

Scientific Objectivity

Julian Reiss* Jan Sprenger[†]

January 10, 2014

Draft version---please do not cite without permission!

*Department of Philosophy, Durham University, 50 Old Elvet, Durham DH1 3HN, United Kingdom. Email: julian.reiss@durham.ac.uk. Webpage: www.jreiss.org.

[†]Contact information: Tilburg Center for Logic, General Ethics and Philosophy of Science (TiLPS), Tilburg University, P.O. Box 90153, 5000 LE Tilburg, The Netherlands. Email: j.sprenger@uvt.nl. Webpage: <http://www.laeuferpaar.de>.

Scientific objectivity is a characteristic of scientific claims, methods and results. It expresses the idea that the claims, methods and results of science are not, or should not be influenced by particular perspectives, value commitments, community bias or personal interests, to name a few relevant factors. Objectivity is often considered as an ideal for scientific inquiry, as a good reason for valuing scientific knowledge, and as the basis of the authority of science in society.

Many central debates in the philosophy of science have, in one way or another, to do with objectivity: confirmation and the problem of induction; theory choice and scientific change; realism; scientific explanation; experimentation; measurement and quantification; evidence and the foundations of statistics; evidence-based science; feminism and values in science. Understanding the role objectivity plays in science is therefore integral to a full appreciation of these debates. As this article testifies, the reverse is true too: it is impossible to fully appreciate the notion of scientific objectivity without touching upon many of these debates.

The ideal of objectivity has been criticized repeatedly in philosophy of science, questioning both its value and its attainability. This article focuses on the question of whether the concept of objectivity plays a guiding role for modern science, whether the ideal of objectivity is desirable at all, and to what extent scientists can achieve it. In line with the idea that the epistemic authority of science relies primarily on the objectivity of scientific reasoning, we focus on the role of objectivity in scientific experimentation, inference and theory choice.

Contents

1	Introduction: Product and Process Objectivity	4
2	Objectivity as Faithfulness to Facts	6
2.1	The View From Nowhere	6
2.2	Theory-Ladenness and Incommensurability	8
2.3	The Experimenter’s Regress	11
2.4	Standpoint Theory, Contextual Empiricism and Trust in Science	12
3	Objectivity as Absence of Normative Commitments and the Value-Free Ideal	15
3.1	Epistemic and Contextual Values	15
3.2	Acceptance of Scientific Hypotheses and Value Neutrality	18
3.3	Science, Policy and the Value-Free Ideal	24
3.4	Feyerabend: The Tyranny of the Rational Method	28
4	Objectivity as Freedom from Personal Biases	31
4.1	Measurement and Quantification	31
4.2	Inductive and Statistical Inference	35
5	Issues in the Special Sciences	44
5.1	Max Weber and Objectivity in the Social Sciences	44
5.2	Contemporary Rational Choice Theory	46
5.3	Evidence-based Medicine and Social Policy	47
6	Instrumentalism to the Rescue?	50
7	Conclusions	52

1 Introduction: Product and Process Objectivity

Objectivity is a value. To call a thing objective implies that it has a certain degree of importance to us and that we approve of it. Objectivity comes in degrees. Claims, methods and results can be more or less objective, and, other things being equal, the more objective, the better. Using the term ‘objective’ to describe something often carries a special rhetorical force with it. The admiration of science among the general public and the authority science enjoys in public life stems to a large extent from the view that science is objective or at least more objective than other modes of inquiry. Understanding scientific objectivity is therefore central to understanding the nature of science and the role it plays in society.

Given the centrality of the concept for science and everyday life, it is not surprising that attempts to find ready characterizations are bound to fail. For one thing, there are two fundamentally different ways to understand the term: **product objectivity** and **process objectivity**. According to the first understanding, science is objective in that, or to the extent that, its products—theories, laws, experimental results and observations, ‘facts’—constitute accurate representations of the external world. The products of science according to this view are not tainted by human desires, goals, capabilities or experience. According to the second understanding, science is objective in that, or to the extent that, the processes and methods that characterize it neither depend on contingent social and ethical values, nor on the individual bias of a scientist. Especially this second understanding is itself multi-faceted; it contains, inter alia, explications in terms of measurement procedures, individual reasoning processes, or the social and institutional dimension of science. The semantic richness of scientific objectivity is also reflected in the multitude of categorizations and subdivisions of the concept (e.g., Megill 1994; Douglas 2004).

If it is true that what is so great about science is its objectivity, then objectivity should be worth defending. The close examinations of scientific practice that philosophers of science have undertaken in the past fifty years have shown, however, that several conceptions of the ideal of objectivity are either questionable or unattainable. The prospects for a science providing a non-perspectival ‘view from nowhere’ or for proceeding in a way uninformed by human goals and values are fairly slim, for example.

This article discusses several proposals to characterize the idea and ideal of objectivity in such a way that it is both strong enough to be valuable, and weak enough to be attainable and workable in practice. We begin with

a natural conception of objectivity: **faithfulness to facts**, which is closely related to the idea of product objectivity. We motivate the intuitive appeal of this conception, discuss its relation to scientific method and discuss arguments challenging both its attainability as well as its desirability. We then move on to a second conception of objectivity as **absence of normative commitments and value-freedom**, and once more we contrast arguments in favor of such a conception with the challenges it faces. The third conception of objectivity which we discuss at length is the idea of **absence of personal bias**. After discussing three case studies about objectivity in scientific practice (from economics, social science and medicine) as well as a radical alternative to the traditional conceptions of objectivity, instrumentalism, we draw some conclusions about what aspects of objectivity remain defensible and desirable in the light of the difficulties we have discussed.

2 Objectivity as Faithfulness to Facts

The idea of this first conception of objectivity is that scientific claims are objective in so far as they faithfully describe facts about the world. The philosophical rationale underlying this conception of objectivity is the view that there are facts ‘out there’ in the world and that it is the task of a scientist to discover, analyze and to systematize them. ‘Objective’ then becomes a success word: if a claim is objective, it successfully captures some feature of the world.

In this view, science is objective to the degree that it succeeds at discovering and generalizing facts, abstracting from the perspective of the individual scientist. Although few philosophers have fully endorsed such a conception of scientific objectivity, the idea figures recurrently in the work of prominent 20th century philosophers of science such as Carnap, Hempel, Popper, and Reichenbach. It is also, in an evident way, related to the claims of scientific realism, according to which it is the task of science to find out the truths about the world, and according to which we have reason to believe in the truth of our best-confirmed scientific theories.

2.1 The View From Nowhere

Humans experience the world from a perspective. The contents of an individual’s experiences vary greatly with the individual’s perspective, which is affected by his or her personal situation, details of his or her perceptual apparatus, language and culture, the physical conditions in which the perspective is made. While the experiences vary, there seems to be something that remains constant. The appearance of a tree will change as one approaches it but, at least possibly, the tree itself doesn’t. A room may feel hot or cold depending on the climate one is used to but it will, at least possibly, have a degree of warmth that is independent of one’s experiences. The object in front of a person does not, at least not necessarily, disappear just because the lights are turned off.

There is a conception of objectivity that presupposes that there are two kinds of qualities: ones that vary with the perspective one has or takes, and ones that remain constant through changes of perspective. The latter are the objective properties. Thomas Nagel explains that we arrive at the idea of objective properties in three steps (Nagel 1986: 14). The first step is to realize (or postulate) that our perceptions are caused by the actions of things on us, through their effects on our bodies. The second step is to realize (or

postulate) that since the same properties that cause perceptions in us also have effects on other things and can exist without causing any perceptions at all, their true nature must be detachable from their perspectival appearance and need not resemble it. The final step is to form a conception of that ‘true nature’ independently of any perspective. Nagel calls that conception the ‘view from nowhere’, Bernard Williams the ‘absolute conception’ (Williams 2011[1985]). It represents the world as it is, unmediated by human minds and other ‘distortions’.

Many scientific realists maintain that science, or at least natural science, does and indeed ought to aim to describe the world in terms of this absolute conception and that it is to some extent successful in doing so (for a detailed discussion of scientific realism, see <http://plato.stanford.edu/entries/scientific-realism/>). There is an immediate sense in which the absolute conception is an attractive one to have. If two people looking at a colored patch in front of them disagree whether it is green or brown, the absolute conception provides an answer to the question (e.g., ‘The patch emits light at a wavelength of 510 nanometers’). By making these facts accessible through, say, a spectroscope, we can arbitrate between the conflicting viewpoints (viz., by stating that the patch should look green to a normal observer in daylight).

Another reason for this conception to be attractive is that it will provide for a simpler and more unified representation of the world. Theories of trees will be very hard to come by if they use predicates such as ‘height as seen by an observer’ and a hodgepodge if their predicates track the habits of ordinary language users rather than the properties of the world. To the extent, then, that science aims to provide explanations for natural phenomena, casting them in terms of the absolute conception would help to realize this aim. Bernard Williams makes a related point about explanation: ‘The substance of the absolute conception (as opposed to those vacuous or vanishing ideas of ‘the world’ that were offered before) lies in the idea that it could nonvacuously explain how it itself, and the various perspectival views of the world, are possible’ (Williams 2011[1985]: 139). Thus, a scientific account cast in the language of the absolute conception may not only be able to explain why a tree is as tall as it is but also why we see it in one way when viewed from one standpoint and in a different way when viewed from another.

A third possible reason to find the view from nowhere attractive is that if the world came in structures as characterized by it and we did have access

to it, we can use our knowledge of it to ground predictions (which, to the extent that our theories do track the absolute structures, will be borne out). A fourth and related reason is that attempts to manipulate and control phenomena can similarly be grounded in our knowledge of these structures. To attain any of the four purposes—settling disagreements, explaining the world, predicting phenomena and manipulation and control—the absolute conception is at best sufficient but not necessary. We can, for instance, settle disagreements by imposing the rule that the person who speaks first is always right or the person who is of higher social rank or by an agreed-upon measurement procedure that does not track absolute height. We can explain the world and our image of it by means of theories that do not represent absolute structures and properties, and there is no need to get things (absolutely) right in order to predict successfully. Nevertheless, there is something appealing in the idea that disagreements concerning certain matters of fact can be settled by the very facts themselves, that explanations and predictions grounded in what’s really there rather than in a distorted image of it.

No matter how desirable, it is clear that our ability to use scientific claims to represent all and only facts about the world depends on whether these claims can unambiguously be established on the basis of evidence. We test scientific claims by means of their implications, and it is an elementary principle of logic that claims whose implications are true need not themselves be true. It is the job of scientific method to make sure that observations, measurements, experiments, tests—the scientific *evidence*—speak in favor of the scientific claim at hand. Alas, the relation between evidence and scientific hypothesis is not straightforward. In the next two subsections we will look at two challenges of the idea that even the best scientific method will yield claims that describe an aperspectival view from nowhere. Subsection 2.4 will challenge the idea that the view from nowhere is a good thing to have.

2.2 Theory-Ladeness and Incommensurability

According to a popular picture, science progresses toward truth by adding true and eliminating false beliefs from our best scientific theories. By making these theories more and more verisimilar, that is, truthlike, scientific knowledge grows over time (e.g., Popper 1963). If this picture is correct, then over time, scientific knowledge will become more objective, that is, more faithful to facts. However, scientific theories often change, and sometimes several

theories compete for the place of the best scientific account of the world.

It is inherent in the above picture of scientific objectivity that observations can, at least in principle, decide between competing theories: if they did not, the conception of objectivity as faithfulness would be a pointless one to have, as we would not be in a position to verify it. This position has been adopted by Karl R. Popper, Rudolf Carnap and other leading figures in (broadly) empiricist philosophy of science. Many philosophers have argued that the relation between observation and theory is way more complex and that influences can actually run both ways (e.g. Duhem 1906, Wittgenstein 1953/2001, Hanson 1958). The most lasting criticism, however, was delivered by Thomas S. Kuhn (1962/1970) in his book ‘The Structure of Scientific Revolutions’.

Kuhn’s analysis is built on the assumption that scientists always view research problems through the lens of a paradigm, defined by set of relevant problems, axioms, methodological presuppositions, techniques, and so forth. Kuhn provided several historical examples in favor of this claim. Scientific progress—and the practice of normal, everyday science—happens within a paradigm that guides the individual scientists’ puzzle-solving work and that sets the community standards.

Can observations undermine such a paradigm, and speak for a different one? Here, Kuhn famously stresses that **observations are ‘theory-laden’** (cf. also Hanson 1958): they depend on a body of theoretical assumptions through which they are perceived and conceptualized. This hypothesis has two important aspects.

First, the *meaning* of observational concepts is influenced by theoretical assumptions and presuppositions. For example, the concepts ‘mass’ and ‘length’ have different meanings in Newtonian and relativistic mechanics; so does the concept ‘temperature’ in thermodynamics and statistical mechanics (cf. Feyerabend 1962). In other words, Kuhn denies that there is a theory-independent observation language. The ‘faithfulness to reality’ of an observation report is always mediated by a theoretical *Überbau*, impairing the role of observation reports as an impartial, merely fact-dependent arbiter between different theories.

Second, not only the observational concepts, but also the *perception* of a scientist depends on the paradigm she is working in. ‘Practicing in different worlds, the two groups of scientists [who work in different paradigms, J.R./J.S.] see different things when they look from the same point in the same direction’ (Kuhn 1962/1970: 150). That is, our own sense data are

shaped and structured by a theoretical framework, and may be fundamentally distinct from the sense data of scientists working in another one. If this picture is correct, it is hard to assess which theory or paradigm is more faithful to the facts, that is, more objective.

The thesis of the theory-ladenness of observation has also been extended to the **incommensurability of different paradigms or scientific theories**, problematized independently by Thomas S. Kuhn (1962/1970) and Paul Feyerabend (1962). Literally, this concept means “having no measure in common”, and it figures prominently in arguments against a linear and standpoint-independent picture of scientific progress. For instance, the Special Theory of Relativity appears to be more objective than Newtonian mechanics because it reduces, for low speeds, to the latter, and it accounts for some additional facts as well. This picture is undermined, however, by two central aspects of incommensurability. First, not only do the observational concepts in both theories differ, but the principles for specifying their meaning may be inconsistent with each other (Feyerabend 1975, 269–270). Second, scientific research methods and standards of evaluation change with the theories or paradigms. Not all puzzles that could be tackled in the old paradigm will be solved by the new one—this is the phenomenon of ‘Kuhn loss’.

A meaningful use of objectivity presupposes, according to Feyerabend, to perceive and to describe the world from a specific perspective, e.g., when we try to verify the referential claims of a scientific theory. Only *within* a peculiar scientific worldview, the concept of objectivity may be applied meaningfully. That is, scientific method cannot free itself from the particular scientific theory to which it is applied; the door to standpoint-independence is locked. As Feyerabend puts it: ‘our epistemic activities may have a decisive influence even upon the most solid piece of cosmological furniture—they make gods disappear and replace them by heaps of atom in empty space’ (1978: 70).

Kuhn and Feyerabend’s theses about theory-ladenness of observation, and their implications for the objectivity of scientific inquiry have been much debated afterwards, and have often been misunderstood in a social constructivist sense. Therefore Kuhn later returned to the topic of scientific objectivity, of which he gives his own characterization in terms of the shared cognitive values of a scientific community. We discuss Kuhn’s later view in section 3. For a more profound coverage, see Section 3 in the SEP entry on the incommensurability of scien-

tific theories (<http://plato.stanford.edu/entries/incommensurability/>) and Section 4.2. in the SEP entry on Thomas S. Kuhn (<http://plato.stanford.edu/entries/thomas-kuhn/>).

2.3 The Experimenter's Regress

Most of the earlier critics of 'objective' verification or falsification focused on the relation between evidence and scientific theories. There is a sense in which the claim that this relation is problematic is not so surprising. Scientific theories contain highly abstract claims that describe states of affairs far removed from the immediacy of sense experience. This is for a good reason: sense experience is necessarily perspectival, so to the extent to which scientific theories are to track the absolute conception, they must describe a world different from that of sense experience. But surely, one might think, the evidence itself is objective. So even if we do have reasons to doubt that abstract theories faithfully represent the world, we should stand on firmer grounds when it comes to the evidence against which we test abstract theories.

Theories are seldom tested against brute observations, however. This too is for good reason: if they were, they'd be unlikely to track the absolute conception. Simple generalizations such as 'all swans are white' are directly learned from observations (say, of the color of swans) but they do not represent the view from nowhere (for one thing, the view from nowhere doesn't have colors). Genuine scientific theories are tested against experimental facts or phenomena, which are themselves unobservable to the unaided senses. Experimental facts or phenomena are instead established using intricate procedures of scientific measurement and experimentation.

We therefore need to ask whether the results of scientific measurements and experiments can be aperspectival. In an important debate in the 1980s and 1990s some commentators answered that question with a resounding 'no', which was then rebutted by others. The debate concerns the so-called 'experimenter's regress' (Collins 1985). Collins, a prominent sociologist of science, claims that in order to know whether an experimental result is correct, one first needs to know whether the apparatus producing the result is reliable. But one doesn't know whether the apparatus is reliable unless one knows that it produces correct results in the first place and so on and so on *ad infinitum*. Collins' main case concerns attempts to detect gravitational waves, which were very controversially discussed among physicists in the 1970s.

Collins argues that the circle is eventually broken not by the ‘facts’ themselves but rather by factors having to do with the scientist’s career, the social and cognitive interests of his community, and the expected fruitfulness for future work. It is important to note that in Collins’s view these factors do not necessarily make scientific results arbitrary. But what he does argue is that the experimental results do not represent the world according to the absolute conception. Rather, they are produced jointly by the world, scientific apparatuses, and the psychological and sociological factors mentioned above. The facts and phenomena of science are therefore necessarily perspectival.

In a series of contributions, Allan Franklin, a physicist-turned-philosopher of science, has tried to show that while there are indeed no algorithmic procedures for establishing experimental facts, disagreements can nevertheless be settled by *reasoned judgement* on the basis of bona fide epistemological criteria such as experimental checks and calibration, elimination of possible sources of error, using apparatuses based on well-corroborated theory and so on (Franklin 1994; 1997). Collins responds that ‘reasonableness’ is a social category that is not drawn from physics (Collins 1994).

The main issue for us in this debate is whether there are any reasons to believe that experimental results provide an aperspectival view on the world. According to Collins, experimental results are co-determined by the facts as well as social and psychological factors. According to Franklin, whatever else influences experimental results other than facts is not arbitrary but instead based on reasoned judgment. What he has not shown is that reasoned judgment guarantees that experimental results reflect the facts alone and are therefore aperspectival in any interesting sense.

2.4 Standpoint Theory, Contextual Empiricism and Trust in Science

Feminist standpoint theorists and proponents of ‘situated knowledge’ such as Donna Haraway (1988), Sandra Harding (1993) and Alison Wylie (2003) deny the internal coherence of a view from nowhere: all human knowledge is at base *human* knowledge and is therefore necessarily perspectival. But they argue more than that. Not only is perspectival the human condition, it is also a good thing to have. This is because perspectives, especially the perspectives of underprivileged classes, come along with certain epistemic advantages.

Standpoint theory is a development of certain Marxist ideas that epistemic position is related to social position. According to this view, workers, being members of an underprivileged class, have both greater incentives to understand social relations better, and better access because they live under the capitalists' rule and therefore have access to the lives of capitalists as well as their own lives. Feminist standpoint theory builds on these ideas but focuses on gender, racial and other social relations.

These ideas are controversial but they draw attention to the possibility that attempts to rid science of perspectives might not only be futile because scientific knowledge is necessarily perspectival, they can also be epistemically costly because they prevent scientists from having the epistemic benefits certain standpoints afford.

If there are no methods that guarantee objective outcomes or objective criteria against which to assess outcomes, what might 'procedural objectivity' consist in? A particularly answer that goes back to Karl R. Popper (1972, 1934/2002) has been taken up and modified by Helen Longino. Popper claimed that 'the objectivity of scientific statements lies in the fact that they can be inter-subjectively tested' (1934/2002: 22), where 'inter-subjectively testable' may be understood as there being verifiable facts with evidential bearing on the theory in question. So Popper does not see the objectivity of a scientific claim in a direct correspondence to facts: rather, the claim must be testable and subject to rational criticism.

Longino (1990) reinforces Popper's focus on intersubjective criticism: for her, scientific knowledge is essentially a social product. Thus, our conception of scientific objectivity must directly engage with the social process that generates knowledge. In response to the failures of attempts to define objectivity as faithfulness of theory to facts, she concludes that social criticism fulfills crucial functions in securing the epistemic success of science. The objectivity of science is no more grounded in correspondence between theory and facts, or in all scientists seeing the same result (called 'concordant objectivity' by Douglas 2011), but in the 'interactive objectivity' that emerges through the scientists' open discourse. Specifically, she develops an epistemology called **contextual empiricism** which regards a method of inquiry as 'objective to the degree that it permits **transformative criticism**' (Longino 1990: 76). For an epistemic community to achieve transformative criticism, there must be:

- 1) **avenues for criticism**: criticism is an essential part of scientific institutions (e.g., peer review);

- 2) **shared standards:** the community must share a set of cognitive values for assessing theories (more on this in Section 3);
- 3) **uptake of criticism:** criticism must be able to transform scientific practice in the long run;
- 4) **equality of intellectual authority:** intellectual authority must be shared equally among qualified practitioners.

Longino’s contextual empiricism can be understood as a development of John Stuart Mill’s view that beliefs should never be suppressed, independently of whether they are true or false (Mill 1859). Even the most implausible beliefs might, for all we know, be true as we are not infallible; and if they are false, they might contain a grain of truth which is worth preserving or, if wholly false, help to better articulate and defend those beliefs which are true (Mill 1859: 72).

Social epistemologists such as Longino similarly see objectivity neither in the products of science (as there is no view from nowhere) nor in its methods (as there aren’t any standards that are valid independently of the contexts of specific inquiries) but rather in the idea that many and competing voices are heard. The underlying intuition is supported by recent empirical research on the epistemic benefits of a diversity of opinions and perspectives (Page 2007).

The turn from scientific results and methods to the social organization of science involves numerous problems. On the one hand, we might ask how many and which voices must be heard for science to be objective. It is not clear for instance whether non-scientists should have as much authority as trained scientists. The condition of equality of intellectual requires only ‘qualified’ practitioners to equally share authority—but who qualifies as ‘qualified’? Nor is it clear whether it is always a good idea to subject every scientific result to democratic approval, as Paul Feyerabend proposed (Feyerabend 1975, 1978). On the other hand, there is certainly no guarantee that democratized science leads to true theories, or even reliable ones. So why should we value objectivity in the sense of social epistemologists?

One answer to this question has been given by Arthur Fine who argues that we value objectivity in this sense because it promotes trust in science (Fine 1998: 17). While there is no guarantee that the process leads to true theories, it is nevertheless trusted because it is fair. We will consider Fine’s views on objectivity in more detail below in Section 6.

3 Objectivity as Absence of Normative Commitments and the Value-Free Ideal

The previous section has presented us with forceful arguments against the view of objectivity as faithfulness to facts and an impersonal ‘view from nowhere’. How can we maintain the view that objectivity is one of the essential features of science—and the one that grounds its epistemic authority? A popular reply contends that science should be **value-free** and that scientific claims or practices are objective to the extent that they are free of moral, political and social values.

3.1 Epistemic and Contextual Values

Before addressing what we will call the ‘value-free ideal’, it will be helpful to distinguish four stages at which values may affect science. They are: (i) the choice of a scientific research problem; (ii) the gathering of evidence in relation to the problem; (iii) the acceptance of a scientific hypothesis or theory as an adequate answer to the problem on the basis of the evidence; (iv) the proliferation and application of scientific research results (cf. Weber 1917/1988).

Most philosophers of science would agree that the role of values in science is contentious only with respect to dimensions (ii) and (iii): the **gathering of evidence** and the **acceptance of scientific theories**. It is almost universally accepted that the choice of a research problem is often influenced by the interests of individual scientists, funding parties, and society as a whole. This influence may make science more shallow and slow down its long-run progress, but it has benefits, too: scientists will focus on providing solutions to those intellectual problems that are considered urgent by society and they may actually improve people’s lives. Similarly, the proliferation and application of scientific research results is evidently affected by the personal values of journal editors and end users, and there seems to be little one can do about this. The real debate is about whether or not the ‘core’ of scientific reasoning—the gathering of evidence and the assessment and acceptance scientific theories—is, and should be, value-free.

An obvious, but unconvincing objection invokes the ‘underdetermination of theory by evidence’ (see <http://plato.stanford.edu/entries/scientific-underdetermination/>). As we have seen above, the relationship between theory and evidence is rather complex. More often than not in the history of science, the existing body of evidence in some domain does not pick out

a unique theoretical account of that domain. ‘Crucial experiments’ do not refute a specific scientific claim, but only indicate that there is an error in an entire network of hypotheses (Duhem 1906/1954). Thus, existing bodies of evidence often *underdetermine* the choice of rival theoretical accounts.

The gap between evidence and theory is filled in by **scientific values**. Consider the curve-fitting problem. When fitting a curve to a data set, the researcher often has the choice between either using a higher-order polynomial, which makes the curve less *simple* but fit the data more *accurately*, or using a lower-order polynomial, which makes the curve *simpler* albeit less *accurate*. Simplicity and accuracy are both scientific values: for instance, econometricians have a preference for solving curve-fitting problems by means of linear regression, thereby valuing simplicity over accuracy.

Philosophers of science tend to regard value-ladenness in this sense as benign. **Epistemic (or cognitive) values** such as predictive accuracy, scope, unification, explanatory power, simplicity and coherence with other accepted theories are taken to be indicative of a good scientific theory and figure in standard arguments for preferring one theory over another. Kuhn (1977) even claims that epistemic values define the shared commitments of science, that is, the standards which scientists have used at all times for assessing the merits of a theory.

A word on terminology. Sometimes epistemic values are regarded as a subset of cognitive values and identified with values such as empirical adequacy and internal consistency that directly bear on the veracity of a scientific theory (Laudan 2004). Values such as scope and explanatory power would then count as cognitive values that express scientific desiderata, but without properly epistemic implications. We have decided, however, to adopt a broader reading of ‘epistemic’ where truth is not the only aim of scientific inquiry, but supplemented by providing causal mechanisms, finding natural laws, creating understanding, etc. In this sense, values such as scope or explanatory power contribute to achieving our epistemic goals. In addition, neat distinctions between truth-conducive and purely cognitive scientific values are hard to come by (see Douglas 2013 for a classification attempt).

Not every philosopher entertains the same list of epistemic values. In Lycan’s (1985) pragmatic perspective, simplicity is included because it reduces the cognitive workload of the scientific practitioner, and because it facilitates the use of scientific theories in dealing with real-world problems. McMullin (2009), on the other hand, does not include simplicity because the notion

is ambiguous, and because there are no conclusive arguments that simpler theories are more likely to be true, or empirically adequate. Subjective differences in ranking and applying epistemic values do not vanish, a point Kuhn made emphatically. This is also one of the reasons for using the term ‘value’ rather than ‘rule’: the assessment of a scientific theory rather corresponds to a *judgment* where different criteria are carefully weighed than to the mechanic application of a rule or algorithm to determine the best theory (McMullin 1982: 17).

In most views, the objectivity and authority of science is not threatened by epistemic, but by **non-cognitive, contextual values**. Contextual values are moral, personal, social, political and cultural values such as pleasure, justice and equality, conservation of the natural environment and diversity. The most notorious cases of improper uses of such values involve travesties of scientific reasoning, where the intrusion of contextual values led to an intolerant and oppressive scientific agenda with devastating epistemic and social consequences. In the Third Reich, a large part of contemporary physics, such as the theory of relativity, was condemned because its inventors were Jewish; in the Soviet Union, biologist Nikolai Vavilov was sentenced to death (and died in prison) because his theories of genetic inheritance did not match Marxist-Leninist ideology. Both states tried to foster a science that was motivated by political convictions (Lenard’s ‘Deutsche Physik’ in Nazi Germany, Lysenko’s anti-genetic theory of inheritance in the Soviet Union), leading to disastrous epistemic and institutional effects.

Less spectacular but numerically more significant cases analyzed by feminist philosophers of science involve gender or racial bias in biological theories (e.g., Okruhlik 1994, Lloyd 2005). Moreover, a lot of industry-sponsored research in medicine (and elsewhere) is demonstrably biased toward the interests of the sponsors, usually large pharmaceutical firms (e.g., Resnik 2007; Reiss 2010). This **preference bias**, defined by Wilholt (2009) as the infringement of conventional standards of the research community, with the aim of arriving at a particular result, is clearly epistemically harmful. Especially for sensitive high-stakes issues such as the admission of medical drugs or the consequences of anthropogenic global warming, it seems desirable that research scientists assess theories without being influenced by such considerations. This is the core idea of the

Value-Free Ideal (VFI): Scientists should strive to minimize the influence of contextual values on scientific reasoning, e.g., in gathering evidence and assessing/accepting scientific theories.

According to the VFI, scientific objectivity is characterized by absence of contextual values and by exclusive commitment to epistemic values in scientific reasoning. See Dorato (2004: 53–54), Ruphy (2006: 190) or Biddle (2013: 125) for alternative formulations.

The next question is then whether the VFI is actually attainable. This is the subject of the

Value-Neutrality Thesis (VNT): Scientists can—at least in principle—gather evidence and assess/accept theories without making contextual value judgments.

While this latter thesis is defended less frequently than the VFI, it serves as a useful foil for discussing its attainability. Note that the VNT is not normative: it only investigates whether the judgments that scientists make are, or could be, free of contextual values.

The VNT is denied by the value-laden thesis, which asserts that contextual values are essential for scientific research.

Value-Laden Thesis (VLT): Scientists cannot gather evidence and assess/accept theories without making contextual value judgments.

The latter thesis is sometimes strengthened to the claim that both epistemic and contextual values are essential to scientific research—and pursuit of a science without contextual values would be harmful both epistemically and socially (see Section 3.4). Either way, the acceptance of the value-laden thesis poses a challenge for re-defining scientific objectivity: one can either conclude that the ideal of objectivity is harmful and should be rejected (as Feyerabend does), or one can come up with a different and refined conception of objectivity (as Douglas and Longino do).

This section discusses the VNT as applied to the assessment and acceptance of scientific hypothesis, the role of the VFI at the interface between scientific reasoning and policy advice, and Paul Feyerabend’s radical attack on the VNT.

3.2 Acceptance of Scientific Hypotheses and Value Neutrality

Regarding the assessment of scientific theories, the VNT is a relatively recent position in philosophy of science. Its rise is closely connected to Reichenbach’s famous distinction between **context of discovery** and **context of**

justification. Reichenbach first made this distinction with respect to the epistemology of mathematics:

‘the objective relation from the given entities to the solution, and the subjective way of finding it, are clearly separated for problems of a deductive character [...] we must learn to make the same distinction for the problem of the inductive relation from facts to theories’. (Reichenbach 1938: 36–37)

The standard interpretation of this statement marks contextual values, which may have contributed to the discovery of a theory, as irrelevant for *justifying* the acceptance of a theory, and for assessing how evidence bears on theory—the relation that is crucial for the objectivity of science. Contextual values are restricted to a matter of individual psychology that may influence the discovery, development and proliferation of a scientific theory, but not its epistemic status.

This distinction played a crucial role in post-World War II philosophy of science. It presupposes, however, a clear-cut distinction between epistemic values on the one hand and contextual values on the other. While this may be *prima facie* plausible for disciplines such as physics, there is an abundance of contextual values in the social sciences, for instance, in the conceptualization and measurement of a nation’s wealth, or in different ways to measure the inflation rate (cf. Dupré 2007; Reiss 2008). More generally, three major lines of criticism can be identified.

First, Helen Longino (1996) has argued that traditional ‘epistemic’ values such as consistency, simplicity, breadth of scope and fruitfulness are not purely epistemic after all, and that their use imports political and social values into contexts of scientific judgment. According to her, the use of epistemic values in scientific judgments is not always, not even normally, politically neutral—she proposes to juxtapose these values with feminist values such as novelty, ontological heterogeneity, mutuality of interaction, applicability to human needs and diffusion of power. Longino’s argument is different from the one in Section 3.1 that scientists may rationally disagree about epistemic values: it casts the entire epistemic/contextual distinction into doubt.

The use of language in descriptions of scientific hypotheses and results poses a second challenge to VNT. As Hilary Putnam has recently argued, fact and value are frequently entangled because of the use of so-called ‘thick’ ethical concepts in scientific descriptions (Putnam 2002). Consider Putnam’s own example, the word ‘cruel’. The statement ‘Susan is a

cruel teacher' entails certain statements about Susan's behavior towards her pupils, perhaps that she gives unnecessarily low grades, puts them on the spot, pokes fun at them or slaps them. It has *descriptive* content. But it also expresses our moral disapproval of Susan's behavior. To call someone cruel is to reprehend him or her. The term has therefore also *normative* content. Thick ethical terms are terms that, like cruel, have a mixed descriptive and normative content. They contrast with 'thin' ethical terms that are purely normative: 'good'/'bad', 'ought'/'must not', 'right'/'wrong' and so on.

Putnam argues at some length that (a) the normative content of thick ethical terms is ineliminable; and (b) thick ethical concepts cannot be factored into descriptive and normative components. Neither of these arguments would, if successful, necessarily cause concern for a defender of the VNT. The existence of terms in which facts and values are inextricably entangled does not pose a threat to scientists who wish to describe their hypotheses and results in a value-free manner: they could simply avoid using thick ethical terms. The crucial question is therefore whether or not scientific hypotheses and the description of results necessarily involves such terms.

John Dupré has argued that thick ethical terms are ineliminable from science, at least certain parts of it (Dupré 2007). Dupré's point is essentially that scientific hypotheses and results concern us because they are relevant to human interests, they will necessarily be couched in a language that uses thick ethical terms. While it will often be possible to translate ethically thick descriptions into neutral ones, the translation cannot be made without losses, and these losses obtain precisely because human interests are involved. According to Dupré, then, there are many scientific statements that are value-free but they are value-free because their truth or falsity does not matter to us (31):

Whether electrons have a positive or a negative charge and whether there is a black hole in the middle of our galaxy are questions of absolutely no immediate importance to us. The only human interests they touch (and these they may indeed touch deeply) are cognitive ones, and so the only values that they implicate are cognitive values.

A third challenge to VNT was posed by Richard Rudner in his influential article 'The Scientist Qua Scientist Makes Value Judgments' (Rudner 1953). Rudner disputes the core of the VNT and the context of discov-

ery/justification distinction: the idea that the acceptance of a scientific theory can in principle be value-free. We now discuss Rudner's argument in some detail.

First, Rudner argues that 'no analysis of what constitutes the method of science would be satisfactory unless it comprised some assertion to the effect that the scientist as scientist **accepts or rejects hypotheses**' (1953: 2). This assumption stems from the practice of industrial quality control and other application-oriented research. In such contexts, it is often necessary to accept or to reject a hypothesis (e.g., the efficacy of a drug) in order to make effective decisions.

Second, he notes that no scientific hypothesis is ever confirmed beyond reasonable doubt—some probability of error always remains. When we accept or reject a hypothesis, there is always a chance that our decision is mistaken. Hence, our decision is also 'a function of the *importance*, in the typically ethical sense, of making a mistake in accepting or rejecting a hypothesis' (ibid.): we are balancing the seriousness of two possible errors (erroneous acceptance/rejection of the hypothesis) against each other.

Hence, ethical judgments and contextual values enter the scientist's core activity of accepting and rejecting hypotheses, and the VNT stands refuted. Closely related arguments can be found in Churchman (1948) and Braithwaite (1953). Hempel (1965: 91-92) gives a modified account of Rudner's argument by distinguishing between judgments of *confirmation*, where values lack any logical relevance, and judgments of *acceptance*. Since even strongly confirming evidence cannot fully prove a universal scientific law, we have to live with a residual 'inductive risk' in inferring that law. Contextual values influence scientific methods by determining the acceptable amount of inductive risk.

But how general are Rudner's findings? Apparently, the result holds true of **applied science**, but not necessarily of **fundamental research**. For instance, Richard Jeffrey (1956) notes that lawlike hypotheses in theoretical science (e.g., the gravitational law in Newtonian mechanics) are characterized by their general scope and not confined to a particular application. Obviously, a scientist cannot fine-tune her decisions to their possible consequences in a wide variety of different contexts. So she should just refrain at all from accepting or rejecting hypotheses and restrict herself to gathering and interpreting the evidence. This objection was foreshadowed by the great statistician, methodologist and geneticist Ronald A. Fisher:

in the field of pure research no assessment of the cost of wrong

conclusions [...] can conceivably be more than a pretence, and in any case such an assessment would be inadmissible and irrelevant in judging the state of the scientific evidence. (Fisher 1935: 25–26, our emphasis)

By restricting scientific reasoning to gathering and interpreting evidence, possibly supplemented by assessing the probability of a hypothesis, and abandoning the business of accepting/rejecting hypotheses, Jeffrey tries to save the VNT in fundamental scientific research, and the objectivity of scientific reasoning.

A related attempt to save the VNT is given by Isaac Levi (1960). Levi observes that scientists commit themselves to certain standards of inference when they become a member of the profession. This may, for example, lead to the statistical rejection of a hypothesis when the observed significance level is smaller than 5%. These community standards may eliminate any room for contextual ethical judgment on behalf of the scientist: they determine when he/she should accept a hypothesis as established. Value judgments may be implicit in the standards of scientific inference, but not in the daily work of an *individual* scientist. Such conventional standards are especially prolific in theoretical research where it does not make sense to specify application-oriented utilities of accepting or rejecting a hypothesis (cf. Wilholt 2013). The VNT, and the idea of scientific objectivity as value freedom, could then be saved for the case of individual scientific reasoning.

Both defenses of the VNT focus on the impact of values in theory choice, either by denying that scientists actually choose theories (Jeffrey), or by referring to community standards (Levi). Douglas (2000: 563–565) points out, however, that the ‘acceptance’ of scientific theories is only one of several places for values to enter scientific reasoning, albeit an especially prominent and explicit one. Many decisions in the process of scientific inquiry may conceal implicit value judgments: the design of an experiment, the methodology for conducting it, the characterization of the data, the choice of a statistical method for processing and analyzing data, the interpretational process findings, etc. None of these methodological decisions could be made without consideration of the possible consequences that could occur. Douglas gives, as a case study, a series of experiments where carcinogenic effects of dioxine exposure on rats were probed. Contextual values such as safety and risk aversion affected the conducted research at various stages: first, in the classification of pathological samples as benign or cancerous (over which a lot of expert disagreement occurred), second, in the extrapolation from the

high-dose experimental conditions to the more realistic low-dose conditions. In both cases, the choice of a conservative classification or model had to be weighed against the adverse consequences for society that could result from underestimating the risks.

These diagnoses cast a gloomy light on attempts to divide scientific labor between gathering evidence and determining the degree of confirmation (value-free) and accepting scientific theories (value-laden). The entire process of conceptualizing, gathering and interpreting evidence is so entangled with contextual values that no neat division, as Jeffrey envisions, will work outside the narrow realm of statistical inference (cf. Biddle 2013)—and even there, doubts may be raised (see Section 4.2).

Philip Kitcher (2011a: 31–40) gives an alternative argument, based on his idea of ‘significant truths’. According to Kitcher, even staunch scientific realists will not hold that science aims at truth as a goal in itself. There are simply too many truths that are of no interest whatsoever—consider all the truths about the areas of triangles whose vertices are three arbitrarily chosen objects (106). Science, then, doesn’t aim at truth *simpliciter* but rather at something more narrow: truth worth pursuing from the point of view of our cognitive, practical and social goals. Any truth that is worth pursuing in this sense is what he calls a ‘significant truth’. Clearly, it is value judgments that help us decide whether or not any given truth is significant.

What Kitcher argues so far is consistent with the traditional view according to which values enter mainly at the first stage of scientific investigation, problem selection. But then he goes on to observe that the process of scientific investigation cannot neatly be divided into a stage in which the research question is chosen, one in which the evidence is gathered and one in which a judgment about the question is made on the basis of the evidence.

Rather, the sequence is multiply iterated. At each stage, the researcher has to decide whether previous results warrant further investigation in the same line of research or whether it would be more fruitful to switch to another avenue, even if the overall goal remains constant. These choices are laden with contextual values.

To add some precision to this idea, Kitcher distinguishes three **schemes of values**: a broad scheme, a cognitive scheme and a probative scheme. The broad scheme of values is the set of commitments around which someone’s life is organised, including personal ideals and goals and those people have for the societies they live in. The cognitive scheme of values concerns the kinds of knowledge a person values, for their own sake or for the sake of their

practical consequences. The probative scheme of values, lastly, concerns more specific questions a researcher finds worth pursuing.

Kitcher now argues that the three schemes mutually interact. Thus, the cognitive scheme might change in response to pressures from the probative and the broad scheme. We give an example of our own. Assume that the cognitive scheme endorses predictive success as an important goal of science. However, the probative scheme finds no available or conceivable strategy to reach this goal in some domain of science, for instance because that domain is characterized by strong non-linear dependencies. In this case, predictive success might have to yield to other forms of scientific knowledge. After all, we would be irrational if we continued to pursue a goal that is unattainable in principle. Conversely, changes in the broad scheme will often necessitate adjustments in the cognitive and probative schemes: changing social goals lead to revaluations of scientific knowledge and research methods.

Science, then, cannot be value-free because no scientist ever works exclusively in the supposedly value-free zone of assessing and accepting hypotheses. Evidence is gathered and hypotheses are assessed and accepted in the light of their potential for application and fruitful research avenues. Both epistemic and contextual value judgments guide these choices and are themselves influenced by their results. More than that, to portray science as value-free enterprise carries a danger with it (2011: 40):

The deepest source of the current erosion of scientific authority consists in insisting on the value-freedom of Genuine Science...

3.3 Science, Policy and the Value-Free Ideal

While the previous discussion focused on the VNT, and the practical attainability of the VFI, little has been said about whether value freedom is desirable in the first place. This subsection discusses this topic with special attention to informing and advising public policy from a scientific perspective. While the VFI, and many arguments for and against it, can be applied to science as a whole, the interface of science and public policy is the place where the intrusion of values into science is especially salient, and where it is surrounded by the greatest controversy. Quite recently, the reproach that climate scientists were pursuing a particular socio-political agenda (the ‘Climategate’ affair) did much to damage the authority of science in the public arena.

Indeed, many debates at the interface of science and public policy are

characterized by disagreements on propositions that combine a factual basis with specific goals and values. Take, for instance, the view that growing transgenic crops carries too much risk in terms of biosecurity, or that global warming needs to be addressed by cutting CO₂ emission. The critical question in such debates is whether there are theses T such that one side in the debate endorses T , the other side rejects it, the evidence is shared, and both sides have good reasons for their respective positions.

According to the VFI, scientists should uncover an epistemic, value-free basis for resolving such disagreements and restrict the dissent to the realm of value judgments. Even if the VNT should turn out to be untenable, and a strict separation to be impossible, the VFI may have an important function for *guiding* scientific research and for minimizing the impact of values on an objective science. In the philosophy of science, one camp of scholars defends the VFI as a necessary antidote to individual and institutional interests, like Hugh Lacey (1999, 2002), Ernan McMullin (1982) and Sandra Mitchell (2004), while others adopt a critical attitude, like Helen Longino (1990, 1996), Philip Kitcher (2011a) or Heather Douglas (2009). These criticisms may refer to the desirability, attainability or the conceptual (un)clarity of the VFI. We begin, however, with the defenders of the VFI.

Lacey distinguishes three components or interpretations of the VFI: impartiality, neutrality and autonomy. **Impartiality** implies that theories are solely accepted or appraised in virtue of their contribution to the epistemic values of science, such as truth, accuracy or explanatory power. In particular, the choice of theories is not influenced by contextual values. **Neutrality** means that scientific theories make no value statements about the world: they are concerned with what there is, not with what there should be. Finally, scientific **autonomy** means that the scientific agenda is shaped by the desire to increase scientific knowledge, and that contextual values have no place in scientific method.

These three interpretations of the VFI can be combined with each other, or stand individually. All of them, however, are subject to severe criticism. First, on the descriptive level, it is clear that autonomy of science often fails in practice due to the presence of external interests, e.g., funding agencies and industry lobbies. Neutrality is questionable in the light of the implicit role of values in social science, e.g., in Rational Choice Theory (see Section 5.2). Impartiality has been criticized in the above discussion of the VNT.

In addition, it has been argued that the VFI is not desirable at all. Feminist philosophers (e.g., Harding 1991; Okruhlik 1994; Lloyd 2005) have

argued that science often carries a heavy androcentric values, for instance in biological theories about sex, gender and rape. The charge against these values is not so much that they are contextual rather than epistemic, but that they are unjustified. The explicit consideration of feminist values may act as a helpful antidote, in stark contrast with the VFI. Moreover, if scientists did follow the VFI rigidly, policy-makers would pay even less attention to them, with a detrimental effect on the decisions they take (Cranor 1993). Given these shortcomings, the VFI has to be rethought if it is supposed to play a useful role for guiding scientific research and leading to better policy decisions.

Douglas (2009: 7-8) proposes that the epistemic authority of science can be detached from its autonomy by distinguishing between **direct** and **indirect roles for values in science**. The assessment of evidence may legitimately be affected indirectly by contextual values: they may determine how we interpret noisy datasets, what is the appropriate standard of evidence for a specific claim, how the severity of consequences of a decision must be assessed, and so on. This concerns, above all, policy-related disciplines such as climate science or economics that routinely perform scientific risk analyses for real-world problems (cf. also Shrader-Frechette 1991). What must not happen, however, is that contextual values trump scientific evidence, or are used as a reason to ignore evidence:

cognitive, ethical and social values all have legitimate, indirect roles to play in the doing of science [...]. When these values play a direct role in the heart of science, problems arise as unacceptable reasoning occurs and the reason for valuing science is undermined. (Douglas 2009: 108)

Douglas's conception of objectivity emphasizes a prohibition for values to replace or dismiss scientific evidence—she calls this **detached objectivity**—but it is complemented by various other aspects that relate to a reflective balancing of various perspectives and the procedural, social aspects of science (ch. 6). Instead of subscribing to the traditional VFI, Douglas suggests to rescue scientific integrity and objectivity by 'keeping values to their proper roles' (175).

That said, Douglas' proposal is not very concrete when it comes to the implementation level, e.g. regarding the way diverse values should be balanced. Compromising in the middle cannot be the solution (Weber 1917/88). First, no standpoint is, just in virtue of being a middle position, evidentially supported vis-à-vis more extreme positions. Second, these middle positions

are also, from a practical point of view, the least functional when it comes to advising policy-makers.

Moreover, the distinction between direct and indirect roles of values in science may not be sufficiently clear-cut to police the legitimate use of values in science. Douglas (2009, 96) distinguishes between values as ‘reasons in themselves’, that is, treating them as evidence or defeaters for evidence (direct role, illegitimate) and as ‘helping to decide what should count as a *sufficient* reason for a choice’ (indirect role, legitimate). But can we always draw a neat borderline? Assume that a scientist considers, for whatever reason, the consequences of erroneously accepting hypothesis H undesirable. Therefore he uses a statistical model whose results are likely to favor $\neg H$ over H. Is this a matter of reasonable conservativeness? Or doesn’t it amount to reasoning to a foregone conclusion, and to treating values as evidence (cf. Elliott 2011, 320–321)?

The most recent literature on values and evidence in science presents us with a broad spectrum of opinions. Steele (2012) bolsters Douglas’s approach by arguing that various probabilistic assessments of uncertainty, e.g., imprecise probability, involve contextual value judgments as well. Betz (2013) argues, by contrast, that scientists can largely avoid making contextual value judgments if they carefully express the uncertainty involved with their evidential judgments, e.g., by using a scale ranging from purely qualitative evidence (such as expert judgment) to precise probabilistic assessments. The issue of value judgments at earlier stages of inquiry is not addressed by this proposal; however, disentangling evidential judgments and judgments involving contextual values at the stage of theory assessment may be a good thing in itself.

Thus, should we or should we not be worried about values in scientific reasoning? While Douglas and others make a convincing case that the interplay of values and evidential considerations need not be pernicious, it is unclear why it *adds* to the success or the authority of science. After all, the values of an individual scientist who makes a risk assessment, need not agree with those of society. How are we going to ensure that the permissive attitude towards values in setting evidential standards etc. is not abused? In the absence of a general theory about which contextual values are beneficial and which are pernicious, can’t we defend the VFI as a first-order approximation to a sound, transparent and objective science? Science seems to require *some* independence from contextual values in order to maintain its epistemic authority.

3.4 Feyerabend: The Tyranny of the Rational Method

The last part of the section on science and values deals with Paul Feyerabend's radical assault on the rationality and objectivity of scientific method (see also <http://plato.stanford.edu/entries/feyerabend/>). His position is exceptional in the philosophical literature since traditionally, the threat for objectivity is located in contextual rather than epistemic values. Feyerabend turns this view upside down: it is the 'tyranny' of rational method, and the emphasis on epistemic rather than contextual values that prevents us from having a science in the service of society. Thus, Feyerabend vociferously denies the VFI and also the VNT by his claim that Western science is loaded with all kinds of pernicious values.

Feyerabend's writings on objectivity and values in science have an epistemic as well as a political dimension. Regarding the first, the leading philosophy of science figures in Feyerabend's young days such as Carnap, Hempel and Popper characterized scientific method in terms of rules for rational scientific reasoning. Some of them, like Popper, devoted special attention to the demarcation of science from 'pseudo-science' and imposture, thereby derogating other traditions as irrational or, in any case, inferior. Feyerabend thinks, however, that science must be protected from a 'rule of rationality', identified with strict adherence to scientific method: such rules only suppress an open exchange of ideas, extinguish scientific creativity and prevent a free and truly democratic science.

In his classic 'Against Method' (1975, chs. 8-13), Feyerabend elaborates this criticism from a famous episode in the history of science: the development of Galilean mechanics and the discovery of the Jupiter moons. In superficial treatments of this episode, it is stressed that an obscurantist and value-driven Catholic Church forced Galilei to recant from a scientifically superior position backed by value-free, objective findings. But in fact, Feyerabend argues, the Church had the better arguments by the standards of 17th century science. Their conservatism regarding their *Weltanschauung* was scientifically backed: Galilei's telescopes were unreliable for celestial observations, and many well-established phenomena (no fixed star parallax, invariance of laws of motion) could at first not be explained in the heliocentric system. Hence, scientific method was not on Galilei's side, but on the side of the Church who gave preference to the old, Ptolemaic world-view. With hindsight, Galilei managed to achieve groundbreaking scientific progress just because he deliberately violated rules of scientific reasoning, because he stubbornly stuck to a problematic approach until decisive theo-

retical and technological innovations were made. Hence Feyerabend's dictum 'Anything goes': no methodology whatsoever is able to capture the creative and often irrational ways by which science deepens our understanding of the world.

The drawbacks of an objective, value-free and method-bound view on science and scientific method are not only epistemic. Such a view narrows down our perspective and makes us less free, open-minded, creative, and ultimately, less human in our thinking (1975: 154). It is therefore neither possible nor desirable to have an objective, value-free science (cf. 1978: 78–79). As a consequence, Feyerabend sees traditional forms of inquiry about our world (e.g., Chinese medicine) on a par with their Western competitors. He denounces appeal to objective standards as a reason for preferring one's own worldview:

there is hardly any difference between the members of a 'primitive' tribe who defend their laws because they are the laws of the gods [...] and a rationalist who appeals to 'objective' standards, except that the former know what they are doing while the latter does not' (1978: 82)

In other words, the defenders of scientific method abuse the word 'objective' for proving the superiority of Western science vis-à-vis other worldviews. To this, Feyerabend adds that when dismissing other traditions, we actually project our own worldview, and our own value judgments, into them instead of making an impartial comparison (1978: 80–83). There is no purely rational justification for dismissing other perspectives in favor of the Western scientific worldview. To illustrate his point, Feyerabend compares the defenders of a strong, value-free notion of objectivity to scientists who stick to the concepts of absolute length and time in spite of the Theory of Relativity. A staunch defense of objectivity and value freedom may just expose our own narrow-mindedness. This is not meant to say that truth loses its function as a normative concept in science, nor that all scientific claims are equally acceptable. Rather, Feyerabend demands that we move toward a genuine **epistemic pluralism** that accepts diverse approaches to searching an acquiring knowledge. In such an epistemic pluralism, science may regain its objectivity in the sense of respecting the diversity of values and traditions that drive our inquiries about the world (1978: 106–107).

All this has a political aspect, too. In the times of the scientific revolution or the Enlightenment, science acted as a liberating force that fought intellectual and political oppression by the sovereign, the nobility or the

clergy. Nowadays, Feyerabend continues, the ideals of value-freedom and objectivity are often abused for excluding non-experts from science, proving the superiority of the Western way of life, and undergirding the power of an intellectual elite. (Here it is important to keep in mind that Feyerabend's writings on this issue date mostly from the 1970s and were much influenced by the Civil Rights Movement in the US and the increasing emancipation of minorities, such as Blacks, Asians and Hispanics.) Feyerabend therefore argues that democratic societies need to exert much greater control over scientific research. Laymen have to supervise science. This includes areas where even defenders of the VFI refrain from demanding value freedom, such as setting up a research agenda, distributing funds, and supervising scientific inquiry. But it also concerns areas that are more central to the VFI, such as the choice of a research method or the assessment of scientific theories. Contrary to commonly held beliefs, lack of specialized training need not imply lack of relevant knowledge (1975: xiii). Feyerabends sums up his view as follows:

a community will use science and scientists in a way that agrees with its values and aims and it will correct the scientific institutions in its midst to bring them closer to these aims (1975: 251)

4 Objectivity as Freedom from Personal Biases

This section deals with scientific objectivity as a form of intersubjectivity—as freedom from personal biases. According to this view, science is objective to the extent that personal biases are absent from scientific reasoning, or that they can be eliminated in a social process. Perhaps all science is necessarily perspectival. Perhaps we cannot sensibly draw scientific inferences without a host of background assumptions, which may include assumptions about values. But scientific results should certainly not depend on researchers’ personal preferences or idiosyncratic experiences. That, among other things, is what distinguishes science from the arts and other more individualistic human activities—or so it is said. A paradigmatic way to achieve objectivity in this sense are measurement and quantification. What has been measured and quantified has been verified relative to a standard. The truth, say, that the Eiffel Tower is 324 meters tall is relative to a standard unit and conventions about how to use certain instruments, so it is neither aperspectival nor free from assumptions, but it is independent from the person making the measurement.

We will begin with a discussion of objectivity, so conceived, in measurement, discuss the ideal of ‘mechanical objectivity’ and then investigate to what extent freedom from personal biases can be implemented in statistical and inductive inference—arguably the core of scientific reasoning, especially in experimentally working sciences.

4.1 Measurement and Quantification

Measurement is often thought to epitomize scientific objectivity, most famously captured in Lord Kelvin’s dictum ‘when you cannot express it in numbers, your knowledge is of a meagre and unsatisfactory kind; it may be the beginning of knowledge, but you have scarcely, in your thoughts, advanced to the stage of *science*, whatever the matter may be’.

Measurement can certainly achieve some independence of perspective. Yesterday’s weather in Durham UK may have been ‘really hot’ to the average North Eastern Brit and ‘very cold’ to the average Mexican, but they’ll both accept that it was 21° C. Clearly, however, measurement does not result in a ‘view from nowhere’, nor are typical measurement results free from presuppositions. Measurement instruments interact with the environment, and so results will always be a product of both the properties of the environment we aim to measure as well as the properties of the instrument.

Instruments, thus, provide a perspectival view on the world (cf. Giere 2006).

Moreover, making sense of measurement results requires interpretation. Consider temperature measurement. Thermometers function by relating an unobservable quantity, temperature, to an observable quantity, expansion (or length) of a fluid or gas in a glass tube; that is, thermometers measure temperature by assuming that length is a function of temperature: $\text{length} = f(\text{temperature})$. The function f is not known *a priori*, and it cannot be tested either (because it could in principle only be tested using a *veridical* thermometer, and the veridicality of the thermometer is just what is at stake here). Making a specific assumption, for instance that f is linear, solves that problem by fiat. But this ‘solution’ does not take us very far because different thermometric substances (e.g., mercury, air or water) yield different results for the points intermediate between the two fixed points 0°C and 100°C , and so they can’t all expand linearly.

In Hasok Chang’s account of early thermometry (Chang 2004), the problem was eventually solved by using a ‘principle of minimalist overdetermination’, whose goal it was to find a reliable thermometer while making as few substantial assumptions (e.g., about the form for f) as possible. It was argued, eventually, that if a thermometer was to be reliable, different tokens of the same thermometer type should agree with each other, and the results of air thermometers agreed the most. ‘Minimal’ doesn’t mean zero, however, and indeed this procedure makes an important presupposition (in this case a metaphysical assumption about the one-valuedness of a physical quantity). Moreover, the procedure yielded at best a reliable instrument, not necessarily one that was best at tracking the uniquely real temperature (if there is such a thing).

What Chang argues about early thermometry is true of measurement more generally. Measurements are always made against a backdrop of metaphysical presuppositions, theoretical expectations and other kinds of belief. Whether or not any given procedure is regarded as adequate depends to a large extent on the purposes pursued by the individual scientist or group of scientists making the measurements. Especially in the social sciences, this often means that measurement procedures are laden with normative assumptions, i.e., values.

Julian Reiss (2008, 2013) has argued that economic indicators such as consumer price inflation, gross domestic product and the unemployment rate are value-laden in this sense. Consumer-price indices, for instance, assume that if a consumer prefers a bundle x over an alternative y , then x is better

for her than y, which is as ethically charged as it is controversial. National income measures assume that nations that exchange a larger share of goods and services on markets are richer than nations where the same goods and services are provided by the government, which is as ethically charged and controversial.

While not free of assumptions and values, the goal of many measurement procedures remains to reduce the influence of personal biases and idiosyncrasies. The Nixon administration, famously, indexed social security payments to the consumer-price index in order to eliminate the dependence of security recipients on the flimsies of party politics: to make increases automatic instead of a result of political negotiations (Nixon 1969). Lorraine Daston and Peter Galison refer to this as **mechanical objectivity**. They write:

Finally, we come to the full-fledged establishment of mechanical objectivity as the ideal of scientific representation. What we find is that the image, as standard bearer of is objectivity is tied to a relentless search to replace individual volition and discretion in depiction by the invariable routines of mechanical reproduction. (Daston and Galison 1992: 98)

The artist Salvador Dalí, no doubt unwittingly, describes his surrealist paintings as a product of mechanical objectivity in Daston and Galison's sense:

In truth I am no more than an automaton that registers, without judgment and as exactly as possible, the dictate of my subconscious: my dreams, hypnagogic images and visions, and all the concrete and irrational manifestations of the dark and sensational world discovered by Freud. (Dalí 1935)

Mechanical objectivity reduces the importance of human contributions to scientific results to a minimum, and therefore enables science to proceed on a large scale where bonds of trust between individuals can no longer hold (Daston 1992). Trust in mechanical procedures thus replaces trust in individual scientists.

In his book *Trust in Numbers*, Theodore Porter pursues this line of thought in great detail. In particular, on the basis of case studies involving British actuaries in the mid-nineteenth century, of French state engineers

throughout the century, and of the US Army Corps of Engineers from 1920 to 1960, he argues for two causal claims. First, measurement instruments and quantitative procedures originate in commercial and administrative needs and affect the ways in which the natural and social sciences are practiced, not the other way around. The mushrooming of instruments such as chemical balances, barometers, chronometers was largely a result of social pressures and the demands of democratic societies. Administering large territories or controlling diverse people and processes is not always possible on the basis of personal trust and thus 'objective procedures' (which do not require trust in persons) took the place of 'subjective judgments' (which do). Second, he argues that quantification is a technology of distrust and weakness, and not of strength. It is weak administrators who do not have the social status, political support or professional solidarity to defend their experts' judgments. They therefore subject decisions to public scrutiny, which means that they must be made in a publicly accessible form.

This is the situation in which scientists who work in areas where the science/policy boundary is fluid find themselves:

The National Academy of Sciences has accepted the principle that scientists should declare their conflicts of interest and financial holdings before offering policy advice, or even information to the government. And while police inspections of notebooks remain exceptional, the personal and financial interests of scientists and engineers are often considered material, especially in legal and regulatory contexts.

Strategies of impersonality must be understood partly as defenses against such suspicions [...]. Objectivity means knowledge that does not depend too much on the particular individuals who author it. (Porter 1995: 229)

Measurement and quantification help to reduce the influence of personal biases and idiosyncrasies and they reduce the need to trust the scientist or government official, but often at a cost. Standardizing scientific procedures becomes difficult when their subject matters are not homogenous, and few domains outside fundamental physics are. Attempts to quantify procedures for treatment and policy decisions that we find in evidence-based practices are currently transferred to a variety of sciences such as medicine, nursing, psychology, education and social policy. However, they often lack a certain

degree of responsiveness to the peculiarities of their subjects and the local conditions to which they are applied (see also section 5.3).

Moreover, the measurement and quantification of characteristics of scientific interest is only half of the story. We also want to describe relations between the quantities and make inferences using statistical analysis. Statistics thus helps to quantify further aspects of scientific work. We will now turn to the question whether or not statistical analysis can proceed in a way free from personal biases and idiosyncrasies.

4.2 Inductive and Statistical Inference

The appraisal of scientific evidence is traditionally regarded as a domain of scientific reasoning where the ideal of scientific objectivity has strong normative force, and where it is also well-implemented into scientific practice. Episodes such as Galilei's observations of the Jupiter moons, Lavoisier's calcination experiments, and Eddington's observation of the 1919 eclipse are found in all philosophy of science textbooks because they exemplify how evidence can be persuasive and compelling to scientists with different backgrounds. The crucial question is therefore: can we identify an 'objective' concept of scientific evidence that is independent of the personal biases of the experimenter and interpreter?

Inferential statistics—the field that investigates the validity of inferences from data to theory—tries to answer this question. It is extremely influential in modern science. For instance, the recent discovery of the Higgs Boson was established through a statistical argument, based on the idea that what is 'significant' for one researcher should also be significant for another. Therefore, this subsection contrasts several schools of inductive/statistical inference and compares them with respect to their objectivity claims.

Logical Probability and Bayesian Inference

In Section 3.2, we have encountered the idea of a scientist who assesses the probability of a hypothesis in the light of the evidence—a conception based on Rudolf Carnap's monumental work 'Logical Foundations of Probability' (1950/62). Carnap is interested in determining the **degree of confirmation** of a hypothesis relative to a given set of observations. To adequately explicate this concept, Carnap first investigates a clear-cut, simple setting: a logical first-order language \mathcal{L} with finitely many monadic predicates and infinitely many individual constants. Carnap then considers all maximally

specific descriptions of states of the world in \mathcal{L} and assigns a measure m over these descriptions. Subsequently, the available observations E are conceptualized as a conjunction of \mathcal{L} -statements. Finally, the degree of confirmation of hypothesis H , represented as an \mathcal{L} -proposition, is calculated according to the formula:

$$c(H, E) = \frac{m(H \wedge E)}{m(E)} \quad (1)$$

In other words, the degree of confirmation of hypothesis H relative to E is the conditional probability of H given E . But is this probability an objective quantity and free from personal bias?

The answer to this question stands and falls with the choice of the m -function. Not all are equally suited: for example, assigning equal weight to all complete state descriptions would not allow for learning from experience. Carnap himself admits that no conclusive argument for a specific m can be made. In later work, he therefore parametrized the choice of a m -function by a parameter λ that expressed the respective impact of logical and empirical considerations on prediction.

Carnap's probabilty logic has often been described as an 'inductive machinery'. We now see where these comments miss the point: they neglect the subjective choices on which the inductive inferences are built. Carnap's approach is objective in so far as the degree of confirmation is intersubjectively compelling once a logical language and appropriate symmetry principles for this language are agreed upon; however, it is subjective in the sense that rational agents may disagree on these principles.

Closely related to Carnap's logical probability framework is the Bayesian approach to confirmation and evidence, developed first by Frank Ramsey. It is outspokenly subjective: probability is used for quantifying a scientist's subjective **degrees of belief** in a particular hypothesis. These degrees of belief are changed by **conditionalization** on observed evidence E and making use of **Bayes' Theorem**:

$$p_{\text{new}}(H) := p(H|E) = p(H) \frac{p(E|H)}{p(E)} \quad (2)$$

The credibility of a hypothesis H in the face of evidence E is then judged by its conditional probability $p(H|E)$, commonly called the **posterior (probability)** of H , as opposed to the **prior (probability)** $p(H)$. These days, the Bayesian approach is extremely influential in philosophy, but also in scientific disciplines such as statistics, economics, and biology. The difference to Carnap lies in the philosophical motivation, and in the different

understanding of confirmation judgments: for Carnap, they are primarily a consequence of certain logical ways of carving the world at its joints; for the Bayesian, they express a genuinely subjective uncertainty judgment. See also <http://plato.stanford.edu/entries/bayes-theorem/>.

Can we ground the objectivity of scientific evidence in a framework that is explicitly based on personal attitudes? Some authors have argued that measuring ‘mere psychological tendencies’ (Fisher 1935: 6–7) such as degrees of belief cannot be relevant for scientific inquiry. Bayesians have supplied several arguments to the effect that subjective probability is not equal to personal bias, which we will review in turn.

1. As argued by Howson (2003) and Howson and Urbach (1993), the Bayesian’s aim is not to determine an intersubjectively binding degree of confirmation, but to provide sound inference rules for learning from experience. In the same way that deductive logic does not judge the correctness of the premises but just advises you what to infer from them, **Bayesian inductive logic** tells you how to change your own attitudes as soon as you encounter evidence. All other updating rules are susceptible to so-called *Dutch books*: betting based on following such rules will lead to sure monetary losses. In addition, convergence theorems guarantee that, as long as novel evidence keeps coming in, the degrees of belief of agents with very different initial attitudes will finally converge (Gaifman and Snir 1982).

Reply: This argument answers a different question. The worry is not about the soundness of the updating process from a prior to a posterior probability. Rather, since any posterior probability can be justified by choosing an appropriate prior, the resulting posterior may still be beset with idiosyncratic bias and manifest social values. The convergence theorems provide no relief since they hold in the limit only, and do not indicate at which point our judgment is free from personal bias.

2. Modern Bayesians distinguish between the **absolute** and **incremental concept of confirmation**. The absolute concept of confirmation measures the credibility of a hypothesis H in the light of evidence E , and it is proportional to the prior of H (see equation (2)). The incremental concept, however, measures the impact of E on H , in terms of the strength of belief change. Several philosophers have argued that there is just one incremental measure of confirmation that satisfies a

set of generally accepted desirable constraints (e.g., Good 1950; Kemeny and Oppenheim 1952; Crupi, Tentori and Gonzalez 2007). If one of these arguments were sound, then the incremental degree of confirmation would provide a bias-free assessment of the evidence.

Reply: This approach is promising, but to date, none of the ‘uniqueness theorems’ for a specific measure of confirmation has gained general acceptance. On the other hand, Bayesian statisticians almost uniformly use the **Bayes factor**, that is, the ratio of prior to posterior odds in favor of a hypothesis. In the case of two parametric hypotheses $H_0 : \theta \in \Theta_0$ vs. $H_1 : \theta \in \Theta_1$, the Bayes factor can be written as:

$$B_{01}(x) := \frac{p(H_0|x)}{p(H_1|x)} \cdot \frac{p(H_1)}{p(H_0)} = \frac{\int_{\theta \in \Theta_0} p(x|\theta)p(\theta)d\theta}{\int_{\theta \in \Theta_1} p(x|\theta)p(\theta)d\theta}. \quad (3)$$

The Bayes factor has a number of desirable properties in inference and model selection (Kass and Raftery 1995), and it also corresponds to a particularly popular confirmation measure, the log-likelihood measure l . Although not free from subjective assumptions, the Bayes factor may be our best candidate for a bias-free assessment of the evidence.

3. The Bayesian approach can eliminate personal bias by imposing additional constraints on an agent’s rational degrees of belief. One way to do so consists in adopting the **Principle of Maximum Entropy** or MaxEnt, going back to Jaynes (1968) and developed philosophically by Jon Williamson (2010). Williamson retains the demand that degrees of belief satisfy the probability axioms, but dismisses the updating by conditionalization in equation (2). Instead, he demands that the agent’s degrees of belief must be in sync with empirical constraints and that, conditional on these constraints, they must be **equivocal**, that is, as middling as possible. This latter constraint amounts to maximizing the entropy of the probability distribution in question. If ω denotes the ‘atoms’ of the relevant σ -algebra, the entropy is given by the term

$$H = - \sum_{\omega \in \Omega} p(\omega) \log p(\omega). \quad (4)$$

Since maximizing the right hand side in (4) leads to a unique solution, subjective bias is eliminated. Instead of going for MaxEnt, one could also assume **objective priors**, that is, prior probabilities which do not represent an agent’s factual attitudes, but are determined by principles

of symmetry, mathematical convenience or maximizing the influence of the data on the posterior (e.g, Jeffreys 1939/80; Bernardo 2012).

Reply: In general, the achievements of the various objective Bayesian approaches come at the expense of weakening their philosophical foundations. For example, the viability of Williamson’s approach depends on the persuasiveness of his arguments in favor of the equivocation norm, which fall, as of today, short of convincing a skeptic (cf. Williamson 2010, ch. 3). Similarly, arguments for objective priors depend on particular ways of conceptualizing inference problems. On the other hand, these objections may be outweighed by the substantial practical benefits of objective Bayesian inference (Sprenger 2012).

Thus, Bayesianism provides not more than a partial answer to securing scientific objectivity from personal idiosyncrasy. On the other hand, the objections to the above proposals are no K.O.-arguments, and further developments of the discussed approaches may help to reconcile transparency about subjective assumptions with objectivity in interpreting statistical evidence.

That said, one may argue that the theories we discussed so far all miss the target. Bayesians primarily address the question of which theories we should rationally **believe** in. The decision procedures reviewed in Section 3.2 respond to the question of what we should **do**. Both analyze the concept of **statistical evidence** from the vantage point of their primary focus—beliefs and decisions. But can’t we quantify the support for or against a hypothesis in an intersubjectively compelling way, without buying into a subjectivist or a behavioral framework? This is the ambition of frequentist and likelihood-based explications of scientific evidence.

Frequentist Inference

The frequentist conception of evidence is based on the idea of the **statistical test of a hypothesis**. Under the influence of the statisticians Neyman and Pearson, tests were often regarded as rational **decision procedures** that minimize the relative frequency of wrong decisions in a hypothetical series of repetitions of a test. As we have seen in Section 3.2., the choice of thresholds in such tests may reflect contextual value judgments and personal bias. Moreover, the losses associated with erroneously accepting or rejecting that hypothesis depend on the context of application which may be unbeknownst

to the experimenter. This speaks for a division of labor where scientists restrict themselves to an **evidential interpretation of statistical tests**, and leave the actual decisions to policy-makers and regulatory agencies.

Such an approach has been developed by Ronald A. Fisher's (1890–1962) and has become the orthodox solution to statistical inference problems. Fisher reasons that if an observed result is 'extreme', then it provides evidence against that hypothesis. In other words, if a result has lower probability under the null hypothesis H than most other possible results, then it undermines the tenability of H :

either an exceptionally rare chance has occurred, or the theory
[=the null hypothesis] is not true. (Fisher 1956: 39)

Then, the strength of evidence against the tested hypothesis is equal to the **p-value**—the probability of obtaining a result that is as least as extreme as the actually observed data. Figure 1 gives a graphical illustration. This probability measures how strongly E speaks against H , compared to other possible results, and the lower it is, the stronger the evidence against H . Fisher interpreted the p-value as 'a measure of the rational grounds for the *disbelief* [in the null hypothesis] it augments' (Fisher 1956, 43). Conventionally, a p-value smaller than .05 is classified as 'significant evidence', a p-value smaller than .01 is called 'very significant evidence', etc. This basic idea is also underlying Mayo's (1996) error-statistical theory of evidence, although she focus on the 'severity' of a test (that is, the observed power).

This concept of evidence is apparently objective, but beset with a variety of problems (see Sprenger 2014 for a detailed discussion). From a practical point of view, it turns out that most researchers have **problems to correctly interpret p-values**. For instance, the p-value is often regarded as a posterior probability of the tested hypothesis (Oakes 1986), or its significance vis-à-vis effect size is exaggerated, although the latter is often the scientifically more interesting quantity (Ziliak and McCloskey 2008). Personal bias enters statistical inference by means of widespread interpretational mistakes. Moreover, the conventional use of the famous 5% threshold for statistical significance is clearly arbitrary and anything but objective. There is no intersubjectively compelling justification why this particular standard of evidence should be used. As statisticians put it, 'God loves the .06 [i.e., a 6% threshold] nearly as much as the .05' (Rosnow and Rosenthal 1989).

From an institutional point of view, the frequentist conception of p-values is problematic as well. Not only does the automatic use of conventional sig-

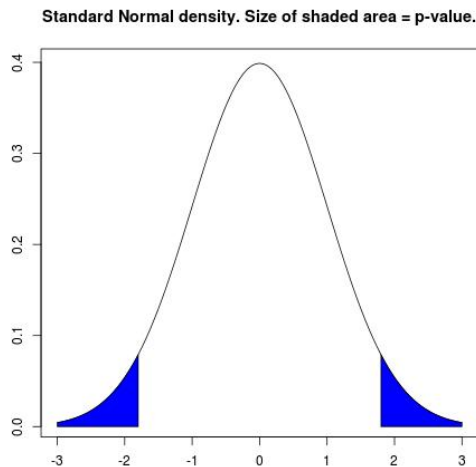


Figure 1: The p-value, represented as the shaded area under the probability density of the tested hypothesis $X \sim N(0, 1)$, for observed data $x = 1.8$ ($p = 0.072$).

nificance levels imply that much valuable research is suppressed, because ‘insignificant’ results have no chance of publication. Ioannidis (2005) even claims that **most published research findings are false**. Indeed, researchers often fail to replicate findings by another scientific team, and periods of excitement and subsequent disappointment are not uncommon in frontier science. Why is this the case? Ioannidis reasons that there are many false causal claims that may be erroneously supported and yield a publishable research finding. The probability that this happens by accident is plausibly larger than the probability that we detect one of the relatively rare true causal relations. This is not only a feature of scientific inquiry in general, but specifically due to the frequentist logic of evidence: the one-time achievement of a significant result provides evidence in the frequentist sense, but it is not a very good indicator for the tenability of a hypothesis.

Finally, there is a principled philosophical objection against the objectivity of frequentist evidence: the **sample space dependence**. That is, in frequentist statistics, the strength of the evidence depends on *which results could have been observed* (but were not observed). For instance, the post-experimental assessment of the evidence has to be changed when we learn about a defect in our measurement instrument, even if that defect is not relevant for the range of the actually observed results! On a Bayesian reading,

this implies that frequentist evidence statements depend on the *intentions* of the experimenter (Edwards, Lindman and Savage 1963; Sprenger 2009): Would she have continued the trial if the results had been different? How would she have reacted to unforeseen circumstances? This feature of frequentist inference jeopardizes claims that frequentist evidence is objective and free from personal bias.

Likelihood-Based Inference

A middle ground between frequentist and Bayesian inference is provided by likelihoodist inference, based on Alan Turing and I.J. Good’s work on breaking the Enigma code during World War II. Their research yielded a theory of evidence where instead of asking to what extent evidence E supports hypothesis H , it is measured how much E favors H over an alternative H' (Good 1950). In this approach, the canonical measure of **weight of evidence** w is explicated as

$$w(E, H, H') = \log \frac{p(E|H)}{p(E|H')} \quad (5)$$

A conception of statistical evidence based on w or monotonic transformations of w is usually called a **likelihood-based theory of evidence**. This is because the probabilities of the actual evidence E under the competing hypotheses are called the **likelihoods of H on E** .

Whenever H and H' imply specific statistical distributions for the observed evidence, this measure can be calculated in an objective, subject-independent way. Therefore, a minority of statisticians (e.g., Royall 1997) uses this measure as a fundamental notion of evidence, and philosophers such as Hacking (1965) and Sober (2008) have applied this concept of evidence to various issues in formal epistemology, inductive inference and philosophy of biology.

However, the scope of w is restricted. In routinely testing the values of the mean or variance of a distribution, H' is often a composite hypothesis whose components are indexed by a parameter θ (e.g., $H : \theta = \theta_0$ is tested against $H' : \theta \neq \theta_0$). A Bayesian would then appeal to subjective belief attitudes toward the individual hypotheses H_θ , calculate $p(E) = \sum_\theta p(H_\theta)p(E|H_\theta)$ and update her beliefs by conditionalization. However, the likelihoodist cannot use subjective probability in order to transform a composite hypothesis into a simple one. Therefore, for composite statistical hypotheses, we either have to compromise the objectivity of $w(E, H, H')$ by

introducing subjective weightings, or we have to weaken the conceptual arguments in favor of w , e.g., by switching to comparing maximum likelihoods instead of average likelihoods.

Summing up our findings, no statistical theory of evidence manages to eliminate all sources of personal bias and idiosyncrasy. The Bayesian is honest about it: she considers subjective assumptions to be ineliminable from scientific reasoning. This does not rule out that contrastive aspects of statistical evidence may be quantified in an objective way, e.g., by using Bayes factors. On that point, she agrees with the likelihoodist, who has perhaps the ‘purest’ model of objective statistical evidence, but also the one with the most restricted scope. The frequentist conception based on p-values still dominates statistical practice, but it suffers from several conceptual drawbacks, and in particular the misleading impression of objectivity. This also has far-reaching implications for fields such as evidence-based medicine, where randomized controlled trials (the most valuable source of evidence) are typically interpreted in a frequentist way. A defense of frequentist inference should, in our opinion, stress that the relatively rigid rules for interpreting statistical evidence facilitate communication and assessment of research results in the scientific community—something that is harder to achieve for a Bayesian.

5 Issues in the Special Sciences

So far everything we discussed was meant to apply across all or at least most of the sciences. In this section we will look at a number of specific issues that arise in the social science, in economics, and in evidence-based medicine.

5.1 Max Weber and Objectivity in the Social Sciences

There is a long tradition in the philosophy of social science maintaining that there is a gulf in terms of both goals as well as methods between the natural and the social sciences. This tradition, associated with thinkers such as the neo-Kantians Heinrich Rickert and Wilhelm Windelband, the hermeneuticist Wilhelm Dilthey, the sociologist-economist Max Weber, and the twentieth-century hermeneuticists Hans-Georg Gadamer and Michael Oakeshott, holds that unlike the natural sciences whose aim it is to establish natural laws and which proceed by experimentation and causal analysis, the social sciences seek understanding (*Verstehen*) of social phenomena, the interpretive examination of the meanings individuals attribute to their actions (Rickert 1929/1986; Windelband 1915; Dilthey 2002; Weber 1904/1949b; Gadamer 1989; Oakeshott 1933).

Understood this way, social science lacks objectivity in more than one sense. One of the more important debates concerning objectivity in the social sciences concerns the role value judgments play and, importantly, whether value-laden research entails claims about the desirability of actions. Max Weber held that the social sciences are necessarily value laden. However, they can achieve some degree of objectivity by keeping out the social researcher's views about whether agents' goals are commendable. In a similar vein, contemporary economics can be said to be value laden because it predicts and explains social phenomena on the basis of agents' preferences. Nevertheless, economists are adamant that economists are not in the business of telling people what they ought to value. Modern economics is thus said to be objective in the Weberian sense of **'absence of researchers' values'**—a conception that we discussed in detail in Section 3.

In his widely cited essay 'Objectivity in Social Science and Social Policy' (Weber 1904/1949a), Weber argued that the idea of an aperspectival social science was meaningless (72/81):

There is no absolutely objective scientific analysis of [...] 'social phenomena' independent of special and 'one-sided' viewpoints

according to which expressly or tacitly, consciously or unconsciously they are selected, analyzed and organized for expository purposes.

All knowledge of cultural reality, as may be seen, is always knowledge from particular points of view.

The reason for this is twofold. First, reality is too complex to admit of full description and explanation. So we have to select. But, perhaps in contraposition to the natural sciences, we cannot just select those aspects of the phenomena that fall under universal natural laws and treat everything else as ‘unintegrated residues’ (73). This is because, second, in the social sciences we want to understand social phenomena in their individuality, that is, in their unique configurations that have significance for us.

Values solve a selection problem. They tell us what research questions we ought to address because they inform us about the cultural importance of social phenomena (76):

Only a small portion of existing concrete reality is colored by our value-conditioned interest and it alone is significant to us. It is significant because it reveals relationships which are important to use due to their connection with our values.

It is important to note that Weber did not think that social and natural science were different in kind. Social science too examines the causes of phenomena of interest, and natural science too often seeks to explain natural phenomena in their individual constellations. The role of causal laws is different in the two fields, however. Whereas establishing a causal law is often an end in itself in the natural sciences, in the social sciences laws play an attenuated and accompanying role as mere means to explain cultural phenomena in their uniqueness.

Nevertheless, for Weber social science remained objective in at least two ways. First, once research questions of interest have been settled, answers about the causes of culturally significant phenomena do not depend on the idiosyncrasies of an individual researcher (84; emphasis original):

But it obviously does not follow from this that research in the cultural sciences can only have results which are ‘subjective’ in the sense that they are valid for one person and not for others. [...] For scientific truth is precisely what is valid for all who seek the truth.

The claims of social science can therefore be objective in our third sense (see Section 4). Moreover, by determining that a given phenomenon is ‘culturally significant’ a researcher reflects on whether or not a practice is ‘meaningful’ or ‘important’, and not whether or not it is commendable: ‘Prostitution is a cultural phenomenon just as much as religion or money’ (81). An important implication of this view came to the fore in the so-called ‘*Werturteilsstreit*’ (quarrel concerning value judgments) of the early 1900’s. In this debate, Weber maintained against the ‘socialists of the lectern’ around Gustav Schmoller the position that social scientists qua scientists should not be directly involved in policy debates because it was not the aim of science to examine the appropriateness of ends. Given a policy goal, a social scientist could make recommendations about effective strategies to reach the goal; but social science was to be value-free in the sense of not taking a stance on the desirability of the goals themselves. This leads us directly to the second relevant sense of objectivity.

5.2 Contemporary Rational Choice Theory

Contemporary mainstream economists hold a view concerning objectivity that mirrors Max Weber’s (see above). On the one hand, it is clear that value judgments are at the heart of economic theorizing. ‘Preferences’ are a key concept of rational choice theory, the main theory in contemporary mainstream economics. Preferences are evaluations. If an individual prefers A to B, she *values* A higher than B (Hausman 2012). Thus, to the extent that economists predict and explain market behavior in terms of rational choice theory, they predict and explain market behavior in a way laden with value judgments.

However, economists are not themselves supposed to take a stance about whether or not whatever individuals value is also ‘objectively’ good in a stronger sense (Paternotte 2011: 307-8):

[...] that an agent is rational from [rational choice theory]’s point of view does not mean that the course of action she will choose is objectively optimal. Desires do not have to align with any objective measure of ‘goodness’: I may want to risk swimming in a crocodile-infested lake; I may desire to smoke or drink even though I know it harms me. Optimality is determined by the agent’s desires, not the converse.

In a similar vein, Gul and Pesendorfer write (Gul and Pesendorfer 2008: 8):

However, standard economics has no therapeutic ambition, i.e., it does not try to evaluate or improve the individual's objectives. Economics cannot distinguish between choices that maximize happiness, choices that reflect a sense of duty, or choices that are the response to some impulse. Moreover, standard economics takes no position on the question of which of those objectives the agent should pursue.

According to the standard view, all that rational choice theory demands is that people's preferences are (internally) consistent; it has no business in telling people what they ought to prefer, whether their preferences are consistent with external norms or values. Economics is thus value-laden, but laden with the values of the agents whose behavior it seeks to predict and explain and not with the values of those who seek to predict and explain this behavior.

Whether or not social science, and economics in particular, can be objective in this—Weber's and the contemporary economists'—sense is controversial. On the one hand, there are some reasons to believe that rational choice theory (which is at work not only in economics but also in political science and other social sciences) cannot be applied to empirical phenomena without referring to external norms or values (Sen 1993; Reiss 2013).

On the other hand, it is not clear that economists and other social scientists qua social scientists shouldn't participate in a debate about social goals. For one thing, trying to do welfare analysis in the standard Weberian way tends to obscure rather than to eliminate normative commitments (Putnam and Walsh 2007). Obscuring value judgments can be detrimental to the social scientist as policy advisor because it will hamper rather than promote trust in social science. For another, economists are in a prime position to contribute to ethical debates, for a variety of reasons, and should therefore take this responsibility seriously (Atkinson 2001).

5.3 Evidence-based Medicine and Social Policy

The same demands calling for 'mechanical objectivity' in the natural sciences and quantification in the social and policy sciences in the 19th century and mid-20th century are responsible for a recent movement in biomedical

research, which, even more recently, have swept to contemporary social science and policy. Early proponents of so-called ‘evidence-based medicine’ made their pursuit of a downplay of the ‘human element’ in medicine plain (Guyatt et al. 1992: 2420):

Evidence-based medicine de-emphasizes intuition, unsystematic clinical experience, and pathophysiological rationale as sufficient grounds for clinical decision making and stresses the examination of evidence from clinical research.

To call the new movement ‘evidence-based’ is a misnomer strictly speaking, as intuition, clinical experience and pathophysiological rationale can certainly constitute evidence. But proponents of evidence-based practices have a much narrower concept of evidence in mind: analyses of the results of randomized controlled trials (RCTs). This movement is now very strong in biomedical research, development economics and a number of areas of social science, especially psychology, education and social policy, especially in the English speaking world.

The goal is that of replacing subjective (biased, error-prone, idiosyncratic) judgments by mechanically objective methods. But, as in other areas, attempting to mechanize inquiry can lead to reduced accuracy and utility of the results.

Causal relations in the social and biomedical sciences hold on account of highly complex arrangements of factors and conditions. Whether for instance a substance is toxic depends on details of the metabolic system of the population ingesting it and whether an educational policy is effective on the constellation of factors within which the policy is to be implemented. If an RCT was conducted successfully, the conclusion about the effectiveness of the treatment (or toxicity of a substance) under test is certain for the particular arrangement of factors and conditions of the trial (Cartwright 2007). But unlike the RCT itself, many of whose aspects can be (relatively) mechanically implemented, applying the result to a new setting (recommending a treatment to a patient, for instance) always involves subjective judgments of the kind proponents of evidence-based practices seek to avoid—such as judgments about the similarity of the test to the target or policy population.

On the other hand, RCTs can be regarded as ‘debiasing procedure’ because they prevent researchers from allocating treatments to patients according to their personal interests, so that the healthiest (or smartest or . . .) subjects get the researcher’s favorite therapy. While unbalanced allocations

can certainly happen by chance, randomization still provides some warrant that the allocation was not done *on purpose* with a view to promoting somebody's interests. *A priori*, the experimental procedure is thus more impartial with respect to the interests at stake. It has thus been argued that RCTs in medicine, while no guarantor of the best outcomes, were adopted by the U.S. Food and Drugs Administration (FDA) to different degrees during the 1960s and 1970s in order to regain public trust in its decisions about treatments, which it had lost due to the thalidomide and other scandals (Reiss and Teira 2013; Teira 2010). It is important to notice, however, that randomization is at best effective with respect to one kind of bias, viz. selection bias. Important other epistemic concerns are not addressed by the procedure but should not be ignored (Worrall 2002).

6 Instrumentalism to the Rescue?

We have seen in Sections 2.4 and 4.1 that scientific objectivity and trust in science are closely connected. We want scientific objectivity because and to the extent that we want to be able to trust scientists, their results and recommendations. One possible lesson to draw from the fairly poor success record of the alternative conceptions of scientific objectivity is that these conceptions have the logical order of the ideas mistaken. They look at some privileged feature of science, define this feature as ‘objectivity-making’ and then leave the issue of whether or not the feature also promotes trust to fate. The obvious alternative is to reverse that order, start with what we want and then look for features that might promote the thing in which we are ultimately interested.

If ultimately it is trust in science we want, we might then define as ‘objective’ any feature of science that promotes trust (cf. Fine 1998: 18). That is, anything goes—as long as the practice promotes trust in science. In contraposition to the three traditional alternatives, we may call this conception **instrumentalism about scientific objectivity**.

From an instrumentalist point of view, defining the objective features of scientific inquiry becomes an empirical and contextual issue. It is empirical in that anything that stands in the right kind of causal relation with public trust will count as an objective feature of science. There is no way to tell *a priori* what these features might be. It is contextual in that there is at least the possibility that these features vary with time, place, discipline and other contextual elements. It may well be, for instance, that one or more of the three traditional understandings have once promoted trust in science, even if they no longer do so. It may also be that it is these features that promote trust in one science but not others. And it may be that trust-making features vary with social and political circumstances—different features may be salient in different stages of development or between war and peace times and so on.

The strong point of this way of thinking about objectivity is that none of the threats discussed in the main body of this article—threats posed by the difficulties with the ‘view from nowhere’, with the value-free ideal, and with biases and idiosyncrasies in science—puts scientific objectivity at risk. There is no reason to think that sciences that represent the world from a perspective, in which non-epistemic values play important roles in scientific decision-making and in which personal elements affect outcomes cannot be trusted by the public. Scientific objectivity in the instrumentalist conception

is thus both valuable, as gives us something worth pursuing, *and* attainable.

At the same time, arguably, instrumentalism about objectivity is more of a research programme than an explication of the concept. Suppose we have a domain of science which, at a particular place and time, fares very well with respect to promoting public trust. How are we to tell which features of these scientific practices are responsible for the success? It is obviously impossible to run experiments. Just observing and comparing historical episodes is not likely to clinch results as there will always be numerous differences between any two domains, historical episodes and places.

Moreover, if objectivity is identified with features that promote trust in science, how do we prevent clever marketing from being a crucial feature of scientific objectivity? Or suppose that a science loses public trust (say, as macro and financial economics did after the Financial Crisis after 2007). What might be effective strategies to regain it? Arguably, instrumentalism raises more questions than it answers.

Yet, instrumentalism faces none of the obvious difficulties of the alternative views, so it might well be a worthwhile research project. The challenge for proponents of traditional views of objectivity consists in showing how specific features of their favorite conception of objectivity (e.g., value freedom) secure the epistemic authority of science independent of contextual constraints, and how these features have been used at all times for supporting the credibility of scientific claims. The instrumentalist and realist research programs may therefore fruitfully complement each other.

7 Conclusions

So is scientific objectivity desirable? Is it attainable? That, as we have seen, depends crucially on how the term is understood. We have looked in detail at three different conceptions of scientific objectivity: faithfulness to facts, value-freedom and freedom from personal biases. In each case, there are at least some reasons to believe that either science cannot deliver full objectivity in this sense, or that it would not be a good thing to try to do so, or both. Does this mean we should give up the idea of objectivity in science?

We have shown that it is hard to define scientific objectivity in terms of a view from nowhere and freedom from values and from personal bias. It is a lot harder to say anything positive about the matter. Perhaps it is related to a thorough critical attitude concerning claims and findings, as Popper thought. Perhaps it is the fact that many voices are heard, equally respected and subjected to accepted standards, as Longino defends. Perhaps it is something else altogether. Perhaps it is a combination of several factors, including some that have been discussed in this article.

However, one should not (as yet) throw out the baby with the bathwater. Like those who defend a particular explication of scientific objectivity, the critics struggle to explain what makes science objective, trustworthy and special. For instance, our discussion of the value-free ideal (VFI) revealed that alternatives to the VFI are at least as problematic as the VFI itself, and that the VFI may, with all its inadequacies, still be a useful heuristic for fostering scientific integrity and objectivity. Similarly, although an ‘unbiased’ science may be impossible, there are many mechanisms scientists can adopt for protecting their reasoning against undesirable forms of bias, e.g., choosing an appropriate method of statistical inference.

Whatever it is, it should come as no surprise that finding a positive characterization of what makes science objective is hard. If we knew an answer, we would have done no less than solve the problem of induction (because we would know what procedures or forms of organization are responsible for the success of science). Work on this problem is an ongoing project, and so is the quest for understanding scientific objectivity.

References

- Atkinson, A. 2001, ‘The Strange Disappearance of Welfare Economics’, *Kyklos* 54(2-3): 193–206.
- Bernardo, J.M., 2012, ‘Integrated objective Bayesian estimation and hypothesis testing’, in J.M. Bernardo et al. (eds.): *Bayesian Statistics 9: Proceedings of the Ninth Valencia Meeting*, 1–68. Oxford: Oxford University Press.
- Betz, G., 2013, ‘In defence of the value-free ideal’, *European Journal for the Philosophy of Science* 2: 207–220.
- Biddle, J., 2013, ‘State of the Field: Transient Underdetermination and Values in Science’, *Studies in History and Philosophy of Science* 44: 124–133.
- Biddle, J., and Winsberg, E., 2010, ‘Value Judgements and the Estimation of Uncertainty in Climate Modeling’, in: *New Waves in Philosophy of Science*, P.D. Magnus and J. Busch, (eds.), Basingstoke: Palgrave MacMillan, 172–197.
- Braithwaite, R., 1953, *Scientific Explanation*, Cambridge: Cambridge University Press.
- Carnap, R., 1950/62, *Logical Foundations of Probability*, Second Edition, Chicago: University of Chicago Press.
- Carnap, R., 1952, *The Continuum of Inductive Methods*, Chicago: University of Chicago Press.
- Cartwright, N., 2007, ‘Are RCTs the Gold Standard?’ *BioSocieties* 2(2): 11-20.
- Chang, H., 2004, *Inventing Temperature*, Oxford, Oxford University Press.
- Churchman, C.W., 1948, *Theory of Experimental Inference*, New York: Macmillan.
- Collins, H., 1985, *Changing Order: Replication and Induction in Scientific Practice*, Chicago (IL), University of Chicago Press.

- Collins, H., 1994, ‘A Strong Confirmation of the Experimenters’ Regress’, *Studies in History and Philosophy of Modern Physics* 25(3): 493-503.
- Cranor, C., 1993, *Regulating Toxic Substances: A Philosophy of Science and the Law*. New York: Oxford University Press.
- Crupi, V., Tentori, K., and González, M., 2007, On Bayesian Measures of Evidential Support: Theoretical and Empirical Issues, *Philosophy of Science* 74: 229–252.
- Daston, L., 1992, ‘Objectivity and the Escape from Perspective’, *Social Studies of Science* 22: 597-618.
- Daston, L. and P. Galison, 1992, ‘The Image of Objectivity’, *Representations* 40(Special Issue: Seeing Science): 81-128.
- Daston, L. and P. Galison, 2007, *Objectivity*, Cambridge (MA), MIT Press.
- Dilthey, W. 2002, *Selected Works, Vol. 3: The Formation of the Historical World in the Human Sciences*, R.A. Makkreel and F. Rodi (eds.), Princeton (NJ): Princeton University Press.
- Dorato, M., 2004, ‘Epistemic and Nonepistemic Values in Science’, in: Machamer and Wolters 2004, 52–77.
- Douglas, H., 2000, ‘Inductive Risk and Values in Science’, *Philosophy of Science* 67: 559–579.
- Douglas, H., 2004: The Irreducible Complexity of Objectivity. *Synthese* 138: 453–473.
- Douglas, H., 2009, *Science, Policy, and the Value-Free Ideal*, Pittsburgh: University of Pittsburgh Press.
- Douglas, H., 2011, ‘Facts, Values, and Objectivity’, Jarvie, I., and Zamora Bonilla, J., (eds.), *The SAGE Handbook of Philosophy of Social Science*, 513–529. London: SAGE Publications.
- Douglas, H., 2013, ‘The Value of Cognitive Values’, *Philosophy of Science* 80: 796–806.

- Duhem, P., 1906/1954. *La théorie physique. Son objet et sa structure*, Paris: Chevalier et Riviere; translated by P.P. Wiener, *The Aim and Structure of Physical Theory*, Princeton/NJ: Princeton University Press, 1954.
- Dupré, J., 2007, ‘Fact and Value’, in Kincaid, Dupré and Wylie 2007, 24–71.
- Edwards, W., Lindman, H., and Savage, L.J., 1963, *Bayesian Statistical Inference for Psychological Research*, *Psychological Review* 70: 450–499.
- Elliott, K., 2011, ‘Direct and Indirect Roles for Values in Science’, *Philosophy of Science* 78: 303–324.
- Feyerabend, P., 1962, ‘Explanation, Reduction and Empiricism’, in H. Feigl and G. Maxwell (ed.), *Scientific Explanation, Space, and Time*, (Minnesota Studies in the Philosophy of Science, Volume III), Minneapolis: University of Minneapolis Press, pp. 28–97.
- Feyerabend, P., 1975, *Against Method*, London: Verso.
- Feyerabend, P., 1978, *Science in a Free Society*, London: New Left Books.
- Fisher, R.A., 1935, *The Design of Experiments*, Edinburgh: Oliver and Boyd.
- Fisher, R.A., 1956, *Statistical Methods and Scientific Inference*, New York: Hafner.
- Franklin, A., 1994, ‘How to Avoid the Experimenters’ Regress’, *Studies in the History and Philosophy of Science* 25: 97-121.
- Franklin, A., 1997, ‘Calibration’ *Perspectives on Science* 5: 31-80.
- Gadamer, H.-G. 1989, *Truth and Method*, 2nd edition, trans. by J. Weinsheimer and D.G.Marshall, New York (NY): Crossroad.
- Gaifman, H., Snir, M., 1982. ‘Probabilities Over Rich Languages, Testing and Randomness’, *Journal of Symbolic Logic* 47, 495–548.
- Giere, R., 2006, *Scientific Perspectivism*, Chicago (IL), University of Chicago Press.

- Good, I.J., 1950, *Probability and the Weighing of Evidence*, London: Charles Griffin.
- Gul, F. and W. Pesendorfer, 2008. ‘The Case for Mindless Economics. in *The Foundations of Positive and Normative Economics: a Handbook*, A. Caplin and A. Schotter (ed.), New York, Oxford University Press: 3-39.
- Guyatt, G., J. Cairns, D. Churchill and others, 1992, ‘Evidence-based medicine. A new approach to teaching the practice of medicine’ *JAMA: The Journal of the American Medical Association* 268(17): 2420-2425.
- Hacking, I., 1965, *Logic of Statistical Inference*, Cambridge: Cambridge University Press.
- Hanson, N.R., 1958, *Patterns of Discovery: An Inquiry into the Conceptual Foundations of Science*, Cambridge: Cambridge University Press.
- Haraway, D., 1988, ‘Situated Knowledges: The Science Question in Feminism and the Privilege of Partial Perspective’, *Feminist Studies* 14: 575-599.
- Harding, S., 1993. ‘Rethinking Standpoint Epistemology: What is Strong Objectivity?’ in *Feminist Epistemologies*, L. Alcoff and E. Potter (ed.), New York (NY), Routledge.
- Howson, C., 2000, *Hume’s Problem: Induction and the Justification of Belief*, Oxford: Oxford University Press.
- Howson, C., and Urbach, P., 1993, *Scientific Reasoning: The Bayesian Approach*, Second Edition, La Salle: Open Court.
- Ioannidis, J.P.A., 2005, Why Most Published Research Findings Are False. *PLOS Medicine* 2(8): e124. doi:10.1371/journal.pmed.0020124 (electronic publication).
- Jaynes, E.T., 1968, ‘Prior Probabilities’, *IEEE Transactions on Systems Science and Cybernetics* SSSC4: 227–241.
- Jeffrey, R., 1956, ‘Valuation and acceptance of scientific hypotheses’, *Philosophy of Science* 23: 237–246.

- Jeffreys, H., 1939/98, *Theory of Probability*, Third Edition, Oxford: Oxford University Press.
- Kass, R., and A. Raftery, 1995, ‘Bayes Factors’, *Journal of the American Statistical Association* 90: 773–790.
- Kemeny, J.G., and Oppenheim, P., 1952, ‘Degree of factual support’, *Philosophy of Science* 19: 307–324.
- Kincaid, H., Dupré, J., and Wylie, A., eds., 2007, *Value-Free Science?: Ideals and Illusions*. Oxford University Press.
- Kitcher, P. 2011a, *Science in a Democratic Society*, New York: Prometheus Books.
- Kitcher, P. 2011b, *The Ethical Project*, Cambridge (MA): Harvard University Press.
- Kuhn, T.S., 1959/77, ‘The Essential Tension: Tradition and Innovation in Scientific Research’, in *The Third (1959) University of Utah Research Conference on the Identification of Scientific Talent*, C. Taylor, Salt Lake City: University of Utah Press: 162–74. Reprinted in *The Essential Tension. Selected Studies in Scientific Tradition and Change*, Chicago: University of Chicago Press.
- Kuhn, T.S., 1962/70, *The Structure of Scientific Revolutions*, Second edition, Chicago: University of Chicago Press.
- Kuhn, T.S., 1977, ‘Objectivity, Value Judgment, and Theory Choice’, in: *The Essential Tension. Selected Studies in Scientific Tradition and Change*, Chicago: University of Chicago Press: 320–39.
- Lacey, H., 1999, *Is Science Value-Free?*, London: Routledge.
- Lacey, H., 2002, ‘The ways in which the sciences are and are not value free. In P. Gardenfors, K. Kijania-Placek, and J. Wolenski (eds), *Proceedings of the 11th international congress of logic, methodology and philosophy of science*, Dordrecht: Kluwer, 513–526.
- Lacey, H., 2005, *Values and Objectivity in Science, and Current Controversy about Transgenic Crops*, Lanham, MD: Lexington Books.
- Lakatos, I., 1978, *The Methodology of Scientific Research Programmes*, Cambridge University Press: Cambridge.

- Laudan, L., 1984, *Values in Science*, Berkeley/Los Angeles: University of California Press.
- Laudan, L., 2004, 'The Epistemic, the Cognitive, and the Social', in: Machamer and Wolters 2004, 14–23.
- Levi, I., 1960, 'Must the Scientist Make Value Judgments?', *Journal of Philosophy* 57: 345–357.
- Longino, H., 1990, *Science as Social Knowledge: Values and Objectivity in Scientific Inquiry*. Princeton: Princeton University Press.
- Longino, H., 1996, 'Cognitive and Non-Cognitive Values in Science: Rethinking the Dichotomy', in *Feminism, Science and the Philosophy of Science*, L.H. Nelson and J. Nelson (eds.), Dordrecht: Kluwer, 39–58.
- Lloyd, E., 2005, *The Case of the Female Orgasm: Bias in the Science of Evolution*, Cambridge/MA: Harvard University Press.
- Machamer, P., and G. Wolters (eds.): *Science, Values and Objectivity*. Pittsburgh: Pittsburgh University Press.
- Mayo, D.G., 1996, *Error and the Growth of Experimental Knowledge*, Chicago & London: The University of Chicago Press.
- McMullin, E., 1982, 'Values in Science', *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association 1982*, 3–28.
- Megill, A., 1994, 'Introduction: Four Senses of Objectivity', in: *Rethinking Objectivity*, A. Megill (ed.), Durham/NC: Duke University Press, 1–20.
- Mill, J. S., [1859] 2003, *On Liberty*, New Haven and London, Yale University Press.
- Mitchell, S., 2004, 'The Prescribed and Proscribed Values in Science Policy', in: Machamer and Wolters 2004, 245–255.
- Nagel, T., 1986, *The View From Nowhere*, New York (NY), Oxford University Press.
- Nixon, R., 1969, 'Special Message to the Congress on Social Security'.

- Oakes, M., 1986, *Statistical inference: A commentary for the social and behavioral sciences*. New York: Wiley.
- Oakeshott, M. 1933, *Experience and Its Modes*, Cambridge: Cambridge University Press.
- Okruhlik, K., 1994, 'Gender Bias in the Biological and Social Sciences', *Canadian Journal of Philosophy* 20 (supplement): 21–42.
- Page, S., 2007, *The Difference: How the Power of Diversity Creates Better Groups, Firms, Schools, and Societies*, Princeton/NJ: Princeton University Press.
- Paternotte, C., 2011. 'Rational Choice Theory'. in *The SAGE Handbook of The Philosophy of Social Sciences*, I. C. Jarvie and J. Zamora-Bonilla (ed.), London, SAGE Publications: 307-321.
- Popper, K.R., 2002 [1934], *Logik der Forschung*, Berlin: Akademie Verlag. English translation as *The Logic of Scientific Discovery*. London: Routledge.
- Popper, K.R., 1963, *Conjectures and Refutations: The Growth of Scientific Knowledge*. New York: Harper.
- Popper, K.R., 1972: *Objective Knowledge: An evolutionary approach*, Oxford: Oxford University Press.
- Porter, T., 1995, *Trust in Numbers: The Pursuit of Objectivity in Science and Public Life*, Princeton, Princeton University Press.
- Putnam, H., 2002, *The Collapse of the Fact/Value Dichotomy and Other Essays*, Cambridge (MA): Harvard University Press.
- Putnam, H. and V. Walsh, 2007, 'A Response to Dasgupta' *Economics and Philosophy* 23(03): 359–364.
- Reichenbach, H., 1938, 'On Probability and Induction', *Philosophy of Science* 5: 21–45.
- Reiss, J., 2008, *Error in Economics: The Methodology of Evidence-Based Economics*, London: Routledge.
- Reiss, J. 2010, In Favour of a Millian Proposal to Reform Biomedical Research, *Synthese* 177(3): 427–47.

- Reiss, J., 2013, *Philosophy of Economics: A Contemporary Introduction*, New York (NY), Routledge.
- Reiss, J., and D. Teira, 2013. Causality, Impartiality and Evidence-Based Policy. in *Towards the Methodological Turn in the Philosophy of Science: Mechanism and Causality in Biology and Economics*. H.-K. Chao, S.-T. Chen and R. Millstein (ed.), New York (NY), Springer.
- Resnik, D.B., 2007, *The Price of Truth: How Money Affects the Norms of Science*, Oxford: Oxford University Press.
- Rickert, H. 1929/1986, *The Limits of Concept Formation in Natural Science*, Cambridge University Press. Abridged translation of Rickert, H. 1929, *Die Grenzen der naturwissenschaftlichen Begriffsbildung. Eine logische Einleitung in die historischen Wissenschaften*, 6th edition, Tübingen: Mohr Siebeck.
- Rosnow, R.L., and Rosenthal, R., 1989, Statistical procedures and the justification of knowledge in psychological science. *American Psychologist* 44: 1276–1284.
- Royall, R., 1997, *Scientific Evidence: A Likelihood Paradigm*, London: Chapman & Hall.
- Rudner, R., 1953, ‘The Scientist Qua Scientist Makes Value Judgments’, *Philosophy of Science* 20: 1–6.
- Ruphy, S., 2006, ‘Empiricism all the way down’: A defense of the value neutrality of science in response to Helen Longino’s contextual empiricism. *Perspectives on Science* 14: 189–213.
- Sen, A., 1993, ‘Internal Consistency of Choice’, *Econometrica* 61(3): 495–521.
- Shrader-Frechette, K., 1991, *Risk and Rationality*, Berkeley/Los Angeles: University of California Press.
- Sober, E., 2008, *Evidence and Evolution*, Cambridge: Cambridge University Press.
- Sprenger, J.M., 2009, ‘Evidence and Experimental Design in Sequential Trials’, *Philosophy of Science* 76: 637–649.

- Sprenger, J.M., 2012, ‘The Renegade Subjectivist: Jose Bernardo’s Reference Bayesianism’. *Rationality, Markets and Morality* 3: 1–13.
- Sprenger, J.M., 2014, ‘Bayesianism vs. Frequentism in Statistical Inference’, forthcoming in: A. Hájek and C. Hitchcock (eds): *Oxford Handbook on Philosophy of Probability*. Oxford: Oxford University Press.
- Steele, K., 2012: ‘The Scientist Qua Policy Advisor Makes Value Judgments’, *Philosophy of Science* 79: 893–904.
- Teira, D., 2010, ‘Frequentist versus Bayesian Clinical Trials’ in *Philosophy of Medicine*, F. Gifford (ed.), Amsterdam, Elsevier: 255–297.
- Weber, M., 1904/1949a, ‘ ‘Objectivity’ in Social Science and Social Policy’. in Weber 1904/1949b, 50–112.
- Weber, M., 1904/1949b, *The Methodology of the Social Sciences*, E. Shils and H. Finch (eds.), New York (NY), Free Press.
- Weber, M., 1917/1988: ‘Der Sinn der “Wertfreiheit” der soziologischen und ökonomischen Wissenschaften’, reprinted in: *Gesammelte Aufsätze zur Wissenschaftslehre*, Tübingen: UTB, 451–502.
- Wittgenstein, L., 1953/2001, *Philosophical Investigations*, translated by G. Anscombe, London: Blackwell.
- Wilholt, T., 2009, ‘Bias and Values in Scientific Research’. *Studies in History and Philosophy of Science* 40: 92–101.
- Wilholt, T., 2013, ‘Epistemic Trust in Science’, *British Journal for the Philosophy of Science* 64: 233–253.
- Williams, B., 1985/2011, *Ethics and the Limits of Philosophy*, London and New York (NY), Routledge.
- Williamson, J., 2010, *In Defense of Objective Bayesianism*, Oxford: Oxford University Press.
- Windelband, W., 1915. *Präludien. Aufsätze und Reden zur Philosophie und ihrer Geschichte*, 5th edition. Tübingen: Mohr Siebeck.
- Worrall, J., 2002, ‘What Evidence in Evidence-Based Medicine?’, *Philosophy of Science* 69: S316–S330.

- Wylie, A., 2003. ‘Why Standpoint Matters’, in *Science and Other Cultures: Issues in Philosophies of Science and Technology*, R. Figueroa and S. Harding (ed.), New York (NY) and London, Routledge: 26-48.
- Ziliak, S.T., and D.N. McCloskey, 2008, *The Cult of Statistical Significance: How the Standard Error Costs Us Jobs, Justice and Lives*, Ann Arbor: University of Michigan Press.

List of Figures

- 1 The p-value, represented as the shaded area under the probability density of the tested hypothesis $X \sim N(0, 1)$, for observed data $x = 1.8$ ($p = 0.072$). 41