

12.1 Introduction

The obsession with effectiveness and education, which has been going on for some years now, has been a much discussed and written about subject in the philosophy of education. In this chapter I want to discuss how statistics and the particular kind of discourse that emerges around them serves to suture the wounds in the discourse of effectiveness culture. I will begin by giving a flavour of how British schools operate and show how performativity functions within them. Lyotard who famously coined the term performativity believed that only resistance to effectiveness culture was to turn to absence and silence. In contrast, Derrida offers a more optimistic metaphysics and his philosophy of language frames the discussion of statistics that follows. There I argue that statistics simultaneously and paradoxically provide the promise of truth and absolute scepticism and that the disorientating effects of this suture over linguistic slippages/bleeding. This, I argue helps to maintain the farcical 'fixing' of statistics that goes on in British schools. Statistics like all numbers can never operate alone; they are bound to policy initiatives that tend to be emblematised through slogans. However much slogans (that become mantras) may, like statistics, suture the bleeding of language, this can never be wholly successful. This is where optimism lies.

12.2 Performativity and the School

Let us begin with a typical 'scene' from British schooling. The first day back after the summer holidays is usually an in-service training day. The first INSET day has everything to do with performance and is a resonant reminder of how teachers' professional subjectivity is generated by it. The first meeting of the year often involves the entire teaching staff watching senior management run through the GCSE results,

I. Munday (✉)
University of Cambridge, Cambridge, UK

P. Smeyers, M. Depaepe (eds.), *Educational Research: The Ethics and Aesthetics of Statistics*, Educational Research 5, DOI 10.1007/978-90-481-9873-3_12,

etc. on a PowerPoint presentation. This involves a comparison with previous results and results that were predicted by various computer programmes based on comparative statistics. Generally speaking, this will either be followed by congratulation for success at meeting targets, or a telling off, thinly veiled in managerial language, for 'failure'. To be an 'effective' professional one must maximise student potential and produce positive statistics.

The picture illustrated above clearly demonstrates that what teachers do is largely regulated by, or in relation to, technological systems that determine an expected outcome/result. The methodologies for achieving these results are homogenised as far as is humanly possible and the shape of the year leads to annual regulation of success. The language that shapes and surrounds these procedures is replete with acronyms, technical terminology and the language of business. This exemplifies what Lyotard refers to as performativity. The grand narratives of modernity (pertaining to truth and justice) have undergone a process of delegitimation and the postmodern world behaves in accordance with a system that has exiled those grand narratives. Such narratives have been supplanted by the logic of performativity, which has taken hold of knowledge.

> The production of proof, which is in principle only part of an argumentation process designed to win agreement from the addressees of scientific messages, thus falls under the control of another language game, in which the goal is no longer truth but performativity – that is the best possible input/output equation. The State and/or company must abandon the idealist and humanist narratives of legitimation in order to justify the new goal . . . (Lyotard, 1984, p. 46)

We can easily see how the scene from schooling described at the start of this chapter embodies Lyotard's description of performativity. Schools put in massive efforts to ensure that they gain value on their statistics and demonstrate the best possible input/output equation. As Gordon Bearn notes, Lyotard, despite giving his assent to the plurality that accompanies the demise of all-encompassing theories, does not celebrate what has appeared in their place. It is interesting that Bearn presents Lyotard's philosophy in 'aesthetic' terms maintaining that 'it is not painted in the slack polychromatic colours of eclecticism' and neither is it 'painted the reassuring black of a glorious tragedy'. Ultimately, we are told, 'Lyotard's philosophy is painted a melancholic grey' (Bearn, 2000, p. 232).

Bearn does not say as much, but Lyotard's philosophy cannot embrace 'tragedy' for several reasons. First, tragedy would imply a nostalgia for something substantial that has now been lost. When Lyotard notes the postmodern incredulity to grand narratives he is not suggesting that all-encompassing systems were ever fit for purpose – they simply appeared credible. Consequently, nostalgia gets us nowhere in either a philosophical or practical sense. Why then, can we not celebrate the current state of society and education?

The problem can perhaps be stated in this way – performativity is a grand narrative of sorts, just a hollowed out one. Performativity 'functions' like a grand narrative, 'If a form of knowledge could not be translated into bits of information, it was bound to become more and more invisible to the system . . .' (ibid., p. 231). Consequently, performativity provides just as overwhelming and brutal a systematic

horizon as any grand narrative that preceded it. 'Openness' and 'diversity' are the order of the day but are only deemed acceptable when read against this horizon. Here we might think of the statistical imperatives faced by universities to 'attract' students from ethnic minorities to their courses.

So, what can be done about performativity replete with its statistical imperatives? Lyotard has no hope that we can fully escape it, though that is not to say that nothing can be done. What Lyotard offers as a response is an extreme ascetic form of negation. The genre of discourse that he refers to as performativity will always win out – our words/ideas will be systematically dragged into it. Therefore, the only alternative form of action is to think the 'blank' and make a radical retreat into the emptiness beyond representation. This way, we negate all desires, goals and purposes (these would invariably fall prey to statistical imperatives). Alone in this emptiness we can bear witness to the 'event', the phrase arising out of the abyss of not-being prior to its being linked to other phrases in the 'genre of discourse' Lyotard calls performativity.

Is this all we can hope for? It seems that Lyotard's approach (by definition) does not get us anywhere at all. This is because the 'event' must always be 'betrayed by putting it into words, colours, a link, a phrase' (Bearn, 2000, p. 238). Bearing witness must always fail (this of course supposes that anything will come out of the emptiness in the first place). We are left with the depressing claim that 'To link some emptiness is necessary' (Bearn, 2000, p. 241). At best we are provided with a pseudo religious worship of absence, 'Peregrinations in the desert', a pilgrimage to nothing. Perhaps we can hope for something more! I want to suggest that Derrida's philosophy of language, which we are about to consider, offers a more optimistic metaphysics that the distortion of statistics cannot do away with. The discussion of Derrida's philosophy will act as a backdrop to the subsequent treatment of statistics.

12.3 Derrida, Difference and Iterability

Derrida wishes to draw attention to the false metaphysics that has held philosophy in its grip. He does this by providing a critique of what he calls logocentrism and phonocentrism. To think logocentrically is to adhere to a belief in truth that is beyond or above history. This vision of truth sees various truths as having the quality of metaphysical presence that is internal to them. This truth is 'internal' and therefore superior to all forms of 'exteriority' whether they are cultural, historical or textual. It seems obvious that language originates with speech – we speak before we can write! It therefore appears equally obvious that speech is somehow closer to the logos. To think in this way is phonocentric.

> As has been more or less implicitly determined, the essence of the *phone* would be immediately proximate to that which within 'thought' as logos relates to 'meaning', produces it, receives it, speaks it, 'composes' it. If, for Aristotle, for example, 'spoken words (ta en te phone) are the symbols of mental experience (pathemata tes psyches)' (De interpretatione, 1, 16a 3) it is because the voice, producer of *the first symbols*, has a relationship of essential and immediate proximity with the mind. (Derrida, 1997, p. 11)

To show how Derrida provides a critique of phonocentrism and logocentrism, it will be helpful at this point to introduce Saussure and his theory of the linguistic sign. The sign is made up of the signifier – an 'acoustic image', and the signified – a concept. Language is a structure made up of signs. For Saussure, the signifier is in an arbitrary relationship with the signified – there is nothing 'dogish' about the word 'dog'. Language is structured according to differences – cat is different from bat by one letter and this difference in the signifier leads to two completely different signifieds. It should be noted that Saussure's theory is clearly phonocentric.

> In every case, the voice is closest to the signified, whether it is determined strictly as sense (thought or lived) or more loosely as thing. All signifiers, and first and foremost the written signifier, are derivative with regard to what would wed the voice indissolubly to the mind or to the thought of the signified sense, indeed to the thing itself . . . (ibid.)

For Saussure, the signified belongs to/is located in the mind. Because the voice seems more immediately connected to our thoughts, it is closer to the signified. Consequently, the 'notion of the sign always implies within itself the distinction between signifier and signified, even if, as Saussure argues, they are distinguished simply as the two faces of one and the same leaf. This notion remains therefore within the heritage of that logocentrism which is also a phonocentrism: absolute proximity of voice and being, of voice and the ideality of meaning' (ibid., p. 12).

Derrida's reading of Saussure is not wholly critical. He agrees with Saussure that the linguistic sign is arbitrary, but argues that Saussure fails to account for the implications of this. Having seen that there is nothing dogish about the word dog, Saussure turns away from the more radical possibilities that accompany this recognition and turns towards the logocentric assumption that there is some fullness of meaning that exists behind the sign, which is then embodied in speech.

Derrida argues that if we take Saussure's argument to its logical conclusion, then the arbitrariness demonstrated by the latter exposes a dislocation, a fundamental otherness that characterises the relationship between the mark of the signifier and the 'content' of the signified – in this sense the signified is in an original sense, absent from the mark. To go one stage further, if we can only think in words, and yet our words as referents bear no necessary relationship to their meaning, then the logocentric hierarchy (thoughts – spoken words – writing as supplement) is in a sense inverted. It would seem that the mark comes first. If we take this logical reordering seriously then it would seem that 'full presence' is in fact an illusory effect of language/something produced by language. Meaning is not out there waiting to be worded. Rather, words as they come into being word the world. Consequently, the signifier does not represent the signified, but brings it into 'presence', into being as an effect. Therefore, the figure of the linguistic sign (with its unity linking signifier and signified) is no longer adequate to the task of understanding how language works and Derrida replaces it with the 'mark'. The 'mark' is an indicator of absence rather than presence. It exists in the form of the trace, the graphic representation of the word prior to its signification.

The account as it has so far been developed strips language of its otherness in regards to metaphysical/logocentric truth. Truth as such is produced by language

rather than being anterior to it. Therefore, notions of difference understood in essentialist terms have been undermined, as truth must now be conceived of as a linguistic 'performance' of some kind. Let us briefly return to Saussure. As we have seen Saussure locates difference in accordance with the distinction between different signifiers and their related 'signifieds'. Once we see that these differences no longer reflect some metaphysical ordering then difference must be understood as something internal to language. Words/concepts differ as effects of language. Therefore, what a mark/word is depends upon all the other marks/words that it is not. Consequently, every word is 'haunted' by what it is not – and must therefore in one sense be radically dislocated and 'other' to itself. Derrida refers to this general difference, prior to all subsequent differences as differance. The difference between difference and differance only works in French because in that language the distinction is inaudible. In choosing to spell the word differently, Derrida is making a point about the priorness of writing (conceived of in the new way described earlier).

So far we have considered a synchronic picture of language. The other dimension of differance can only be accounted for if we move beyond this to a diachronic picture of language in use. As Derrida notes, for words to mean anything at all they have to be repeatable and repeated, iterable and reiterated, they have to be 'used' and used over time. For Derrida, words are subject to an internal force and movement. It is not the case that any single context determines the force of words. Rather, an unlimited number of possible contexts are internal to the words themselves. Therefore, the word or concept is never at one with itself, it is always other to itself. Derrida insists that he is not referring to the polysemy of language but its iterability. If we perceive the forces in language to be external, then the context determines meaning. By showing that words carry their contexts with them, Derrida demonstrates that the fullness and completeness, the presence ascribed to the speech situation is never fully realised; never mind the force of the present context, the words will not allow it:

> . . . a written sign carries with it a force that breaks with its context, that is, with the collectivity of presences organizing the moment of its inscription. This breaking force (force de rupture) is not an accidental predicate but the very structure of the written text. . . by virtue of its essential iterability, a written syntagma can always be detached from the chain in which it is inserted or given without causing it to lose all possibility of functioning, if not all possibility of 'communicating' precisely. One can perhaps come to recognise other possibilities in it by grafting it onto other chains. No context can entirely enclose it. Nor any code, the code here being both the possibility and impossibility of writing, of its essential iterability (repetition/alterity). (Derrida, 1988, p. 9)

12.4 Language, Numbers and Statistics

Having provided a short account of Derrida's philosophy of language, I now want to explore the role of statistics in performativity more directly and the ways in which the utilisation of statistics involves doing something with words. Numbers (and this is something we possibly forget) are words and moreover they are iterable – we

might talk of 20 bags of potatoes or 20 people shot for insurrection – the number 20 finds itself in different contexts in which it will resonate in different ways. Moreover, throughout the rich history of language various numbers have become 'significant'. There are plenty of examples of this, some of which involve numbers in combination, some of which require (or perhaps do not require) explanation. Let us list some: 13 (unlucky for some) 666 (the number of the beast), 69 (I am not sure if this has resonance outside Britain, though the pervasiveness of English suggests that it might), 99 (red balloons), 9/11, 6,000,000 (Jews killed in the Holocaust), 10 (the number usually worn by the best player in an international football team). Certain numbers take on a moral status. An ethical significance attaches itself to some of these numbers, notably 666, 69 and 6,000,000. Moreover, one of the numbers is a statistic (we'll come back to this shortly, though we might note that 6,000,000 is usually written as six million as though there were something obscene about using polynomial/positional notation here).

Now, clearly, the numbers used above are iterable, they will find themselves in all manner of contexts in which their previous contexts will necessarily play some sort of role. Moreover, they gained their performative force from being repeated. One cannot mention 6,000,000 anything without thinking about the Holocaust. If the number 69 appears in the national lottery this will provoke a degree of smirking and general titillation around the country. The number 999 is the British number of the emergency services but turn it on its head and . . . even the most confirmed atheist may look over her shoulder just to make sure that a little red man with a tail and horns isn't standing behind her.

Though numbers are iterable, they are iterable in a very specific way. Moreover, they are not iterable in the same way that other kinds of word are. When we looked at what can happen with the number/word 20, we must note that the fact that 20 can take on a different resonance is due to the words that surround it. The iterability of 20 here has nothing to do with the word/numeral itself. Moreover, the 'meaning' of 20 does not change at all. In the case of 666, assuming one is not (too?) superstitious, the meaning that is attributed to it has nothing to do with where it has been before. In contrast let us take a word like 'queer' which has come to denote homosexuality. Of course there is nothing about the word itself that relates to homosexuality, its passage from an insult expressing moral disapproval (homosexuality here is supposedly, strange, unedifying and unnatural) to being reclaimed by the gay community whereby the transgressive image of homosexuality is celebrated, reflects the complexity and richness of the iterable life of the word. Even apparently 'functional' words like 'both' 'and' either and 'or' express an extraordinary richness if one considers the role they have played in explaining deconstruction. This sort of richness cannot apply to numbers. In his chapter, Standish (see Chapter 14, this volume) captures what is at stake here when he notes that numbers are untranslatable.

> . . . there is the peculiar place of numbers in contemporary natural languages. Numbers, it seems, can appear in translations of texts in a way that is entirely without loss or distortion, and in this respect they are unlike the (more obviously) linguistic aspects of those texts. But in a sense this is to say that they cannot be translated at all – given that all translations between natural languages do involve loss and distortion in some degree. To some, this

> will confirm the view that numbers achieve an ideal clarity of meaning, and indeed it is partly such a thought that lay behind the experiments with language in which Leibniz and others engaged. It is partly what computers achieve (and depend on). But to others, this will demonstrate the ways in which numbers fall short of the very qualities of meaning upon which our thought and being and our accounting for ourselves, are sustained . . .. (Standish, 2010, p. 214)

It is perhaps unnecessary to note that sympathy here lies with Standish's 'others' and almost certainly, with Standish himself. Let us return to 666. In a sense what has happened to it is emblematic of the false metaphysics vis-à-vis language that Derrida's philosophy tries to overturn. The 'attribution' of significance to 666 interestingly demonstrates a common misunderstanding of the ways in which language ordinarily works – 666 (the signifier) 'signifies' the devil (signified). As Derrida shows us, this kind of logocentrism is precisely undermined by differance and language's translatability/untranslatability. The fact that the significance of 666 has to be artificially attributed to it reveals the barrenness of numbers. They do not give birth to any kind of 'becoming'. Numbers are marks, but somehow not the 'trace' of anything. In this sense, they tell a half-truth about language, as they are not the progeny of the metaphysics of presence. For Derrida, the metaphysics of presence is not a wholly bad thing – it is what allows meaning to take place. Paying uncritical lip service to it causes all the problems.

So far, we have looked at numbers and iterability. Statistics are of course numbers but not just any old numbers. Statistics supposedly represent something that is there, they offer the promise of a truth that is etched into the fabric of existence. Having looked at the some of the ways in which numbers can become iterable, and how this reflects a false metaphysics vis-à-vis language, we might say that something similar is going on with the symbolic practice of statistics. Just as 666 'signifies' the devil, statistics are supposed to 'signify' some truth about the world. Consequently, statistics as numbers look 'clean'. Now, the process of accumulating statistics is bound to be open to the linguistic slippages that apply to iterability as it generally works in language. However, when statistics are shown, and it is interesting that statistics are something we ordinarily 'show' (a point about aesthetics?) the processes behind the accumulation of statistics are not usually made explicit. This is true for debating politicians, or for school managers introducing statistics to teachers. The function that statistics perform in many of their common usages is to render invisible the complex processes that led to their status as data. This is what gives statistics the lure of purity – they suture over all the messy stuff that led to their generation and hide the 'meaningful' process that takes place before and below. I want to draw explicit attention to the metaphor of the suture (a metaphor that Derrida regularly employs) here. Sutures are made up of non-toxic substances that can be absorbed by the body. Statistics, which, as numerals, 'appear' pure, have been absorbed into educational practice?

That said teachers are often critical of the data supplied on their classes. They will commonly complain about the targets their students are supposed to meet, 'student x has been predicted a level 5 – she's never going to get that! It's all a load of rubbish'. Here the teacher shows suspicion towards the statistics generated, but it

is unclear whether they think that a better treatment of the data would have yielded more accurate results; the implication seems to be that it would. The data that is given leads teachers to be suspicious, but this is always caught up in the dialectic with a truth that a better approach to statistics might yield. Please note that I am not suggesting that any student is capable of anything, but that judging 'potential' should not be reduced to checking things off against data.

Let us not be too quick here. Paradoxically, though statistics present the promise of truth, they also present the threat of absolute scepticism. Though we might become inured to the performative force of statistics in our everyday working lives, this is never absolute. We are always dimly aware that statistics can be produced to make any case whatsoever. What this demonstrates is the paradoxical double bind accompanying statistics that can work in disorienting ways on teachers through their discourses and practices. Such disorientation arguably serves a purpose. If teachers are preoccupied with whether or not their statistics are accurate, or if they are suspicious of statistics generally, they may be less inclined to consider whether or not their teaching should be thought of in terms of effectiveness. I recently marked a group of essays on 'inclusion' in which the students (who were trainee teachers) often pointed to statistical imperatives (as regards target grades) as an obstacle to wider concerns about 'including pupils'. These wider concerns related to the importance of effective teaching strategies to accommodate learning differences and the promotion of 'self esteem techniques'. They made it clear that we should not forget that the teaching assistant was the most important 'learning resource'. What is most interesting about the views expressed in these essays is that the students clearly see no connection between effectiveness when thought of in terms of meeting statistical targets and the discourse of 'effectiveness' per se. Of course, the very 'techniques' they champion are those that are employed to gain value on statistics. Consequently, we might say that just as statistics can take on a suturing function through the promise of the constative, they also serve to divide experience. They are set up as the enemy of better teaching, yet the very notion of what better teaching involves is inextricably bound up with statistics. The ultimate effect of this simultaneous suturing and division of experience is suturing at a meta-level. Performativity is served by the paradoxical nature and use of statistics.

12.5 Fixing the Stats

Furthering our discussion of statistics and language, 'truth' and scepticism, let us note that whereas words cannot be fixed statistics famously can. This is not to say that words cannot be manipulated, they cannot be 'fixed' in either a neutral or 'ethical' sense – iterability and interpretation will not allow it. The fact that statistics can be fixed is due to their suturing effects resulting from the barrenness of numbers (the surface effect) and the ways in which statistics partake in the fluctuation between the promise of the constative and absolute scepticism. Let us see how farcical things have become within English schooling.

In England, the situation has intensified as regards statistical imperatives. Earlier this year, schools were told that that 30% of students must achieve 5 C grades or more including English and mathematics. This process goes by the name of the 'National Challenge'. In keeping with media-friendly heroic sloganeering, this sounds like a jolly game show in which people from all round the country may participate via their television sets and mobile phones. 'Failing' schools (many of whom do not possess cohorts that are even predicted to meet this target) must find a way to boost their statistics and take on the 'National Challenge'. There are a number of ways of doing this that I am familiar with.

Given the fact that English and maths are primary requirements in reaching the magic number, schools can fill up the timetable with a greater number of English and maths lessons. Then, if one lacks a sufficient number of teachers to deliver these lessons (there is such a great demand for English and mathematics teachers that 'failing schools' can often not attract enough of them) take non-specialist teachers (skilled in 'behaviour management') from 'non-essential' areas such as physical education or history to teach English and maths. Such teachers will at least be able to ensure that something gets taught. Consequently, supply teachers can cover the non-essential subject areas. Of course such subject areas can potentially all but disappear. This takes us on to ways of fixing the stats. There are numerous vocational courses that schools now offer which are equivalent to GCSEs in terms of their statistical value. So, for example, rather than doing a GCSE in French, one might do an NVQ Business French. Everybody knows such courses are easier for children to do well in. They are often dominated by coursework activities that can be repeated until decent grades are achieved. Consequently, failing schools are dismantling the framework of a traditional liberal education to gain value on statistics. In some schools, lower achieving students are being 'prevented' from taking GCSEs in virtually all subjects bar Maths and english.

What goes on in 'failing' schools also takes place in 'successful' schools, but to a lesser extent and therefore a less explicit way. Schools that soar above the 30% 5 A–C target will still have weaker students who will take courses equivalent to GCSEs. Though nobody says as much, 'successful' schools will often be those whose leaders are smart and savvy enough to play the system well. Such schools appear to be offering value to their clients and they help the government to play its part in a closed economy of exchange. It is in nobody's interest (apart from the students and broader, wider aims of education) to bring close critical attention to the farcical nature of what is going on. The government need to show that standards are improving. Numbers, as long as one doesn't consider what they relate to, can give this impression.

12.6 Statistics and Slogans

Having considered the performative force of statistics let us refer back to one of the numbers mentioned earlier – 6,000,000 (referring to Jews killed in the Holocaust). I have never seen this number (given the context/s it normally appears in) written

in numerals. This might be due to the fact that the number is somewhat inexact. However, there seems to something obscene about using numerals in this instance that goes beyond their relation to reality, something disgusting about the equation of numbers with such horror. One might even go further and say that writing 6,000,000 out in words sutures over the horror of the Holocaust and that there is something unethical about humanising it in this way. What I am trying to touch on here is that bleakness accompanies numbers/statistics, whether they appear as words or numerals.

This perhaps explains why policies designed to boost the stats are given names like 'Every Child Matters' (a British invention) and 'No child left behind'. Both these phrases have become slogans. 'Every Child Matters' need not mean (and from a humanist perspective should not mean) 'every child must gain added value on their minimum target grade'. There is a sense in which the phrase 'every child matters' in its internal linking carries a humanist quality to it. In contrast 'no child left behind' has a rather more forced quality to it that shows up its' sloganeering – the phrase is adapted from military imagery of leaving no man behind on the battlefield (though there is also perhaps a hint of Hansel and Gretel about it). In at least two respects, these slogans have a suturing effect in that (1) they hide what they are about and (2) they become mantras – they dissolve into the public body. However, importantly, slogans are more vulnerable than statistics even when (or perhaps particularly when) they are married to them. This is due to the iterability of language mentioned earlier – for a slogan to be a slogan it has to be repeated. It will draw on words that have a rich history. Consider the fact that both slogans include the word 'child'. Much of the recent discourse in British schooling had abandoned this word due to its progressive connotations and replaced it with 'student'. Policy makers and politicians were presumably unaware of this word's origins: 'A student, let us remember, is originally a lover (Latin, *studere* – to love)' (Standish, 2005, p. 60). What we are bearing witness to here is the aforementioned 'iterability' of language.

I am going to tell a story to reinforce what is perhaps at stake here. I recently gave a lecture to teacher trainees on 'behaviour for learning'. Prior to this lecture I was asked to refer to the 'Every Child Matters' agenda, as Ofsted inspectors would check that the students had been repeatedly introduced to it. I did not feel too happy about this and grudgingly referred to it in passing. The 'phrase' elicited a groan from an audience who are still quite new to sloganeering. Suddenly I found myself defending the notion that every child matters with a discussion of the fact that when many of the students come across a certain kind of 'child' and that individual will bear little resemblance to the sorts of human being (civil, non-violent, articulate) that they would previously have encountered (many of the students are privately schooled Cambridge graduates), resisting the temptation to write that child off will require considerable mental and emotional strength. Putting things this way introduces the term 'every child matters' into a context that is neither technical nor humanist (in any sentimental sense). Perhaps, in certain respects, this story exemplifies Derrida's account of professing.

> Derrida explores ways in which the idea of profession requires something tantamount to a pledge, to the freely accepted responsibility to profess the truth. The professor enacts this performative continually in her work: what she says is testimony to the truth; as work it is necessarily an orientation to a to-come. The academic work of professing must then be something more than the (purely constative) statement of how things are. (Standish, 2001, p. 18)

Here Derrida's argument is subtle: professing the truth is about orientating one's audience towards something that neither professor nor student (these terms should perhaps be seen treated in a rather abstract way) can necessarily predict in which the constative (the truth) will always be performed as the to-come (an aspect of the future) rather than a statement of that which is secure or originary. To take the truth for granted is to ignore the performative aspects of language that are bound up with its iterability (language's otherness to itself). There is therefore both an active and passive dimension to professing that is both backward and forward looking. We cannot speak from nowhere, but where we 'are' will undergo transformation – we are at home and not at home. The somewhere that we speak from must in some way reveal the mark of the past and the future.

There is an ethical dimension to this orientation to the 'to-come' which goes beyond a simple celebration of openness. 'Presence' makes other ways of seeing invisible. To be open to what is 'to come' is 'to do justice to the "other" of presence'. This shows that 'the point of deconstruction is not negative or destructive but first and foremost *affirmative* . . . It is an affirmation of what is excluded and forgotten; an affirmation of what is *other*' (Biesta, 2009, p. 395). The emergence of this otherness, what Derrida calls 'the arrivant' is unforeseeable and 'impossible': 'For Derrida "the impossible" is not what is *im*possible but what cannot be foreseen as a possibility' (ibid). We are called on to show 'hospitality' towards the impossible other that waits at the door. If we ignore this, communication can only be viewed as a 'strictly mechanical, a strictly calculable and predictable process' (Biesta, 2009, p. 399), which is exactly what performativity encourages us to see it as.

12.7 Conclusion

Some readers may feel that what has been left out of the account, as it is presented, is direct engagement with the ways in which statistics are generated. After all, statistics only create the 'impression' that they operate in accordance with a false metaphysics of signification. What statistics stand in for or 'signify' is open to the rich iterable processes of language. Though carrying out research in to the mechanism of statistics creation is a worthwhile task, what I have tried to show is that the performative aesthetic force of statistics encourages us to refrain from looking below the surface – statistics are something we 'show'. This is because what we encounter with statistics is a 'surface' in which depth is not always already inscribed. This is true of all numbers.

We might also note that advocates of performativity may mount a defence in regard to some of the claims made during the section 'fixing the stats'. The fact that some schools do not possess cohorts expected to achieve 30% 5 GCSEs is no excuse for failure. Schools in the lowest income bracket have managed this, so every school should be capable of success. If schools cannot attract teachers then it is their own fault for being so ineffective and if physical education teachers are the only people who command enough respect to teach literacy and numeracy skills, then why not let them do it (teachers should be facilitators shouldn't they?). Are history, geography and physical education that necessary to the future careers or general functioning of students' lives or indeed the functioning of society? Is the notion that a liberal education will be the best kind of education either coherent or sufficiently up to date? Perhaps it is better that students take courses more in keeping with either their dispositions or ability levels. Maybe insisting that some students take vocational courses helps them to save them from themselves. That way, at least they might get some form of qualification to put in their portfolios at the end of their schooling. If vocational courses really are easier to do, well, then, this problem can be ironed out by making them more difficult. There is an ethical aspect to this kind of defence. Schools should be accommodating their 'clients' needs and demands. It can be shown that 'failing' schools are failing their students.

We might note here that this defence is predicated on a number of assumptions regarding (1) what good teaching is, (2) what is good for students, (3) what is good for society and (4) the importance of being 'contemporary' or 'fashionable'. Such assumptions mark the metaphysics of presence as it inscribed through performativity. They are haunted by what may arrive – the 'arrivant' is also a 'revenant'. Let us also note that easy equivalent GCSE courses have been offered for years now. It seems unlikely that they will be made more difficult.

Lastly let us note that statistics need words (words, that is that are not numbers) to humanise them. Though this often takes the rather inhuman form of the 'slogan', slogans like all uses of language will contain elements that are iterable, their past and future contexts belong to them. The suturing effects of the marriage of statistics and slogans will not prevent language from bleeding. This chapter celebrates such bleeding not in the name of humanism, but rather in the name of the 'inhuman' other that may arrive at any time.

References

Bearn, G. (2000). The university of beauty. In P. Dhillon & P. Standish (Eds.), *Lyotard: Just education* (pp. 230–258). London: Routledge.

Biesta, J. J. (2009). Witnessing deconstruction in education: Why quasi-transcendentalism matters. *Journal of Philosophy of Education, 43*, 391–404.

Derrida, J. (1988). *Limited Inc*. Evanston, IL: Northwestern University Press.

Derrida, J. (1997). *Of grammatology*. Baltimore: John Hopkins University Press.

Lyotard, J. -F. (1984). *The postmodern condition*. Manchester: Manchester University Press.

Standish, P. (2005). Towards an economy of higher education. *Critical Quarterly, 47*, 53–71.

Standish, P. (2010). Calling education to account. In P. Smeyers & M. Depaepe (Eds.), *Educational research: The ethics and aesthetics of statistics* (pp. 207–216). Dordrecht: Springer.

Chapter 13
A Bubble for the Spirit Level: Metricophilia, Rhetoric and Philosophy

Richard Smith

We must lobby for energetic, unencumbered collection of epidemiologic data—from surveillance to association studies to clinical trials—on the basis that such data will bring future benefits. Would the public want to give up avian influenza surveillance? Are our communities happy to continue to add chemicals to the environment and never conduct human studies to assess their potential toxicity? Would people vote for public policies and accept medical interventions based solely on results in rats?

Ness and Rothenberg (2007)

On you go now! Run, son, like the devil
And tell your mother to try
To find me a bubble for the spirit level
And a new knot for this tie.

Seamus Heaney, *The Errand*

13.1 Introduction: The Measurement of Love

In this chapter I shall be critical of the craving for statistics and measurement in social science, the faith that they always and without question give us the answers on which to base policy in education as in other areas of life: it is therefore important to emphasise from the start that I am not against statistics, only the craving for them and unreflective faith in them. It is much the same as with the more familiar case of science. Without science our lives would be vastly less safe and less comfortable. It is common to instance dentistry as a knock-down example; and when I travel in an airplane I expect the science of aerodynamics to get me there in one piece. But this does not justify scient*ism*, the colonising by science of every other form of thought and the assumption that whatever problem we have, the solution can only be in some way a scientific one.

R. Smith (✉)
University of Durham, Durham, UK

P. Smeyers, M. Depaepe (eds.), *Educational Research: The Ethics and Aesthetics of Statistics*, Educational Research 5, DOI 10.1007/978-90-481-9873-3_13,

'Statisticism', on the analogy of 'scientism', has recently been proposed to capture the 'obsessive devotion to, or veneration for, statistical evidence as the *sine qua non* of genuinely scientific knowledge' (Lamiell, 2009). However, the word does not trip off the tongue. I suggest *metricophilia*, to do justice to the fact that statistics are only part of the fascination with measurement more generally; there is more of a suggestion in the word of something pathological (compare, say, necrophilia). I found it interesting to type 'love of measurement' into Google, partly to see if 'metricophilia' has any currency. Most of the first few pages return sites devoted not to the love of measurement, but to the measurement of love. The unit of measurement of love is 'The number of times you think of your loved one during the day. Also kisses are a good way to tell' (www.answerbag.com). A book by Oliver C.S. Tzeng, *Measurement of Love and Intimate Relations*, offers in its subtitle *Theories, Scales, and Applications for Love Development, Maintenance, and Dissolution*.[1] An article tells us that 'The silence surrounding the psychological investigation of love has been broken. Not only is love being examined from a theoretical framework but better empirical tools are being developed and used' (Elkins & Smith, 1979, p. 7). And so on.

In what follows, I analyse three prominent and influential cases of metricophilia, all responses to the well-known failure of increasing affluence to bring higher levels of well-being in the developed societies of the (largely western) world. These responses, in the order in which I shall discuss them, are, first, a short newspaper article on happiness by Richard Layard, sometimes referred to as the UK 'happiness czar'. The second is the commission on the Measurement of Economic Performance and Social Progress (often referred to as the Stiglitz Commission, after its chair, Joseph Stiglitz). The third response is that of Wilkinson and Pickett in their widely noticed book *The Spirit Level: Why More Equal Societies Almost Always Do Better* (2009). This is an attempt by two epidemiologists to correlate inequality of income in developed and better-off countries of the world with all kinds of problems such as drug and alcohol abuse, mental illness, lower levels of educational performance, obesity, violence and murder.

In all three cases I shall argue that their prevailing metricophilia leads to oversimplification and reductionism, and that there is a strong tendency to ignore or underplay crucial philosophical questions in the faith that better metrics and statistics, and more of them, will tell us all that we need to know. I also identify, in the first case I shall analyse in particular, the use of various and luxuriant forms of rhetoric: a phenomenon that is all the more odd because it is a prime example of the sophistry that the dispassionate objectivity of metrics, a favourite child of the Enlightenment, promised to dispel.

13.2 Levelling Happiness

Numerous studies attest to the failure of increasing affluence, in the developed societies of the world, to bring increased well-being. Much of this work was stimulated by Richard Easterlin's 1974 paper, 'Does Economic Growth Improve the Human Lot? Some Empirical Evidence', and the paradox that rising national wealth brings

no increase in happiness is often referred to as the Easterlin paradox (see also, in particular, Offer, 2006). Wilkinson and Pickett (2009, p. 3), to whom I return in detail below, offer as good a summary as any:

> It is a remarkable paradox that, at the pinnacle of human material and technical achievement, we find ourselves anxiety-ridden, prone to depression, worried about how others see us, unsure of our friendships, driven to consume and with little or no community life.

Also familiar, I take it, are the studies that track the effects on children of this sad state of affairs (see, for example, UNICEF, 2007). We thus start from the momentous insight that wealth does not always bring happiness.

This is the context in which Richard Layard is committed to the project of measuring happiness so that we can see what, if not wealth, brings happiness in its train. This project of course commits him to the view that all kinds of pleasure, satisfaction, flourishing, fulfilment, etc. are commensurable in terms of his favoured common currency. If they were different they could not be compared and ranked. A certain process of levelling, a flatness, a lack of diversity, is therefore implied. A newspaper article written by Layard, from *The Guardian*, 13 September 2009,[2] has the title 'This is the greatest good', and its subtitle reads 'We have only one true yardstick with which to measure society's progress: happiness'. Even the title thus asserts that the problem is about just one quality, happiness, rather than showing a willingness to explore the complexities of human well-being and flourishing. We should note too that the claim that there is 'only one *true* yardstick' (for this is clearly where the emphasis is to be placed) manages rhetorically to smuggle in the assumption that there can indeed be a standard against which different kinds of happiness can be measured and compared: 'everyone takes this for granted', the claim seems to imply, 'the only question is just which is the *best* standard or yardstick'. Layard begins the article,

> What is progress? That is the question President Sarkozy of France has posed to a distinguished commission. It is exactly the right question, and the future of our culture depends on the answer.

This is throughout a heavily rhetorical text, and we need to employ strategies that go beyond the usual academic range in order to resist its spell. (Nietzsche (1990) recommends, 'happy distrust, pleasure in mockery' as signs of critical health (*Beyond Good and Evil,* 4, p. 154).) A president, then, has asked a distinguished commission; and his question, apparently consisting of three stark and simple words, has the simplicity of innocence and authority at once. We see him, sombre perhaps in his double-breasted suit, rising to ask his question of the balding and grey-headed figures seated on three sides of a vast, polished table. 'It is exactly the right question', Layard writes, and it is hard not to imagine the distinguished commissioners nodding in agreement. The king in his simple wisdom, born of generations of noble lineage, asks his deceptively simple question. We, the readers of the newspaper article, know our humble place. We are in no position to disagree or criticise, particularly since 'the future of our culture depends on the answer'.

Layard tells us that 'progress must be measured by the overall quality of people's lives' and not by GDP (gross domestic product). We must focus on 'how people feel: are they happy and contented?' This is the 'overarching good' that we need to

identify, and twice in the first three paragraphs Layard invokes the philosophers of the Enlightenment in support of his claim that the supreme good is happiness. The ghosts of Bentham and John Stuart Mill loom behind the shoulders of the distinguished commission. The 'noble philosophy of the Enlightenment' is cited for the third time, in support of the view that 'every human being wants to be happy'. A sentence that falls neatly into two iambic pentameters suggests that the writer, or speaker, might come from Shakespeare.

> So it is time to reassert the noble
> Philosophy of the Enlightenment.

The metre – a different kind of metric this, of course – requires us to pronounce 'noble' as one word (this is common in Shakespeare, e.g. 'Your wife, your son, these senators, the nobles', *Coriolanus* III. ii). After the slightly blurred effect of this line ending, the second pentameter, 'Philosophy of the Enlightenment', which is perfect in the coincidence of ictus and accent, strikes the reader with redoubled force. The Philosophy of the Enlightenment is indeed reasserted, as if it were indubitable.

It follows from our acceptance of this philosophy, apparently, that,

> progress is measured by the overall scale of human happiness and misery. And the right action is the one that produces the greatest happiness in the world and (especially) the least misery.

Of this Layard states, 'I can think of no nobler ideal'. There is a curious sequence of thought here running from President Sarkozy through the distinguished commission – surely by now they must be a Distinguished Commission – to the Enlightenment and its noble philosophers, and thus to their, and Layard's own, noble ideal. From President and Distinguished Commission, with their glimpses of ermine and the nobility which we have been rather heavily directed to notice, we are invited somehow to arrive at the democracy of the greatest happiness principle, in which, as Layard notes, 'everybody counts equally'.

Layard proposes a 'campaign for the Principle of the Greatest Happiness'. The last sentences of this paragraph, and the whole of the following one, repay careful reading:

> We desperately need a social norm in which the good of others figures more prominently in our personal goals. Today's excessive individualism removes so much of the joy from family life, work and even friendship.
>
> There have been objections to this principle, which can be answered. But even some sympathisers prefer the term 'flourishing' to 'happiness'. Why is this? I fear it reflects a streak of puritanism – that happiness ought to come from some sources rather than others. But in the world's great literature, people discuss whether they are happy, not whether they are flourishing. When we discuss the quality of life, we should use the words that people use to describe themselves.

The first oddity is that the 'excessive individualism' of our time is taken as a reason why we should sign up for a campaign (or principle, or philosophy) that treats the happiness of communities as an aggregate of the happiness of individuals. The second oddity consists in Layard's repudiation of the idea of 'flourishing'. This is said to reflect a 'streak of puritanism': the thought being that a particular sort of person,

associated historically with the seventeenth century ascetics who were the regicides of the English Civil War as well as closers of theatres and general kill-joys, dislikes happiness because it sounds too, well, *happy*. The interesting point is that by criticising 'flourishing' Layard manages to add plausibility to the assumption that there really is a single principle, one aim that we all pursue, only just not this one. The third oddity lies in the final two sentences of the second paragraph above. To start with, it would not be difficult to give many examples of people in 'great literature' talking and reflecting on how their lives are going in terms other than those of happiness. It is in fact one of the distinctive merits of the arts, particularly literature and film, that they offer a richer and more fine-grained treatment of human well-being and suffering than such thin terms as 'flourishing' and 'happiness' contain. *Anna Karenina,* for instance, may start with a famous and striking declaration about happiness and unhappiness, but much of the novel turns on the particular kinds of contentment, complacency or self-deception in which the principal characters find what we might call, unsatisfactorily, their happiness.

Two examples from Tolstoy's novel demonstrate this in an illuminating way. The first is from Part I, Chapter 10. Oblonsky is having dinner with Levin, who comes from the countryside and is ill at ease in the luxurious city restaurant. Levin says,

> . . . there is something monstrous about what we are doing now. In the country we try to get our meals over as quickly as possible, so as to get on with our work, but you and I are doing our best to make our dinner last as long as we can; we therefore have oysters . . .
>
> Why, of course, objected Oblonsky. But the whole aim of civilisation is to make everything a source of enjoyment.
>
> Well, if that is so, I'd rather be a savage.

The second example is from Part III, Chapter 13 Karenin has been considering how to respond to his wife's unfaithfulness.

> Though Alexey Alexandrovitch [i.e. Karenin] was perfectly aware that he could not exert any moral influence over his wife, that such an attempt at reformation could lead to nothing but falsity; though in passing through these difficult moments he had not once thought of seeking guidance in religion, yet now, when his conclusion corresponded, as it seemed to him, with the requirements of religion, this religious sanction to his decision gave him complete satisfaction, and to some extent restored his peace of mind. He was pleased to think that, even in such an important crisis in life, no one would be able to say that he had not acted in accordance with the principles of that religion whose banner he had always held aloft amid the general coolness and indifference. As he pondered over subsequent developments, Alexey Alexandrovitch did not see, indeed, why his relations with his wife should not remain practically the same as before. No doubt, she could never regain his esteem, but there was not, and there could not be, any sort of reason that his existence should be troubled, and that he should suffer because she was a bad and faithless wife. 'Yes, time will pass; time, which arranges all things, and the old relations will be reestablished', Alexey Alexandrovitch told himself; 'so far reestablished, that is, that I shall not be sensible of a break in the continuity of my life. She is bound to be unhappy, but I am not to blame, and so I cannot be unhappy'.

Karenin cannot see, or does not want to see, that a life based on seeking 'peace of mind' and in which the comfortable ways of his existence are not troubled is a life of complacency rather than anything worth the name of happiness; and the casuistry

of the final sentence of his reflections, with its telling double negative, directs us to the thought that people's use of the word 'happy' to describe or think of themselves and their hopes may be a large part of their problem and not, as Layard supposes, something in which we are required to follow them.

13.3 Homogenising Well-Being

To sum up, Layard's project is committed to the measurement of happiness. The online version of his article helpfully contains a hyperlink: 'there has been a huge increase in our ability to *measure happiness* and in our knowledge of its causes'. The link is to a BBC webpage with the title 'The science of happiness'. This tells us, 'scientists say they can actually measure happiness . . . Neuroscientists are measuring pleasure. They suggest that happiness is more than a vague concept or mood; it is real. . . . Social scientists measure happiness simply by asking people how happy they are'. Yet, this scientific project, that will end in metrics and statistics, requires, as we have seen, a good deal of specious argument and the support of rhetorical trickery to make its opening moves. But it is, in a sense, worse than this. The problem with assuming that we can measure happiness and unhappiness on a single scale is that by thinking of all the complex forms of human delight as happiness, and all forms of misery, loss of meaning, alienation, failure to flourish, absence of sense of self-worth (etc.) as unhappiness, we risk missing the connections between individual experience and the economic and social conditions that may lie behind it. The hedonic dimension becomes all; the causes of unhappiness are seen as lying in the way a person views the world rather than in the world itself, that is the world of wealth distribution, economic and social class, etc. It is thus unsurprising to find that Layard proposes that the solution to the crisis of unhappiness lies in a major expansion of CBT (Cognitive Behavioural Therapy), described – naturally – as an 'evidence-based approach' (Layard, 2005).

Layard's response to the 'crisis of unhappiness' can be set out schematically as follows. It shows

(i) a tendency to view the problem as about just one quality, happiness, ignoring the complexities of human well-being and flourishing;
(ii) a tendency to look for a single *cause* for the general malaise, even when the diversity of the kinds of unhappiness is acknowledged;
(iii) a faith that more and better metrics and statistics is what is required.

These points are interconnected. Points (i) and (ii) are similar, though clearly not the same: Point (i) asserts that you have a problem with happiness, rather than, say, with being underpaid and subjected to many forms of stress, while (ii) may accept that happiness, like unhappy families, comes in many different forms, yet has at root a single source or cause. The ambition to measure happiness and the lack of it will be facilitated by the assumption that happiness can be measured on a single scale:

thus (iii) is connected with (ii), but even more so with (i). We shall see this schema working in similar ways in the other two responses.

The report of the Stiglitz Commission,[3] which was the result of President Sarkozy's initiative (above), and is also hyperlinked from Layard's article, gives further plentiful evidence of what we might call metricophilia. It is a long document: I confine myself to the executive summary. Para. 2 reads as follows:

> . . . statistical indicators are important for designing and assessing policies aiming at advancing the progress of society, as well as for assessing and influencing the functioning of economic markets. Their role has increased significantly over the last two decades. This reflects improvements in the level of education in the population, increases in the complexity of modern economies and the widespread use of information technology. In the 'information society', access to data, including statistical data, is much easier. More and more people look at statistics to be better informed or to make decisions. To respond to the growing demand for information, the supply of statistics has also increased considerably, covering new domains and phenomena.

There is a strong sense here that, as J.-F. Lyotard (1984) predicted, what cannot be encoded in the bits and bytes of information technology comes to be seen as of only marginal importance. More complex societies – or at any rate economies – require ever more statistical indicators, it seems; but we may wonder whether these will do justice to the complex societies they are required to measure. I return to this point below.

The central shift in the work of the Commission is away from measuring economic production. Yet, there is some ambivalence concerning what the shift moves us to. Sometimes it seems to be towards measuring income and consumption, and sometimes towards measuring well-being. It is the move to an interest in well-being rather than GDP that is often hailed as the crucial step forward the Commission has taken. This move is encapsulated in recommendation 1: 'When evaluating material well-being, look at income and consumption rather than production'. The phrase 'material well-being' encapsulates the ambivalence, for sometimes talk is of well-being *tout court*, while at other times it is the 'material' that is central. On the one hand, traffic jams may increase GDP through an increase in the consumption of fuel, but they hardly improve quality of life (para. 5). On the other hand, the paragraphs that follow recommendation 1 speak exclusively of material living standards (paras 22 and 23: the next two paragraphs speak of 'living standards', but by now it is clear that this is shorthand).

A bold headline before para. 28 declares that 'Well-being is multi-dimensional'. The Commission identifies a number of key dimensions that 'should be considered simultaneously': material living standards; health; education; personal activities including work; political voice and governance; social connections and relationships; environment, present and future; insecurity, economic and physical (para. 28). Many of these 'are missed by conventional income measures' (ibid.). Here, it might seem, we have a welcome acknowledgement that people value an incommensurable diversity of goods. However, recommendation 6, which follows this paragraph, begins by insisting that 'Quality of life depends on people's *objective* conditions and capabilities' (my emphasis). The next paragraph (para. 29) writhes uncomfortably

between admitting that issues of quality of life are inevitably very subjective (and by implication diverse), and the assumption that more and better statistics will bring the answer. I set out some of the principal statements, either *verbatim* or very nearly so, in tabular form.[4]

Objective measures	Both	Subjective, diverse values
Go beyond people's self-reports and perceptions	Measuring the features of quality of life requires both subjective and objective data	(By implication) Include self-reports and perceptions
Include measures of 'functionings' and freedoms		What really matters . . . is people's freedom to choose the life they value
The challenge is to identify gaps in available information and to invest in statistical capacity		Such choice is a value judgement rather than a technical exercise

The general drive towards stipulative homogeneity, however, cannot be resisted, and it emerges with full force in para. 32, 'While assessing quality-of-life requires a plurality of indicators, there are strong demands to develop a single summary measure'. This will 'aggregate across quality-of-life dimensions' (recommendation 9; the sudden move to hyphenisation seems itself to confer a kind of unity). Subjective measures of the quality of life should also be considered (para. 33): this dimension of well-being might include 'happiness, satisfaction, positive emotions such as joy and pride, and negative emotions such as pain and worry' (para. 34). It is immediately assumed that there can be 'quantitative measures of these subjective aspects' (ibid.), which are then in the next sentence referred to as 'subjective *measures*' (my emphasis). No doubt these confusions reflect the need for the Commission to please its diverse members. But it is difficult to see how a 'single summary measure' can fail to do violence to the extraordinary range of goods that people in fact value, to the many and incommensurable dimensions of human happiness and well-being. Thus, the Stiglitz Commission, like Richard Layard, threatens to contribute to the general levelling, the homogenising of diversity: what Jean-François Lyotard (1984) calls a totalising 'nostalgia of the whole and the one'.

13.4 In Search of the Bubble

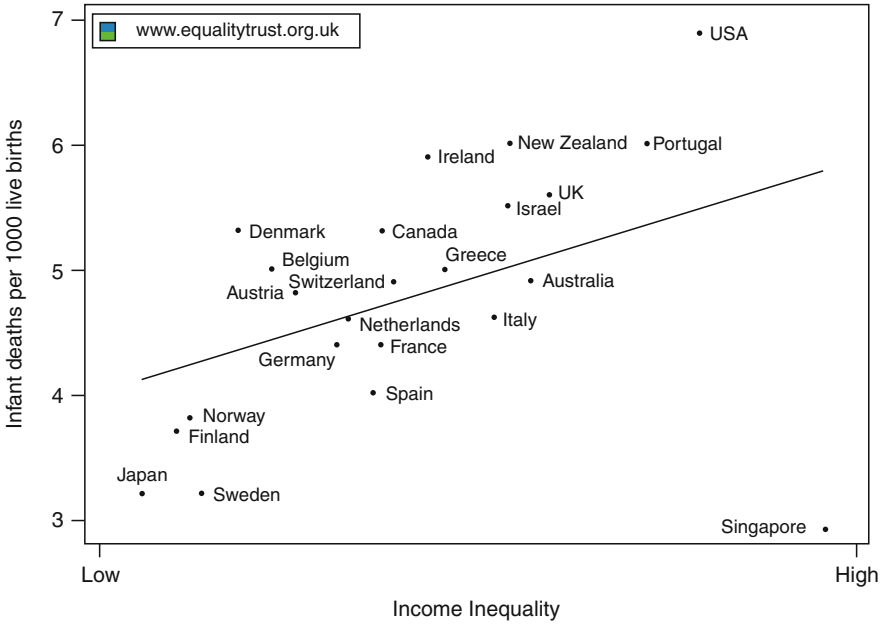
The attractions of epidemiology – a branch of statistics usually understood as the study of factors affecting the health and well-being of whole populations – are powerful. Epidemiology is usually seen as reinstating the significance of the political and social policy dimension of our lives, in the face of studies that over-emphasise individual pathology (and thus solutions such as CBT). The spectacular successes of epidemiology include the identification of the cause of cholera by John Snow in

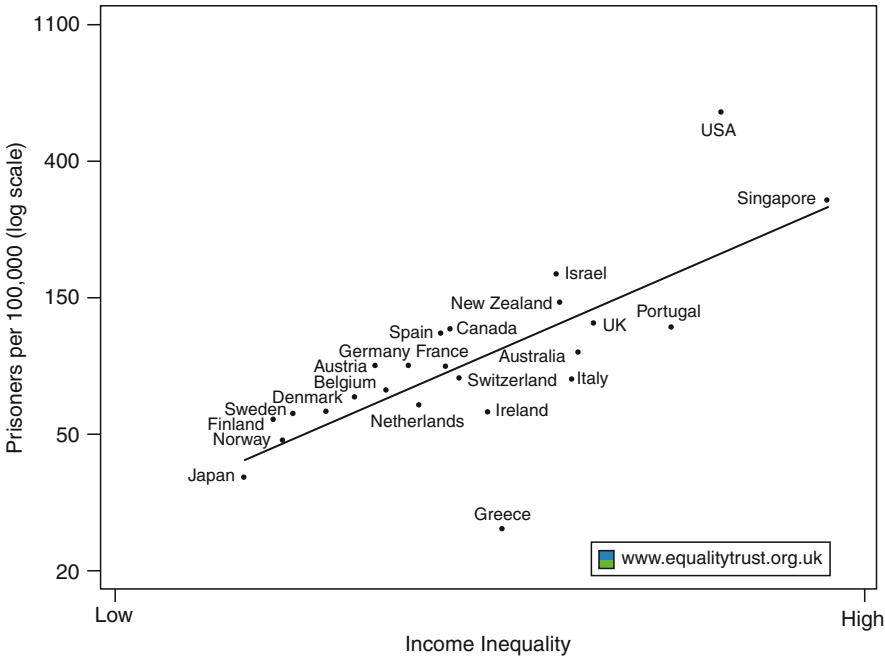
the nineteenth century and the establishment of the connection between smoking and lung cancer in the twentieth. There are well-known criticisms of epidemiology, such as that methodological errors tend to result in 'false positives', that epidemiologists are naïve or complacent about methodology, and that they do too little to counter over-interpretation of epidemiological data by the media and the public (Feinstein, 1989). There is also a line of thought that, as illustrated by the quotation from Ness and Rothenberg prefixing this chapter, epidemiologists frequently overstate the credentials of their discipline and offer in their place a distinctive form of rhetoric.

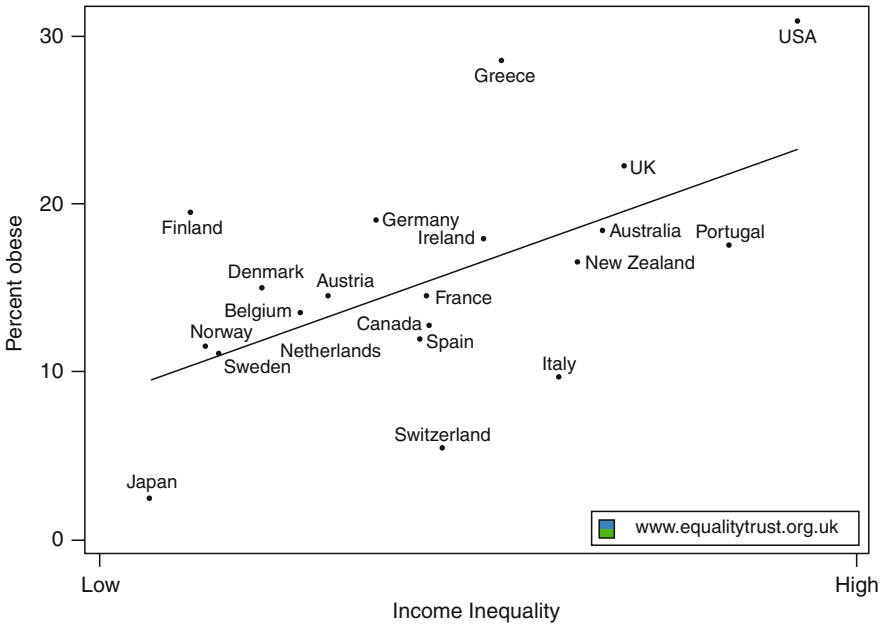
In what follows I attempt to analyse the claims of *The Spirit Level: Why More Equal Societies Almost Always Do Better* (Wilkinson & Pickett, 2009). I focus on this book for several reasons. First, its emphasis on problems that can be remediated in the social world, as opposed to in people's heads by way of CBT or other therapies, is on principle very welcome. Second, however, it seems to have the capacity to defeat the critical faculties of many reviewers (for example 'a bank of evidence against inequality that is impossible to deny', 'its unarguable battery of evidence': both from the *Guardian*, 14 March 2009, *Review* p. 6). Third, it leaves unsolved crucial questions about causality: about just how inequality 'gets under the skin' to bring about the social and other effects that Wilkinson and Pickett identify. To ask questions about causality is to search for the bubble of this chapter's title.

Wilkinson and Pickett started with the question why health within any population typically deteriorates the further down the socio-economic scale you go. The differences are at first sight puzzling since they cannot be explained in terms of gross national product (GNP) or spending on health care. Some countries that spend heavily on health care (such as Norway) do very well in terms of physical health and life expectancy, while others that also invest heavily in health care (such as Switzerland) do significantly worse. Crucially, it is not only the poor whose health suffers. In unequal countries the health of the rich, as well as their levels of educational attainment and all the other variables investigated, suffers too. The explanation given in *The Spirit Level* focuses on inequality: both poor and rich suffer, in unequal (and developed, broadly better-off) societies, from a range of problems more than people in equal ones. The problems include drug and alcohol abuse, mental illness, lower levels of educational performance, obesity, violence and murder. More people are sent to prison; infant mortality levels are higher; people are less inclined to trust their fellow citizens; there is less concern, as reflected in government spending, for the developing world.

The writers' method is to plot variables against levels of inequality in developed nations, thus presenting evidence in the form of scatter diagrams (sometimes called 'scattergrams' or 'scatterplots'). They then draw a computer-generated regression line through the co-ordinates, which always inclines in a way that reveals the guilty hand of inequality. This they describe as a 'best-fit line', which they only include 'if the relationship would be very unlikely to occur by chance' (p. xv). There are nearly 40 such scatter diagrams (as well as numerous other graphs and charts). I include several here to illustrate the approach.[5]


www.equalitytrust.org.uk
Infant deaths per 1000 live births
7
6
5
4
3
USA
New Zealand
Portugal
Ireland
UK
Israel
Denmark
Canada
Belgium
Greece
Switzerland
Australia
Austria
Netherlands
Italy
France
Germany
Spain
Norway
Finland
Japan
Sweden
Singapore
Low
High
Income Inequality


Prisoners per 100,000 (log scale)
1100
400
150
50
20
USA
Singapore
Israel
New Zealand
Portugal
UK
Spain
Canada
Germany
France
Australia
Austria
Switzerland
Italy
Belgium
Denmark
Sweden
Finland
Norway
Netherlands
Ireland
Japan
Greece
www.equalitytrust.org.uk
Low
High
Income Inequality


30
20
10
0
Percent obese
USA
Greece
UK
Finland
Germany
Ireland
Australia
Portugal
New Zealand
Denmark
Austria
France
Belgium
Canada
Norway
Spain
Netherlands
Sweden
Italy
Switzerland
Japan
www.equalitytrust.org.uk
Low
High
Income Inequality


80
60
40
20
0
Most people can be trusted (% agreeing)
www.equalitytrust.org.uk
Sweden
Denmark
Norway
Finland
Netherlands
New Zealand
Japan
Switzerland
Australia
Canada
Germany
Spain
USA
Austria
Ireland
Belgium
Italy
UK
Greece
Israel
France
Singapore
Portugal
Low
High
Income Inequality

It is an interesting question whether these diagrams themselves constitute a kind of rhetoric. The vividness with which evidence is presented – why, there are the Scandinavian countries again! – has the effect of silencing awkward questions. But there are such questions to be asked. For instance, is it possible that the data are skewed by 'outliers', particularly Japan at one extreme and the USA at the other? The assumption must be that all the countries that appear have relevantly similar societies, or there is no basis for comparison. Yet, Japan is a very different society from the USA, being characterised for instance by Shinto traditions, strong family ties, deference (particularly to the older generation) and conformity. It is different again from the Scandinavian societies that model the correlation between equality and other social goods. If Japan and the USA are removed from the graphs then the pattern of co-ordinated points looks far less regular. Wilkinson and Pickett might respond that this actually strengthens their point. It doesn't matter *how* you achieve income equality: however you do so, the other benefits flow.

Rather more tellingly, the causal relationship here is obscure. If more equal societies lock up a smaller proportion of their population in prison, is this less because greater equality somehow *produces* such an outcome and more because the kind of society that values high levels of equality is also likely to be cautious about the effects of imprisonment on relatively minor offenders, along with being more enthusiastic about the promotion of green initiatives such as recycling (see the scatter diagram on p. 228)? This point is made sharply by one reviewer, who concludes that the book's analysis is 'diminished by the sense that the authors have simply scoured the data for signs of malignancy in unequal societies' (Reeves, 2009).

Wilkinson and Pickett are well aware that the causal relationship needs to be illuminated. They address it directly in Chapter 3, 'How inequality gets under the skin', and return to it at intervals throughout the book, especially in Chapter 13, 'Dysfunctional societies', and its subsection 'Causality'. In general their answers turn on the idea of social status and self-esteem. Unequal income leads to unequal status, and in a world where people are alert to and anxious about where they are positioned on the social ladder, mental and physical health are both affected by this anxiety: self-esteem suffers. Where the ladder or hierarchy is particularly steep the middle classes suffer no less than disadvantaged groups.

A highly inconvenient fact here, which Wilkinson and Pickett fully acknowledge, is that over the time-scale under consideration, self-esteem also 'showed a very clear long-term upward trend. It looked as if, despite the rising anxiety levels, people were also taking a more positive view of themselves over time' (p. 36). Surely, it would seem, anxiety about status should be reflected in lower, not higher, self-esteem. Wilkinson and Pickett's solution is to distinguish 'healthy' self-esteem from the defensive kind found in those prone to violence, racism and insensitivity to others. This latter kind is fragile and more akin to 'whistling in the dark' (p. 37); we might compare Ruth Cigman's discussion of 'psychological fraudsters' (Cigman, 2004). In the context this looks like a rather desperate strategy on Wilkinson and Pickett's part to save the explanation in terms of status. At the least we need some careful exploration of psychological self-aggrandisement (self-promotion, bravado, etc.) and its opposites (self-deprecation, humility, diffidence, modesty). Wilkinson

and Pickett touch on this in a short section (pp. 44–45), noting that the former qualities are more prevalent in the (inegalitarian) USA while the latter are found in (egalitarian) Japan. But again it is not clear which way the causal relationship works. Have the Japanese created a more equal society because their culture traditionally values modesty, or is it rather the case that in such a society self-promotion is naturally deprecated as an offence against the general egalitarian outlook, which is respectful of people's sensitivities over status?

A second part of Wilkinson and Pickett's answer concerns trust. Inequality corrodes trust. In more income-equal states of the USA more people agree that 'most people can be trusted'. Sweden has high levels of trust, Portugal low levels (see last graph above). People with higher levels of trust live longer (p. 57); trust is crucial and may be a matter of life or death in disasters such as Hurricane Katrina (ibid.). But, again, there is no clear conclusion concerning causality. Where we think of our neighbours as trustworthy fellow-citizens, people like us, we may no doubt be more inclined to support measures that reduce inequality. A sense of human solidarity flinches from exploitation and major disparities of income. On the other hand, where such disparities exist we may view our fellow-citizens as 'them', the feckless, untrustworthy poor who are always with us. Wilkinson and Pickett admit that their argument for causality could be stronger: 'In this chapter, we have shown that levels of social trust are connected to income inequality, but of course showing a correlation is not the same thing as showing causality' (p. 61). They write, 'In summary, we can think of trust as an important *marker* of the ways in which greater material inequality can help to create a cohesive, co-operative community' (p. 62, my emphasis). The word 'marker' here serves as an admission that the causal argument is shaky.

It is because the argument is shaky, I take it, that towards the end of the book Wilkinson and Pickett include speculative sections on evolutionary psychology, mirror neurons and oxytocin. For example, *Homo sapiens* is descended from two different species of ape: chimpanzees and bonobos. Chimpanzees (to put it briefly) are status-conscious and hierarchical, while bonobos are more interested in sex, using sex to defuse conflicts and tension. Humans, apparently, have inherited the bonobo section of DNA that regulates social and sexual behaviour. We are no doubt meant to conclude that trusting and caring are more natural for us human beings, rather than to wonder why we have (naturally, it might seem) built such hierarchical societies, or why we, unlike the bonobos, do not 'engage in sexual activity – including mutual masturbation – frequently and in any combination of sexes and ages' (p. 201).

13.5 Conclusion

The important general point to make here, however, takes us back to the earlier sections of this chapter and my complaint that the prevailing metricophilia has the effect of levelling and homogenising: of requiring us to think of all variations of

self-esteem, happiness or well-being as essentially forms of the same thing. The kind of work that needs to be done here is at least as much philosophical as it is statistical. We need to think more carefully about the kinds of self-esteem that are built on a realistic appraisal of the world, and the kinds that consist of little more than a transient 'feeling good', brought about perhaps by psychological manipulation (Smith, 2002). Happiness, in turn, is neither monolithic nor an unalloyed good: the more lasting and profound pleasures do not bring happiness of the same kind as moments of joy or the shorted-lived delight caused by eating a good meal (or ingesting other substances). The higher pleasures need to be distinguished from the lower ones, and that work is irreducibly philosophical, as John Stuart Mill taught us (Mill, 1863). As for trust, that other crucial intermediary in Wilkinson's and Pickett's account, we need an analysis of trust that at the least allows us to make a distinction between trust and the various kinds of scepticism. Without scepticism there would have been no Cartesian radical doubt and thought since the seventeenth century would have taken a very different direction. With more scepticism, on the other hand, there might have been no invasion of Iraq. If – to take a topical example from the UK – trust in Members of Parliament and bankers is at a record low, it is not obvious that it is more *trust* that is needed. We find these issues explored in Onora O'Neill's 2002 BBC Reith Lectures (O'Neill, 2002): the exploration is again philosophical, as it is bound to be.

The more we try to address the causal relationships between inequality of income and other elements of our social lives the more we find ourselves trying to understand just what it is – trust, status, self-esteem, solidarity – that 'gets under the skin', as the title of Chapter 3 has it. But the more we seek such understanding the more we see that it is ideas such as trust and so on that are crucial. It is a fuller and more sophisticated insight into these ideas that we need, rather than yet another graph or set of statistics to clinch the matter once and for all. If I conclude that Wilkinson and Pickett's book is best read as a reminder of where more philosophical work needs to be done, that is not meant to be faint praise; it is a way of indicating how epidemiology and philosophy can work fruitfully together. Wilkinson and Pickett thought of calling their book *Evidence-Based Politics* (p. ix): the glitter of statistics and the reassurance that some people find in possibilities of 'objective measurement' must not be allowed to obscure the fact that philosophical thinking is evidence too, and evidence that, on the showing of the examples treated in this chapter, we clearly need.

Notes

1. 'This is the most comprehensive and state-of-the-art work on the measurement of love. It includes descriptions of a theoretical paradigm and two love models, a common mythological framework for theory development and evaluation, an introduction to over 40 theories for love development, maintenance, and dissolution, the integration of research principles and strategies, the compilation of 26 popularly used scales, and illustrations of three empirical research programs in measuring love and intimate relations. The compilation of scales provides information on purpose, function, constructs, psychometric properties, subject response formats,

and actual measurement items'. Retrieved 7 February 2010 from http://www.greenwood.com/catalog/C4273.aspx

2. See http://www.guardian.co.uk/commentisfree/2009/sep/13/happiness-enlightenment-economics-philosophy/print. Retrieved 5 February 2010.
3. Retrieved 2 November 2009, from http://www.stiglitz-sen-fitoussi.fr/en/index.htm
4. The paragraph reads, in full:

> The information relevant to valuing quality of life goes beyond people's self-reports and perceptions to include measures of their 'functionings' and freedoms. In effect, what really matters are the capabilities of people, that is, the extent of their opportunity set and of their freedom to choose among this set, the life they value. The choice of relevant functionings and capabilities for any quality of life measure is a value judgment, rather than a technical exercise. But while the precise list of the features affecting quality of life inevitably rests on value judgments, there is a consensus that quality of life depends on people's health and education, their everyday activities (which include the right to a decent job and housing), their participation in the political process, the social and natural environment in which they live, and the factors shaping their personal and economic security. Measuring all these features requires both objective and subjective data. The challenge in all these fields is to improve upon what has already been achieved, to identify gaps in available information, and to invest in statistical capacity in areas (such as time-use) where available indicators remain deficient.

5. The graphs appear on pp. 82 (infant deaths), 148 (prisoners), 93 (obesity) and 52 (trust).

References

Cigman, R. (2004). Situated self-esteem. *Journal of Philosophy of Education, 38*, 91–105.

Easterlin, R. A. (1974). Does economic growth improve the human lot? Some empirical evidence. In P. A. David & M. W. Reder (Eds.), *Nations and households in economic growth: Essays in honour of Moses Abramovitz* (pp. 89–125). New York: Academic.

Elkins, G. R. & Smith, D. (1979). Meaning and measurement of love: A review of humanistic and behavioral approaches. *Humanist Educator*, *18*, 7–12.

Feinstein, A. R. (1988, March 10). Scientific standards in epidemiologic studies of the menace of daily life. *Science, 243*(4896), 1255–1256.

Gunnell, J. G. (2009). Can social science be just? *Philosophy of the Social Sciences, 39*, 595–621.

Lamiell, J. T. (2009). Why was there no place for personalistic thinking in 20th century psychology? *New Ideas in Psychology.* Retrieved February 5, 2010, from doi:10.1016/j.newideapsych.2009.02.008

Layard, R. (2005). Annexes to 'happiness'. Retrieved February 5, 2010, from http://cep.lse.ac.uk/layard/annex.pdf

Layard, R. (2009). This is the greatest good. *The Guardian, 13* Sept.

Lyotard, J. -F. (1984). *The postmodern condition: A report on knowledge* (G. Bennington & B. Massumi, Trans.). Manchester: Manchester University Press.

Mill, J. S. (1863) *Utilitarianism.* Online. Retrieved February 5, 2010, from http://www.constitution.org/jsm/utilitarianism.htm

Ness, R. B., & Rothenberg, R. (2007, December 12). Critique of epidemiology: Changing the terms of the debate. *Annals of Epidemiology, 17*, 1011–1012.

Nietzsche, F. (1990). *Beyond good and evil* (R. J. Hollingdale, Trans.). London: Penguin.

Offer, A. (2006). *The challenge of affluence: Self-control and well-being in the United States and Britain since 1950*. Oxford: Oxford University Press.

O'Neill, O. (2002). *A question of trust.* Cambridge: Cambridge University Press.

Reeves, R. (2009, April 5). *The Observer.* Retrieved February 5, 2010, from http://www.guardian.co.uk/books/2009/apr/05/spirit-level-richard-wilkinson

Smith, R. (2002). Self-esteem: The kindly apocalypse. *Journal of Philosophy of Education, 36*, 87–100.

Tolstoy, L. (first pub. 1873–1877). *Anna Karenina*. Retrieved February 5, 2010, from http://www.literature.org/authors/tolstoy-leo/anna-karenina/part-03/index.html

UNICEF. (2007). *Child poverty in perspective: An overview of child well-being in rich countries, Innocenti Report Card* 7. Florence: UNICEF Innocenti Research Centre.

Wilkinson, R., & Pickett, K. (2009). *The spirit level: Why more equal societies almost always do better*. London: Allen Lane.

Chapter 14
Calling to Account

Paul Standish

14.1

The playwright Harold Pinter had some fun with the way people speak. Here, in 1958, is Mr. Fibbs, the factory manager, and Mr. Wills, the shop steward—the former anticipating that the men are about to down tools in demand for more pay or better conditions, the latter about to surprise him.

Fibbs sits resolutely at his desk. There is a knock on the door, and Wills enters. Surely the workers must appreciate the new amenities that have been provided. So what is it this time? Wills at first faintly is apologetic. The men are very grateful for the amenities. They just don't like the products. But they are beautiful products, Fibbs remonstrates. He has been in the business a life-time and has never seen such beautiful products. What exactly don't they like? Wills becomes more insistent. There is the brass pet cock, for instance. They don't like the look of the brass pet cock? (Fibbs is flabbergasted.) Then there's the hemi unibal spherical rod end. (Where could you find a finer hemi unibal spherical rod end?) The workers have gone very vicious about the high-speed taper shank spiral flute reamers . . . (Fibbs is shattered) . . . the gunmetal side outlet relief with handwheel, the nippled connector and the nippled adaptor and the vertical mechanical comparator. (No!) They tremble when they even speak about the jaw for Jacob's chuck for use on portable drill. And as for male elbow adaptors, tubing nuts, grub screws, internal fan washers, dog points, half-dog points, white metal bushes—they've taken against the whole lot. They hate and despise Fibbs' cherished parallel male stud couplings. (Surely not my lovely parallel male stud couplings.) And they have gone right off straight flange pump connectors, and back nuts, and front nuts, *and* the bronzedraw off cock with handwheel and the bronzedraw off cock without handwheel! With or without . . . So, asks Fibbs, finally broken: What do they want to make instead? And Wills replies: brandy balls.

P. Standish (✉)
Institute of Education, London, UK

P. Smeyers, M. Depaepe (eds.), *Educational Research: The Ethics and Aesthetics of Statistics*, Educational Research 5, DOI 10.1007/978-90-481-9873-3_14,

Some students' notes on the Web[1] explain that, in *Troubles in the Works*, Pinter is 'using humour and irony': the 'absurd dialogue and situation refers to the inhuman, ridiculous mechanical nature of industry'. The dialogue conveys 'the sense of entrapment that people can feel when they do meaningless work'. The 'ridiculousness symbolises the uselessness of human activity'. The manager is faced with 'dissatisfaction and protest at everything he believes in. This totally unexpected criticism gnaws at the foundations of his life'. Moreover, the 'true intentions' of the two men are never given away.

Such earnestness from such a short sketch! Yet, what this gloss seems entirely to miss is the way that the sketch could scarcely work without the manager's savouring of the names of the machine parts that the workers are so cruelly rejecting. One can imagine his delight at the listing of these in the company catalogue. One can imagine—one can hear!—his fluency in articulating these terms; and the shop steward is no less adept in this—how else would he be credible? During rehearsals for a revival of Pinter's early sketches at the Theatre Royal, Haymarket, in 2007, the writer and actor, Sean Foley, commented, 'You realise that, while he's unique as a writer, he also belongs in a particular line of British comedy ... These are classic comedy sketches, some of them written for revues and cabaret nights, and there's this strain of surrealism—he got there 12 years before Monty Python'.[2] *Trouble in the Works* was first written for television but censored by the BBC in case it fomented trade union unrest. It was only the censored version, ending with 'Brandy balls', not 'Trouble' that made it into the published scripts.

The sketch is an early short work by Pinter, but as he came more to fame and his characteristic style became better known, the critics coined the words 'Pinteresque' and 'Pinterese' not just for his famous 'pauses' but for his particularly facility in capturing something of the rhythms and vocabulary of ordinary speech, including the way that people linger on phrases, light upon a word, reiterate and repeat. As the sketch shows, this includes an appreciation of the kinds of technical language that, in various ways, have come to feature in our ordinary interactions, at work, in our leisure activities and in our everyday understanding of ourselves. Pinter wanted to reveal the unacknowledged poetic impulse in such language, especially given poetry's repression in so much of the culture as a whole (perhaps in Anglophone culture above all). And this impulse is seen to run through the memorisation of bus routes, the menus in a restaurant, the inventory of the machinery of torture and batting averages in cricket. One can scarcely understand the sketch if one does not attend to the alliterations, the rhythms, the repetitions that are, as I said, so obviously savoured.

14.2

The sketch is firmly located in a particular, now largely bygone industrial world. Fibbs' elegy to male elbow adaptors, tubing nuts, grub screws, internal fan washers and dog points evokes a particular kind of engineering practice: even if the sketch is

set in a factory producing small machine parts, one imagines it to be peopled with overalled figures checking the production line, oiling the moving parts and with, somewhere in the background, the skilled craftsman working at the lathe. Fibbs' romanticisation of this is a reflection of something other than the exploitation of labour. In the 50 years since Pinter's sketch was written, the changing forms technology has assumed have realised new possibilities of the poetic. What is the nature of these changes in technology, and how do they cast light on changes in language?

Here there is a bigger story to tell, and Heidegger's story is bigger than most. The common perception that Heidegger is anti-technology can readily be supported by selective quotation from his various writings related to the topic, but this distorts his position in significant respects. Heidegger takes technology to be one mode of revealing of the world, and undoubtedly it is one that enables us to do remarkable things, improving our lives in various respects. The problem with technology is that, unlike other modes of revealing, it has the tendency to colonise other ways of thought and being, ultimately to take over our lives as a whole. The modern tendency to think of difficulties in terms of problems for which we should seek technical solutions is evidence that this is so, and numerous examples of this could be given. It lies behind the tendency to think of practical engagement in teaching and learning, or, say, in social care, in terms of 'interventions'. I do not propose to expand on this here but rather to draw attention to the larger changes in thought and being that Heidegger takes to be at stake, for these have a bearing on how we understand and account for ourselves.

In what sense is technology a mode of revealing, and how does it push aside other possibilities of thought? One feature of a technological understanding that has frequently been highlighted is the way that it tends to amplify one aspect of our perception of the world at the same time as it subdues or hides other modes of awareness and thought. This can be seen easily with the example of the telescope. The telescope brings near what is far away, but the nearness is dependent exclusively upon the sense of sight, while the senses of hearing, touch and smell, which might be fully involved in more immediate, holistic perception, are thereby rendered more or less redundant. Thus, as some aspect of the world is revealed, other aspects withdraw. That withdrawal affects the way that the world is for us.[3]

Heidegger claims that it is necessary to understand what is at stake here in relation to a change from the Greek to the Roman worlds, which is realised most crucially in the difference between Greek and Latin. Thus, the Greek words *physis* and *hyle* translate into the Latin *natura* and *materia* at the cost of a change at the level of the metaphysical. The Latin prompts an understanding of the world in terms of a lifeless array of material *objects* to which a thinking *subject* is juxtaposed. To see the world in terms of *physis*, by contrast, was to see it as inherently dynamic but also dependent for its coming to light on the *logos* of human beings within it. The late paintings of Van Gogh, where the vehemence of the brush marks reveals cornfield and tree and mountain and cloud swirling up in a dynamic unity, charged with a kind of electrical intensity, perhaps convey something of this. But any sense of the dynamic relation between human being and world implied by *logos*, any connotation of the way, the *Tao*, comes to be displaced by a more disembodied and

abstract *ratio*, with its characteristic calculative modes of setting things in order. This tendency of thought reaches a new point of self-consciousness in the modern period with the work of Leibniz. Heidegger devotes a book to the principle of sufficient reason: *nihil est sine ratione* (Heidegger, 1991). The calculative force of *ratio* is explained by Heidegger as a kind of reckoning, and this shapes the modern sense of risk-assessment and accounting. (Heidegger points out, somewhat wryly, that Leibniz was also the inventor of life assurance.) Such thinking not only lays the way for contemporary accounting practices, but determines the way that we account for ourselves (which is to say, how we understand ourselves)—a thought developed most obviously in the writings of Michel Foucault.

In the course of the lecture series that constitutes *The Principle of Reason*, Heidegger quotes the words of Mozart, who we find reflecting on the way that, while riding in a carriage or sleepless at night, a tune would come to him, and with it a harmony, and a development that would unfold as a symphony. He refers also to Hölderlin, to Stefan George and to Angelus Silesius' line, 'A rose is without reason why . . .' Silesius was some 20 years older than Leibniz, and it is perhaps surprising to find that Leibniz refers to him frequently in his writings. These then are gestures to the different, receptive thinking with which Heidegger seeks to expose the false pretensions of the absolute principle: nothing is without reason.

The stakes here could scarcely be higher. Just as we saw that the Greek sense of *physis* was inseparable from the operation of *logos*, it can be said that the manifestation of being depends upon a certain operation of language. Being has a history. In the 1950s, Heidegger takes his account of that history forward by identifying a newly prominent feature of his times. The greatest threat, he can now say, is not the atomic bomb but an atomism in thinking itself. He elaborates this thought specifically in relation to cybernetics, which depends in large part on the development of the binary logic with which Leibniz experimented. It is not difficult to see the plausibility of this in terms of the ways that the methods of ordering that computers generate and facilitate—calculation, databases, spreadsheets—themselves affect how human beings structure their work and their lives as a whole. Prominent in these methods are forms of accounting and projection.[4]

In fact, there is a problem with seeing cybernetics as a specific technological innovation because of its pervasive effects on technology as a whole; this is a change that manifestly affects our everyday world, where the computer's progressive, comfortable embedding in everyday life brings with it new forms of withdrawal. What needs to be noticed about these effects is the fact that the shift to 'black-box' technology hides from us how things work, in ways we have grown happy to accept. A conventional metal car key, for example, functions as a kind of lever in a manner that is readily understood; the workings of the electronic fob, hidden in its ergonomically designed plastic casing, are such as most people do not understand, nor have any inclination to understand. The withdrawal is something in which we progressively collude.

It is worth tracing the background to this modern atomism also in philosophy itself. For it is precisely a kind of 'logical atomism' that Bertrand Russell advances. Russell extends a line of thought that derives in part from Leibniz and also from British empiricism. His thinking in this respect involves not only a certain

metaphysics, but an account of a philosophical methodology—that is, of a kind of analysis, perhaps the archetypal kind of analysis. The metaphysics consists in the belief that the world is fundamentality a plurality of independently existing things exhibiting qualities and standing in relations. These are the atomic facts of the world, and a proper philosophical approach involves uncovering such facts and understanding their relations: it requires reconstructing complex vocabularies in terms of simpler ones—ultimately a kind of reductionism. The mind of God, if such there could be, would then be in possession of the full inventory of these atomic facts as well as all their possible combinations.

Russell's logical atomism was obviously influential for the early Wittgenstein but also, more significantly, for logical positivism, which thrived in the decades that ensued, with its notorious relegation of ethics and aesthetics to the category of 'non-sense'. The limitations of this were subsequently exposed in philosophy in a variety of ways, and it would be good to think that matters ended there. Whether or not it is born again in the contemporary vogue for 'scientific naturalism', the more important aspect of its influence is, however, surely its inheritance in everyday life and in various forms of theorising. As Hilary Putnam has put this,

> Every one of you has heard someone ask, 'Is that supposed to be a fact or a value judgment?' The presupposition of this 'stumper' is that if it's a 'value judgment' it can't possibly be a [statement of] 'fact', and a further presupposition of this is that value judgments are subjective (Putnam, 2002).

He goes on to illustrate this with reference to the economist Lionel Robbins, who was a key government adviser on policy during the Great Depression. Robbins writes,

> If we disagree about ends it is a case of thy blood or mine—or live or let live according to the importance of the difference, or the relative strength of our opponents. But if we disagree about means, then scientific analysis can often help us resolve our differences. If we disagree about the morality of the taking of interest (and we understand what we are talking about), then there is no room for argument (Robbins, 1932, p. 132).

It is a sad truth that Putnam's 'stumper' is still heard often enough today, and that the assumptions that Robbins was making about facts and values in the 1930s have scarcely been rooted out from contemporary educational research.

What one sees here as a manifestation of this contemporary form of *ratio*—which computers, to the extent that they can be said to reason, epitomise and model—is the severance of the process of reasoning from the content of the reasoning, a possibility of which Leibniz had dreamed and with which he assiduously experimented (see Eco, 1995, Chapter 14).

14.3

Thoreau had some fun with practices of accounting when, in the chapter of *Walden* called 'Economy', he presented some pages of accounts. Thus, having remarked sardonically, in the preceding paragraphs, on the operations of state banks, he lists his outgoings on food for 8 months as follows:

Rice	$1.73 1/2	
Molasses	$1.73	Cheapest form of the saccharine.
Rye meal	$1.03 3/4	
Indian meal	$0.99 3/4	Ceaper than rye.
Pork	$0.2	
[etc.]		

(Thoreau, 1992b, p. 40).

There is no reason to doubt this as an accurate record of Thoreau's income and expenditure or that it provides a snapshot of sorts of the way he was spending his life at the time, and certainly the frugality of his expenditure, in contrast presumably to the operations of the banks, is part of his point. One can readily imagine that some readers will then take this at face value—as the somewhat quaint detailing of a life back in the woods, a retreat from the pressures of society. But there is a pointed lack of frugality in the writing of this chapter, some four or five times longer than any of the other 17 chapters that make up *Walden*, and this surely underscores the fact that the period of time that Thoreau spent living at Walden Pond—something under 2 years—and his writing of this account of his life are his 'experiment in living', which is to say his attempt to find a different economy for our lives. There is every reason then for this chapter to exceed the measure of all the others.

Thoreau's project is no less than to reappraise what Heidegger a century later will call our 'building dwelling thinking' (Heidegger, 1971), and this extends to a consideration of our relations to our neighbours, near and far: it questions how we account for ourselves. The recurrence throughout the text of a vocabulary of counting and accounting (of 'capital', 'interest', 'profit', 'loss', 'return'), supported by a consideration of what amounts to expenditure and investment and of what interest there might be in this, helps to show the way that this is simultaneously a meditation on language's relation to things. *Walden* asks, in effect, the Biblical question of what it shall profit a man to gain the world if he loses his soul, a question that Thoreau's contemporary readers in New England could hardly have missed. But it asks this without conventional religious presumptions but rather in relation to a right economy of living. And this, as we saw, is wrongly understood if it is imagined to imply merely an escape from society. Thoreau builds his hut about a mile from his neighbours and presents himself as a kind of example to them. But there comes a time when it is important to leave. He leaves the woods, as he explains at the end of the book, for the same reason he went there: there is a time to move on and to find something new. For, as Emerson has said, finding is founding. We can but sojourn in our lives; there is no ultimate or original dwelling place. And in contrast to this, Heidegger can appear positively nostalgic—nostalgia being, after all, a homesickness. Heidegger is drawn back to the origins, Greek thought, of the German language and, disastrously, of blood and belonging. Thoreau is without Heideggerian nostalgia. Hence, *Walden* can make vivid something like the qualities that a Greek *physis* and *logos* might suggest—though with an acknowledgement of

ways of thought that extend beyond the European inheritance and with a realisation, in the process, of a distinctively American possibility of being. Heidegger for sure would be horrified, and yet this anticipates, and in some ways exceeds, so much of what he said.

In the end Thoreau is simply practical. Amongst the townsfolk of Concord there are rumours of the extraordinary depth of Walden Pond, with fantastic stories about what can be seen in the winter if a hole is bored through the ice. Certainly Thoreau's accounts of the pond express a kind of enchantment. But in response to the rumours he rows a boat out and lowers a plumb line. The depth, he records, is 125 m.

14.4

Let us return to our opening office scene. If we imagine Fibbs today, it is not only the case that we can scarcely envisage the kind of confrontation with a shop steward that Pinter's sketch ironically adapts: it is also that we cannot think of him delighting in the factory's products. Even his workers are likely in many respects to be insulated from the things they are making by the black-box technology that governs so much of what they do. The poetry of Fibbs' work must now be located elsewhere.[5] While the technical names of machine parts once tripped off his tongue, he is now more likely to delight in his proficiency in marketing speak (given ethereal substance by Powerpoint) and in his familiarity and facility with the spreadsheets, bar charts, and flow diagrams that he will confidently gloss. Statistics will figure prominently in this.

The purpose of the present book is to consider the ethics and aesthetics of statistics. Statistics are present in our lives today in unprecedented ways, and they intrude insistently and in some ways insidiously into our modes of accounting for ourselves. In the process they hold a peculiar fascination, and it is helpful to understand this in aesthetic terms. Thus, there is a kind of poetic appeal for the sports enthusiast—similarly, no doubt, for the trader in the money markets (or even, one hesitates to make the comparison, for those who excitedly track the UK Research Assessment Exercise (RAE))—in the recording of numerical data of performance. This is plainly evident in the banners of statistics that accompany, for example a televised rugby match, which show the percentage of possession, the ratio of successful and unsuccessful conversion attempts, the number of tries, and so on. The examples proliferate, of course, through the various ways we govern ourselves. The key point here is the prominence of numbers in various forms.

Can anything more specific be said about the nature of this appeal? Here I would like to risk a number of suggestions.

In the first place, it is common to look to statistics to reveal 'underlying trends', as though there were a hidden reality to which the statistics might point. This is a complex kind of reification. There is a formal appeal, which is redolent of a certain strain of Platonism, coupled with the kind of atomism that was mentioned above.[6] What is at issue here is not the predictive power of statistics, whose value in medical

research none of us seriously doubts and whose place in the practice of education is not really to be questioned, albeit that specific findings merit careful critical scrutiny.[7] The point that is being made is rather that the prevalence of statistics in our lives alters our sense of the real, and this involves a kind of distortion of experience that is particularly harmful for educational practice. I take this harm to lie in a surreptitious nihilism, which downgrades the actuality of our experience in favour of an idealisation. In teaching and learning it distorts the sense of the point of what one is doing, for teacher and learner. Second, the provision of statistics is seen as tantamount to offering hard and often incontrovertible evidence, nowithstanding the fact that statistics inevitably effect the withdrawal in many respects of the reality to which they relate. The manner in which a technological understanding amplifies one aspect of the world at the same time as it hides another has itself become part of the way we have learned to see things, and it determines progressively what we expect of experience. This progressive expectation dovetails with structures of curriculum design that are themselves conditioned by other aspects of information technology, specifically its readiness for programming and means-ends thinking, and its facility with certain architectures of classification.

Third, there is the peculiar place of numbers in contemporary natural languages. Numbers, it seems, can appear in translations of texts in a way that is entirely without loss or distortion, and in this respect they are unlike the (more obviously) linguistic aspects of those texts. But in a sense this is to say that they cannot be translated at all—given that all translations between natural languages do involve loss and distortion in some degree. To some, this will confirm the view that numbers achieve an ideal clarity of meaning, and indeed it is partly such a thought that lay behind the experiments with language in which Leibniz and others engaged. It is partly what computers achieve (and partly what they depend on). But to others, this will demonstrate the ways in which numbers fall short of the very qualities of meaning by which our thought and being, and our accounting for ourselves are sustained: it is as if they offer a promise of totality, losing sight of the partiality of which we and our practices are inevitably constituted, and this in spite of the fact that, in the end, numbers and statistics cannot have meaning, cannot exist, outside the contexts of our natural languages.[8] This is then to say that there is an internal relationship between translation in its necessary imperfections and the very possibility of language, of meaning at all.

Fourth, while Wittgenstein wisely tells us to return our words to the circumstances in which they ordinarily occur in the language game, we need to acknowledge that many of our ordinary practices now involve forms of virtuality that sever them from the kinds of context that Wittgenstein had in mind. Virtuality is partly the product of the technological advances that have been under consideration, and so there are kinds of withdrawal that are inevitably part of what has become our ‘natural’ experience. Hence, it seems that the ways we must find of living well with these practices cannot simply follow the insights of Wittgenstein or Heidegger in these respects. Our nature is convention, and our convention is increasingly cybernature. Hence, the naturalisation of these practices is itself a form of writing the world.

Fifth, statistics and the kinds of accounting that go with it need to be understood not just as reflections or recordings of how the world is but as modes of writing the world. The poetry to which the new Fibbs responds comes from an excess in language that simultaneously projects a new possibility of being, for better or worse. To recognise that the world is written in this way is to begin to understand the libidinal excess with which our language is necessarily charged (never merely representing, always at the same time producing the world)—where 'language' is understood to extend through these technologies and the cyberpractices they engender. The savouring of the names in the Pinter sketch, their quickfire repetition and their sexual connotations accentuate this libidinal excess.

It may be thought that the line of argument presented here is tantamount to a call for 'social accounting', where it is accepted that it is not just the 'bottom line' of finance that can accurately reflect the way things are or how they should be.[9] Many of the practices that social accounting advocates have in recent years been adopted, but to some extent these are taken up in the way that hotels have environmental policies with regard to the washing of towels; and to some extent they formalise the kinds of things that decent managers have always considered anyway. In my view social accounting just scratches the surface. It is a different economy of living that is needed, and to understand this we need to take account of the kinds of changes that Heidegger and Thoreau—much better than, and before, Heidegger—draw to our attention.

Notes

1. Online, accessed on 25 February 2010, at http://www.touchdown-online.nl/.../Example%20discussion%20of%20a%20play.doc
2. See Stephanie 'Merritt, Harold Pinter, king of comedy', *The Observer*. Online at: http://www.guardian.co.uk/stage/2007/jan/14/theatre.comedy (Accessed on 25 February 2010).
3. For a fuller discussion of Heidegger and technology, see Blake, Smeyers, Smith, and Standish (2000, Chapter 1) and Standish (1997, 1999, 2000).
4. An ironic footnote to this history is the distraction that was caused in education by the advent, during the 1980s, of computing into the curriculum in schools and colleges. Computer study was then the flagship subject that schools earnestly introduced, and that increasing numbers of parents came to expect. Students took their first, very basic steps in computer programming and, in applications, became adept in the use of such commands as Control-KD—the way to open a new document in Wordstar, circa 1985. The comparative irrelevance of the former to the way that computing actually developed in society and the sheer obsolescence of the latter did not prevent them from distracting attention from where the most significant changes were happening—precisely in those modifications in ways of seeing and experiencing the world that are here at issue.
5. Exit Fibbs, enter David Brent. A very distant relation of Fibbs, David Brent is regional manager at the Slough offices of Wernham Hogg, manufacturers of office supplies. In contrast to the factory background implicit in the Pinter sketch, the television series, *The Office*, presents a listless office environment where, in place of background images of men in overalls at machine lathes and the noise of the production line, we see men and women gazing dully at their monitors and hear the soporific hum of the of the photocopier. Managers are focused not on the 'beauty'

of the products but on their own presentational skills and their expertise in human resource management. David Brent sees his own approach as inspirational. He thinks outside-the-box. He brings a new philosophy to work. But his crass ineptness, his cringe-making attempts at people management, and his barely concealed hang-ups and insecurities—absurd and funny as these are—in the end reinforce the sense of the inescapable emptiness of this office-world.

6. This is particularly pertinent to questions of labelling and identity (see Cigman & Davis, 2008). But it can also be seen in what is closer to home for many readers of this book in the gradings and evaluations that go with research assessment. Thus, to look back to an earlier iteration of the UK RAE, one looked around one's department and wondered whether while Jones was surely a 4 and Smith perhaps a 5, Evans was really no more than a 3b.
7. See Paul Smeyers (this volume) for a nuanced account of how the place of statistics in the practice of education might properly be understood.
8. For an elaboration of this point in relation to a Derridean account of meaning, see Ian Munday (this volume).
9. For a broad endorsement and survey of such conceptions of accounting, see Gray (2002). For a more critical discussion and an argument in favour of a more communitarian approach, see Lehman (1999). Both have been influential articles.

References

Blake, N., Smeyers, P., Smith, R., & Standish, P. (2000). *Education in an age of nihilism*. London: RoutledgeFalmer.

Cigman, R., & Davis, A. (Guest Eds). (2008). New philosophies of learning [Special Issue]. *Journal of Philosophy of Education, 42*(3/4).

Eco, U. (1995). *The search for the perfect language*. London: Fontana Press.

Gray, R. (2002). The social accounting project and accounting, organizations and society: Privileging engagement, imaginings, new accountings and pragmatism over critique? *Accounting, Organizations and Society, 27,* 687–708.

Heidegger, M. (1971). Building dwelling thinking. In *Poetry, language, thought* (A. Hofstadter, Trans.). New York: Harper Colophon Books.

Heidegger, M. (1991). *The principle of sufficient reason* (R. Lilly, Trans.). Bloomington: Indiana University Press. (Originally published in German in 1957)

Lehman, G. (1999). Disclosing new worlds: A role for social and environmental accounting and auditing. *Accounting, Organizations and Society, 24*, 217–241.

Pinter, H. (1996). Troubles in the works. In *Plays: 2*. London: Faber and Faber.

Putnam, H. (2002). *The collapse of the fact/value dichotomy and other essays.* Cambridge, MA: Harvard University Press.

Robbins, L. (1932). *An essay on the nature and significance of economic science*. London: Macmillan.

Standish, P. (1997). Heidegger and the technology of further education. *Journal of Philosophy of Education, 31,* 439–460.

Standish, P. (1999). Only Connect: computer literacy from Heidegger to cyberfeminism. *Educational Theory, 49*, 417–435.

Standish, P. (2000). Fetish for effect. In N. Blake & P. Standish (Eds.), *Education at the interface: Philosophical questions concerning on-line education* (Special Issue). *Journal of Philosophy of Education, 34*(4), 151–168.

Thoreau, H. D. (1992a). *Walden* and r*esistance to civil government* (Edited W. Rossi). New York: Norton and Company.

Thoreau, H. (1992b). *Walden.* Harmondsworth: Penguin.

About the Authors

David Bridges is professor emeritus at the University of East Anglia (where he was previously pro-vice-chancellor), professorial fellow at the University of Cambridge, faculty of education and emeritus fellow of St Edmund's College Cambridge. He is honorary vice-president of the Philosophy of Education Society of Great Britain and a council member of both the British and European Educational Research Associations – positions which reflect the way in which his academic career has bridged empirical research and philosophical work. Among the publications which have arisen out of this dual interest are *Fiction Written Under Oath? Essays in Philosophy and Educational Research* (Kluwer 2003); *The Ethics of Educational Research* (Blackwell 2003) edited with Mike McNamee; *Philosophy, Methodology and Educational Research* (Blackwell 2007) edited with Richard Smith; and *Evidence Based Educational Policy* (Wiley Blackwell 2009) edited with Paul Smeyers and Richard Smith.

Kathleen Coessens' research is situated at the border of science and art. It deals with human creativity and cultural representations looked at from an embodied, epistemological and philosophical perspective. She graduated in piano and chamber music from the Conservatory of Brussels and the Ėcole Cortot at Paris; at the Vrije Universiteit Brussel, she studied philosophy, sociology and psychology. She was awarded her doctorate in 2003. Her thesis was called *The human being as a cartographer – coping with the already epistemized world*. A reworked publication will follow in 2010 with the title: *The Human Being as a Cartographer*. She is professor and post-doctoral researcher at the Vrije Universiteit Brussel (VUB, Centre for Logic and Philosophy of Science), senior researcher at the Orpheus Research Centre in Music, Ghent, and guest professor at the Conservatory (Artesis Hogeschool), Antwerpen where she teaches 'semiotics', 'sociology of artistic practice' and 'arts and performance culture'. Her recent publications include *The Artistic Turn – A Manifesto* (2009, Leuven University Press) written with Darla Crispin and Anne Douglas.

Marc Depaepe, born in 1953, is full professor at the Katholieke Universiteit Leuven, where he teaches courses in the history of education and in the history of behavioural sciences at the campuses Kortrijk and Leuven. He has been

P. Smeyers, M. Depaepe (eds.), *Educational Research: The Ethics and Aesthetics of Statistics*, Educational Research 5, DOI 10.1007/978-90-481-9873-3,

co-editor-in-chief of *Paedagogica Historica*, since 2005 and a member of the editorial board of *Zeitschrift für Pädagogische Historiographie* (Switzerland), *Historia de la Educación. Revista Interuniversitaria* (Spain), *History of Education Review* (Australia) and *Ricerche Pedagogiche* (Italy). His research interests are theory, methodology and historiography of education; history of the educational sciences; educational history in Belgium; colonial history of education and the history of migration and education. He has published articles and books on these topics in several languages.

Jeroen J.H. Dekker is full professor of History and Theory of Education at the University of Groningen and president of the Groningen Research School for the Study of the Humanities. In 1998 and 2005, he was visiting professor at the History Department of the European University Institute in Florence. He is currently a visiting scholar at the European Institute based at Columbia University in New York. He is former president of the International Association for the History of Education (ISCHE), co-editor-in-chief of *Paedagogica Historica,* and visiting member of the Editorial Board of *History of Education*. He is the author of *The Will to Change the Child. Re-education Homes for Children at Risk in Nineteenth Century Western Europe* (2001), of *Het verlangen naar opvoeden* [The Educational Aspiration] (2006), and of numerous articles in scientific journals.

Karen François is a post-doctoral researcher at the Centre for Logic and Philosophy of Science –Vrije Universiteit Brussel, Belgium. Her doctoral research (2008) was on the politics of mathematics. She now focuses on statistical and mathematical literacy, ethno-mathematics, the intersection of science and politics, social constructivism and actor-network-theory. She has published several papers, the most recent of which are (2010) Ethnomathematics as Human Right, in *The Philosophy of Mathematics Education Journal, Special Issue On Critical Mathematics Education*, Paul Ernest (Ed.), No. 25; Ethnomathematics and the Philosophy of Mathematics, in *Education*, Benedikt Löwe and Thomas Müller (Eds.) written with Bart Van Kerkhove and *Philosophy of Mathematics: Sociological Aspects and Mathematical Practice. Series: Texts in Philosophy*. London: College Publications. She also co-edited a book with Jean Paul Bendegem called *Revolutions in Mathematics*. *More than 30 Years after Crowe's (1975) "Ten Laws". A New Interpretation*; further with Gert Goeminne, The Thing Called Environment: What It Is and How to Be Concerned With It, in the *Oxford Literary Review*, 31(1).

Norbert Grube received his Ph.D. in 1997 in history at the University of Hamburg. The subject of his thesis was the history of elementary schools in northern Germany 1730–1830. Since 2006 he has been research associate at the Pestalozzianum Research Institute for the History of Education. He is a member of the editorial board of the critical edition of *Letters to Pestalozzi*. His habilitation project focuses on the governmental application of survey research in West Germany during the Cold War.

Elias Hemelsoet is a doctoral researcher at the Department of Educational Theory, Ghent University, Belgium. His research focuses on the right to education in relation to equal educational opportunities for irregular migrant children.

Naomi Hodgson is a doctoral student at the Institute of Education, University of London. Her research, which is informed by the work of Michel Foucault, is focussed on the role of education in the construction of European citizenship in current policy and practice. Her research is funded by the UK Economic and Social Research Council (ESRC; Award number PTA-031-2005-00032).

Edwin Keiner is professor of Foundations of Education at the Friedrich-Alexander-University of Erlangen-Nürnberg, Germany, and director of the Institute of Educational Research. He is the chairperson of the section 'Foundations of Education' and of the commission Research on Education Research (Wissenschaftsforschung) within the German Association of Educational Research (DGfE). From 2003 to 2006 he was the spokesperson for the convenors of all networks of the European Education Research Association (EERA) at the council. His research areas are comparative history and the sociology of education and educational research in the context of processes of Europeanization; the relationship between educational research and educational practice; the use and application of educational research knowledge and the transformation of educational knowledge due to new media.

David F. Labaree is a professor at the Stanford University School of Education, Stanford, CA, USA. He was president of the US History of Education Society (2004–2005) and vice-president for Division F (history of education) of the American Educational Research Association (2003–2006). His research focuses on the history of American education. His new book is *Consuming School: How Personal Ambition Trumps Social Mission* (Harvard University Press, 2010).

Ian Munday taught English in a British secondary school for 5 years. He is currently teaching associate in the Philosophy of Education at the University of Cambridge and a Ph.D. student at the Institute of Education in London. His research interests are in performativity and performatives in the work of J.L. Austin, Judith Butler, Stanley Cavell, Gilles Deleuze and Jacques Derrida.

Paul Smeyers is research professor of Philosophy of Education at Ghent University and part-time professor at K. U. Leuven. He teaches philosophy of education and methodology of the *Geisteswissenschaften* (qualitative/interpretative research methods). He has a wide involvement in philosophy of education (more than 250 publications). He holds, or has held several positions in the International Network of Philosophers of Education (president since 2006). For almost a decade, he has chaired the Research Community *Philosophy and History of the Discipline of Education: Evaluation and Evolution of the Criteria for Educational Research* established by the Research Foundation Flanders, Belgium (Fonds

voor Wetenschappelijk Onderzoek – Vlaanderen). Together with Nigel Blake, Richard Smith and Paul Standish he co-authored *Thinking Again. Education after Postmodernism* (Bergin & Garvey, 1998), *Education in an Age of Nihilism* (Falmer Press, 2000) and *The Therapy of Education* (Palgrave Macmillan, 2007) and co-edited *The Blackwell Guide to the Philosophy of Education* (2003). Together with Marc Depaepe he co-edited several books published by the Research Community (Dordrecht: Springer). With Michael Peters and Nick Burbules he co-authored *Showing and doing. Wittgenstein as a Pedagogical Philosopher* (Paradigm Publishers, 2008).

Richard Smith is professor of Education at the University of Durham, United Kingdom. He is founding editor of *Ethics and Education* and associate editor of the *Journal of Philosophy of Education*. His most recent book is (with David Bridges and Paul Smeyers, joint eds.) *Evidence-based Educational Policy: What Evidence? What Basis? Whose Policy?* (Blackwell, 2009).

Ulrike Stadler-Altmann is associate professor at the Institute of Education Research, Friedrich-Alexander-University Erlangen-Nürnberg, Germany. Her research interests address empirical research on school and instructional designs, the professional work of teachers and students' self-concepts. She is also interested in gender issues.

Paul Standish is professor and head of Philosophy of Education at the Institute of Education – London. His work expands across a range of topics in education, and he is particularly interested in the creative tensions between analytical and continental approaches to philosophy. He is editor of the *Journal of Philosophy of Education*.

Jean Paul Van Bendegem is full professor at the Vrije Universiteit Brussel where he teaches courses in logic and philosophy of science. He is director of the Center for Logic and Philosophy of Science (www.vub.ac.be/CLWF/) where ten researchers are currently working. He has been president of the National Center for Research in Logic (http://www.lofs.ucl.ac.be:16080/cnrl/), founded in 1955 by, amongst others, Chaïm Perelman and Leo Apostel. He is the editor of *Logique et Analyse* (http://www.vub.ac.be/CLWF/L&A/). His research focuses on the philosophy of strict finitism and on the development of a comprehensive theory of mathematical practice. Recent publications include: *Perspectives on Mathematical Practices. Bringing together Philosophy of Mathematics*, *Sociology of Mathematics, and Mathematics Education* (jointly edited with Bart Van Kerkhove, Dordrecht: Springer/Kluwer Academic, 2006), *Philosophical Dimensions in Mathematics Education* (jointly edited with Karen François, New York: Springer, 2007, Mathematics Education Library 42) and *Mathematical Arguments in Context* co-authored with Bart Van Kerkhove (*Foundations of Science*, 2009). Personal website: http://www.vub.ac.be/CLWF/members/jean/index.shtml

Author Index

Subject Index