Methodological and Cognitive Biases in Science: Issues for Current Research and Ways to Counteract Them

Manuela Fernández Pinto
Universidad de los Andes

Arguments discrediting the value-free ideal of science have left us with the question of how to distinguish desirable values from biases that compromise the reliability of research. In this paper, I argue for a characterization of cognitive biases as deviations of thought processes that systematically lead scientists to the wrong conclusions. In particular, cognitive biases could help us understand a crucial issue in science today: how systematic error is introduced in research outcomes, even when research is evaluated as of good quality. To conclude, I suggest that some debiasing mechanisms have great potential for countering implicit methodological biases in science.

1. Introduction

Philosophers of science have traditionally understood problems of biases in science in terms of the negative influence that personal, social, or political interests have for the research process. According to this Baconian view, science ought to be value-free in order to warrant the objectivity of the results. As philosophy of science moves away from this value-free ideal, acknowledging the different ways in which non-epistemic values play an inevitable role in scientific reasoning and practice (Longino 2002; Douglas

The author would like to thank two anonymous reviewers for their careful reading and revision of this paper. Special thanks to Torsten Wilholt and Jamie Shaw for our discussions about biases in science, and their thoughtful comments on earlier drafts. Thanks also to the audiences of the Conferencia Inaugural de Posgrados de Filosofía at Universidad San Buenaventura (Colombia) and the Seminario del Departamento de Lógica, Historia y Filosofía de la Ciencia at UNED (Spain), where earlier versions of the paper were presented. This paper is a product of the research project "A Philosophical Investigation of Biasing Mechanisms in Scientific Research" based at Leibniz Universität Hannover and Universidad de los Andes and funded by the Deutsche Forschungsgemeinschaft, Grant no. 462891071.

Perspectives on Science 2023, vol. 31, no. 5

https://doi.org/10.1162/posc_a_00589

2009), different questions arise regarding biasing mechanisms in scientific inquiry. How should we understand the distinction between the inevitable values involved in scientific practice and the biases that steer scientific research away from its epistemic goals? Which types of bias are involved in scientific inquiry? How do they operate (e.g., are they implicit or explicit)? How do these biases impact the epistemic goals of research? How can such biases be identified and what are potential measures to counteract them?

From a very different perspective, behavioral and cognitive psychologists have shown how the evolution of cognitive mechanisms have helped our species navigate efficiently a sea of information in order to survive. They have also uncovered the ways in which such cognitive mechanisms might be leading us today to constantly arriving at the wrong conclusions. Amos Tversky and Daniel Kahneman (1974) famously argued that humans do not make decisions rationally most of the time, but instead systematically commit the same types of mistakes, due to our cognitive biases. If this is correct, how can we make better decisions when we are "wired" to commit such mistakes? And in the case of science, how are scientists supposed to overcome such biasing mechanisms so that they do not compromise their research results? Are there any successful debiasing techniques that can help us move in the right direction?

In this paper, I argue that in order to identify and manage biases that compromise scientific results it is crucial to pay attention to research from contemporary psychology, which points us to cognitive mechanisms that systematically deviate our decisions, leading us to the wrong conclusions.[1] Acknowledging that cognitive biases affect scientific decision-making, much more than has been previously admitted, is key for understanding some of the problems that confront research today, and finding ways to counteract them. To conclude, I suggest that debiasing mechanisms, such as cognitive forcing tools, have great potential for countering at least some biases in science.

The paper is divided as follows. The second section presents a brief historical overview of the science and values debate leading to the distinction between values and biases. The third section presents a crucial issue with biases in science today: to understand how systematic error is introduced in research outcomes, even when research is evaluated as of good quality. In the fourth section, I introduce cognitive biases as a possible explanation for the problem highlighted in the third section. Finally, the fifth section

1. Notice that not everything we call "bias" falls within the domain of psychology. However, for the purposes of this paper, I will focus only on cognitive biases studied in psychology.

analyses recent research on debiasing techniques and suggests how to implement them in scientific research and education.

2. From "Science and Values" to "Values vs. Biases"

According to the traditional conception of science, science ought to be value-free. In this view, the objectivity of scientific results can only be achieved when scientists leave their values "at the door," for values are seen as corruptive, compromising the epistemic goals of research. For the past thirty years, philosophy of science has moved slowly but steadily away from this value-free ideal, acknowledging the different ways in which non-epistemic values play an inevitable role in scientific reasoning and practice.

A number of philosophers of science have argued against the value-free ideal. One of their strategies has been to challenge the conceptual framework that the ideal presupposes. For instance, some authors have challenged the distinction between epistemic and non-epistemic values (Longino 1995; Solomon 2001; Douglas 2009), others have denied we can make a clear-cut division between the internal and the external aspects of scientific research (Rudner 1953; Anderson 2004; Douglas 2009), and even more radically, some have questioned the fact/value distinction (Dewey 1938; Anderson 2004; Clough 2015).

A more straightforward strategy, not necessarily incompatible with the previous one, has been to show that social and political values (should) play a legitimate role in scientific research, not only before and after the development of scientific research (e.g., when making decision about funding certain lines of research over others, or determining how to apply research products or results), but also during scientific practice as such (e.g., when framing hypotheses, designing experiments collecting and interpreting data, etc.). With respect to this second strategy, two arguments (or lines of argument) have been particularly influential: arguments from the underdetermination of theories by evidence, according to which social and political values are needed to close the gap between theory and evidence that underdetermination leaves open (Nelson 1990; Longino 1990; Kourany 2003) and inductive risk arguments, according to which scientists use social or ethical values to judge the risk of erring when accepting or rejecting a hypothesis (Rudner 1953; Douglas 2009). More recently, a third line of argument has been developed, challenging the lexical priority of evidence, i.e., challenging the privileged epistemic status that arguments from underdetermination and inductive risk give to evidence over values, and adjudicating social values a primary role in scientific practice (Anderson 2004; Kourany 2010; Brown 2013).

Feminist philosophers of science, in particular, have been crucial to the critique of the value-free ideal. From important reformulations of the underdetermination argument (Nelson 1990; Longino 1990; Harding 1998; Kourany 2003) to the more radical challenge of the lexical priority of evidence (Anderson 2004; Kourany 2010), feminist philosophers have shown that many sexist and androcentric values have historically permeated scientific research and that more diverse and feminist values are needed for the improvement of scientific knowledge. Thus, most feminist philosophers of science have defended the importance of identifying appropriate social and political values for scientific research.

Philosophy of science has moved away from the value-free ideal, but new challenges have arisen on the way. If social values have a role to play in science, what exactly is this role? Certainly, they should not be allowed to play any role, for this compromises the empirical adequacy of scientific research. Obviously, research must not privilege certain values over the empirical evidence. As Anderson clearly states:

> Deep down, what the objectors find worrisome about allowing value judgments to guide scientific inquiry is not that they have evaluative content, but that these judgments might be held dogmatically, so as to preclude the recognition of evidence that might undermine them. We need to ensure that value judgments do not operate to drive inquiry to a predetermined conclusion. This is our fundamental criterion for distinguishing legitimate from illegitimate uses of values in science. (Anderson 2004, p. 11)

Accordingly, the acceptance of the value-ladeness of science comes hand-in-hand with new questions about the legitimacy of values in science: which values are legitimate for scientific inquiry?, what roles can values legitimately play in scientific research?, and how should we understand the distinction between the inevitable values involved in scientific practice and the biases that drive scientific research away from its epistemic goals? In this way, the conceptual distinction between values and biases can be useful for better understanding the new challenges that the rejection of the value-free ideal introduces for philosophers of science.

Wilholt (2009) has made an important contribution to this debate showing that biases can compromise research results, and thus be regarded as epistemologically detrimental, even when acknowledging that science is value laden. His analysis of preference bias, i.e., when research results unduly reflect the researcher's preference over other possible results (2009, p. 92), as an epistemic shortcoming will help guide our broader analysis of biases in science. Although Wilholt acknowledges that the concept "bias" is polysemic, being used in different ways both in science and philosophy (2009, p. 92), in general, the concept implies some sort of epistemic shortcoming, more specifically, an introduction of error that deviates the scientific process from legitimate results. Wilholt, for instance, characterizes preference bias as "the infringement of an explicit or implicit conventional standard of the respective research community in order to increase the likelihood of arriving at a preferred result" (2009, p. 99). Other authors have defined bias as "systematic error" (Gluud 2006; Greenland and Lash 2008), "deviations beyond chance" (Ioannidis 2005), or "deviation from the truth beyond random error" (Ioannidis 2017).[2]

Of course, the science and values debate, in which different arguments against the value-free ideal have been provided, has been framed in terms of "values" rather than "biases." After all, the main argument of the critics of the ideal is that values have a role to play in scientific research, and that they can even be epistemically beneficial. Feminist philosophers of science, for example, have made a clear defense of feminist values, as important vehicles for the achievement of scientific knowledge, e.g., through a feminist standpoint (Harding 1986), by diversifying the values of the scientific community (Longino 2002), or by promoting more general democratic values (Kourany 2010). So, we can understand why it was important for the critics of the value-free ideal to shift the conversation from talking about any values as bias, to acknowledging a proper role for values in science.

However, once we move beyond the value-free ideal, questions regarding the epistemically detrimental effects of some values in scientific inquiry remain. For it is still the case that some values, sometimes, have a negative influence in the research process, insofar as they introduce systematic errors, deviating research from its epistemic goals. In this sense, some values (or preferences, or things we privilege), can have a biasing effect in the research process.[3]

Accordingly, one lesson we can take from the science and values debate is that while values can play a legitimate role in scientific inquiry, they can also play an illegitimate role, moving scientists away from their epistemic goals (as feminist philosophers have shown). In such cases, values have a biasing effect in research and should be properly handled so that they don't compromise the production of scientific knowledge. Holman and

2. From a broader characterization of bias in philosophy of science, see Bueter (2022).

3. Notice that in this sense biases can result from the influences of values in scientific research, but also from other, perhaps accidental, causes. Although I am mostly interested in the relation between values and biases in this paper, it is important to keep in mind that biases might also emerge from other sources.

Wilholt call this "the new demarcation problem" (2022). In this sense, philosophers of science who have argued against the value-free ideal can also argue for debiasing mechanisms in scientific inquiry without being inconsistent.

3. Biases in Science

Resnik (2000) provides a preliminary taxonomy of biases in research: distinguishing (i) biases that emerge from human values (e.g., political ideologies or religious beliefs), (ii) psychological prejudices (e.g., anchoring bias), (iii) biases from social, cultural or economic conditions (e.g., financial biases), and (iv) biases from flawed methodological assumptions (e.g., craniometry's assumption that intelligence depends on brain size and shape). While I consider this taxonomy an important first attempt at categorizing biases in science, there is room for improvement. First, the category of psychological biases might be better understood in terms of cognitive biases, as the current literature in social psychology suggests (see, e.g., Mercier and Sperber 2017), to emphasize that they are the result of the evolution of our cognitive capacities, i.e., of how our brains work and not other, broader psychological factors. Second, categories (i) and (iii) actually refer to broad social values that are not clearly distinguished in Resnik's taxonomy, they also individuate biases on the basis of their cause, rather than stipulating where biases inhere (psychology) or what the content of a bias is (iv). And finally, category (iv) needs to be expanded to include not only flawed assumptions, but also flawed methodological decisions more generally.

As previously mentioned, scientists usually understand biases as "deviations beyond chance" (Ioannidis 2005) or "systematic errors" (Greenland and Lash 2008), stemming from choices made during the research process. For the purposes of this paper, I will call these biases proper of the scientific context, *methodological biases*, following Resnik's type (iv) (2000). Examples of methodological biases include confounding bias (distortion of results due to a confounding variable), selection bias (violation of the selection validity conditions), publication bias (selection of more papers with positive outcomes for publication), response bias (tendency to answer untruthfully in surveys), and the like. Scientists have classified, studied, and learned to manage this sort of methodological biases (e.g., Sackett 1979; Vineis 2002; Chavalarias and Ioannidis 2010; Lash et al. 2014). However, other methodological biases, such as biases introduced through non-representative samples or misleading data presentation, are less understood (Bero and Rennie 1996), given that they cannot be identified through quality assessment statistical tools.

As the latest meta-analyses have systematically showed, industry-sponsored studies are significantly more likely to obtain results favoring sponsors than independently funded research (Bekelman et al. 2003; Lexchin et al. 2003; Sismondo 2008; Lundh et al. 2017). Surprisingly, the same meta-analyses have also shown that industry-sponsored studies have lower risk of bias (e.g., of biases being introduced in the process of double-blinding the study), and their methodological quality is at least as good as, sometimes even better than, the quality of independent studies.[4] We know then that financial conflicts of interest do not necessarily lead to scientific fraud, but the precise mechanisms through which biased results enter the scientific process here is not yet clear. Of course, scholars have offered different hypotheses or possible explanations of how this could be happening (inferior comparators, biased coding of events, selective report of favorable outcomes, spined conclusions, publication bias, and so on), and suggested that multiple factors (social, political, economic, etc.) could be playing a role in biasing research results beyond quality assessment tools (Lexchin et al. 2003, p. 1169; Lundh et al. 2013, p. 13). The precise mechanisms at play in a particular case are much harder to prove. However, the fact that scientific fraud is not necessarily related to these biases is also coherent with the results of empirical studies showing that scientists are less likely to commit overt fraud, and that a great number of cases of questionable research practices might not be intentional (Fanelli 2009).

When conducting research, scientists address a number of methodological decisions, emerging from different stages of the research process: establishing the central question that the study aims to answer, designing a study to answer such question, conducting the study, drawing conclusions from the study, and finally publishing the study (Bero and Rennie 1996, p. 211). When designing a clinical study, for example, scientists have to choose the specific patient population for the trial, the specific comparator (e.g., a placebo or an existing available treatment) against which the new treatment will be tried, they also have to determine the dosage for both the control and the treatment groups, and they have to specify an outcome or endpoint for measuring, among others. Different considerations have to be

4. Of course, "research quality" in these studies is assessed according to the available quality assessment tools, which have been designed to measure specific risks of bias (e.g., blinding, drop-out, sample-size), while other risks are left completely unassessed (e.g., comparator choice, outcome reporting, publication bias). So while a study can appear to be high quality and low risk of bias according to the quality assessment tools, it can certainly be suffering from other biasing mechanisms that remain invisible even when checked with the traditional filters.

taken into account in order to make such decisions: time and budget constraints, geographical constraints, laboratory constraints, scientific talent and expertise available, what is the best and most efficient combination of choices to answer the question posed, etc. Despite the multiplicity of options, decisions regarding experimental design have a spectrum of *epistemically legitimate choices*, and methodological biases appear "precisely when making decisions beyond the spectrum of what is epistemically (or methodologically, if you prefer) appropriate, jeopardizing the reliability of the results" (Fernández Pinto 2019, p. 204).

The following example might help illustrate this problem. *Comparator bias* is a type of methodological bias that arises when choosing comparison groups and doses. In particular, comparator bias emerges "when treatments known to be beneficial are withheld from patients participating in controlled trials" (Mann and Djulbegovic 2013, p. 30). Given that new treatments are not compared to the best available therapies, comparator bias leads to suboptimal trial results, and thus represents an epistemic shortcoming.

Comparator bias can appear in many forms. To begin with, new treatments in clinical trials can be compared to a placebo or to an effective available treatment. Placebo-controlled trials are important when trying to determine the efficacy of new treatments, but they are not recommended when effective alternative treatments are already available. Accordingly, the *Helsinki Declaration of Ethical Principles for Medical Research* (1964) states that new treatments should always be tested against the best proven interventions, with only a few exceptions: (i) when no proven treatment is available, or when (ii) "for compelling and scientifically sound methodological reasons the use of any intervention less effective than the best proven one, the use of placebo, or no intervention is necessary to determine the efficacy or safety of an intervention" (WMA 1964).

Despite this, only about half of the drugs approved by the FDA present evidence of comparison with an already existing alternative treatment for market authorization (Goldberg et al. 2011). In addition, recent studies reveal that authors of clinical trials are not aware of relevant systematic reviews and previous trials when designing studies on new treatments (Mann and Djulbegovic 2013, p. 32).

Issues with placebo-controlled trials involve not only ethical issues regarding the fact that control subjects are not given in many cases the best available therapy, but also epistemic issues regarding the fact that trial results won't tell us whether the new treatments are better or not to the ones already on the market. In other words, placebo-controlled trials might tell us whether a treatment is better than nothing, but this is not optimal knowledge when other effective treatments are available: "What we really want to know, what would be significant in terms of advancing current knowledge, is whether the new treatment is better than the best available one" (Fernández Pinto 2019, p. 204).

Comparator bias also arises when choosing the available treatment to compare with the new treatment, as well as when choosing the relevant doses for comparison. The use of suboptimal alternative therapies, i.e., available therapies that have been proven not to be the best on the market, as well as the use of higher or lower doses than the standard doses in the alternative treatments have been tactics successfully used to prove either the efficacy or the benefits of new treatments (Bero and Rennie 1996; Smith 2005; Mann and Djulbegovic 2013). In such cases, comparator bias moves us away from achieving relevant knowledge as well.

However, many times a comparator bias cannot be clearly identified through quality assessment tools, making it possible for a research study to appear of good or even excellent quality even when it has this bias. Since comparators are chosen during the design process, and they aren't commonly justified in the publication process, the decision introducing the comparator bias is very likely to remain hidden from scrutiny. Granting that scientists do not want to bias their results deliberately, and thus leaving cases of overt scientific fraud aside, comparator bias might be unconsciously introduced by scientists who are not even aware of the biasing effects of their comparator choices. This is a disastrous combination epistemically speaking: scientists are likely unaware of their biases and the decision is kept hidden from third-party scrutiny. A more detailed understanding of cognitive biases and their mechanisms in the context of scientific inquiry would perhaps contribute to untangling this problem, as will become clear in the next section.

4. Cognitive Biases in Science

We must acknowledge that science is prone to the same cognitive biases that affect human behavior. The same cognitive system that allows us to understand and explain the world around us, also sets limits to the possibilities of our knowing. Human cognitive capacities have adapted to make optimal decisions under environmental pressures, using different heuristics and biases (Tversky and Kahneman 1974; Greenwald and Banaji 1995; Fazio 2007; Gendler 2008; Payne and Gawronski 2010; Mercier and Sperber 2017). While this seems to work well most of the time—we are after all efficient decision-makers—such mechanisms can easily be misapplied, leading us to unwarranted and sometimes blatantly wrong conclusions (Lilienfeld et al. 2009). Cognitive biases affect scientific research in different ways. For instance, scientists are prone to *asymmetric attention bias*—double-checking unexpected results, while giving a free-pass to expected ones—and to just-so storytelling—giving ungranted post hoc justifications, or "stories," for results of data-analysis (Nuzzo 2015).

According to the traditional view of science, scientists ought to evaluate the evidence supporting or rejecting a hypothesis independently from their previous beliefs. The philosophy of Karl Popper (1963) clearly illustrated this idea in his treatment of the demarcation problem: proper science, contrary to pseudoscience, is falsifiable; proper scientific theories should make risky predictions, i.e., hypothesis contrary to expectations, which should withstand the most rigorous attempts to refute them. Scientists, thus, should in principle aim hard to falsify their hypotheses instead of trying to confirm them. As we have learned from contemporary psychology, however, the human mind works in a very different way. In fact, our previous beliefs greatly influence our appreciation of new beliefs.

In this respect, a well-studied example of a cognitive bias that has a clear influence in scientific research is confirmation bias, also known as expectation bias. Confirmation bias is the tendency to believe or pay attention to evidence that confirms our expectations or beliefs, while ignoring or rejecting evidence that disconfirms or goes against our beliefs or expectations. As a cognitive bias, confirmation bias affects all human reasoning, including scientific reasoning.

Thus, contrary to Popper's view, scientists might be more likely to design and conduct studies that confirm their hypotheses, than to find evidence that disconfirms them.

Explanations for the underlying mechanisms of confirmation bias include the desire to believe, information-processing biases, positive-test strategies, conditional reference frames, and error avoidance (Nickerson 1998). Evidence of the existence of confirmation bias in science comes both from the history of science (Nickerson 1998; Jeng 2006) as well as from empirical studies in different disciplines (Fugelsang et al. 2004; Marsh and Hanlon 2007). A good example of the former is Eddington's expedition to confirm Einstein's prediction that light would be bent by the gravitational field of the sun, a prediction that could be empirically verified by taking photographs of the sun during an eclipse. Accordingly, Eddington embarked on an expedition to West Africa to make the relevant observations of a total solar eclipse to occur on May 29, 1919.

As the official story goes, the evidence collected by Eddington during the eclipse, and later accepted by the Royal Society in London, was key in providing empirical confirmation of Einstein's theory of general relativity and more generally for the acceptance of the new theory worldwide. However, later revisions of the historical record (e.g., Collins and Pinch 1993) have uncovered important measurement errors as well as the discarding of unfitting photographs, in particular the eighteen plates from the Sobral expedition to Brazil, where a second team was sent to register the 1919 eclipse from a different location, without proper justification.

Although the theory has been amply verified during later observations, Eddington's expedition in 1919 can be now considered a case of confirmation bias: he already knew the results he expected to get, and he got there regardless of all the noise in the evidence (Nickerson 1998). As Collins and Pinch note:

> … there was nothing inevitable about the observations themselves until Eddington, the Astronomer Royal, and the rest of the scientific community had finished with their after-the-fact determinations of what the observations were to be taken to be. Quite simply, they had to decide which observations to keep and which to throw out in order that it could be said that the observations had given rise to any numbers at all. (Collins and Pinch 1993, p. 51)

If human decision-making is systematically prone to bias, scientific decision-making is prone to bias as well. Of course, scientists have developed many interesting mechanisms to deal with such bias, such as randomization, double-blinding, and peer-review (Resnik 2014). However, as I have tried to show, there are some instances of decision-making in the research process that are still hidden from any third-party scrutiny and that rely, mistakenly, on the individual scientist's rational capacities. We now know, however, that individual scientists are bad judges of their own biases (Nuzzo 2015), and that they are left in a very vulnerable position when their decisions are left unchecked. In particular, they are prone to introducing biases in research, such as the comparator bias, not because they want to bias their results deliberately, but because they might be unaware of the cognitive biases implicitly guiding their decisions, e.g., a confirmation bias, guiding their choice of comparator. In this way, a systematic error might be introduced in the research process even without the scientist being aware of the problem, just because of the scientist's own cognitive mechanisms.

Acknowledging that individual scientists are prone to cognitive biases, as any other human being, is the first step to understanding how a series of biases might be populating scientific research today, as some meta-analyses suggest (Lundh et al. 2017), but with no deliberately fraudulent behavior involved (Fanelli 2009). Of course, the fact that scientists are mostly unaware of such biases does not mean that these biases are not problematic. They systematically lead to inadequate results, compromising the epistemic goals of science. In this sense, they ought to be identified and managed. Their implicit character just makes it much more difficult to do so.

5. How to Counteract Biases in Science

To start thinking about how to counteract biases in research, it is crucial to acknowledge that biases can be the result of implicit or explicit cognitive attitudes. While implicit attitudes tend to operate automatically and outside our awareness, explicit attitudes are the result of cognitive deliberation and agents are often aware of them (Briñol et al. 2009). Cognitive biases appear as intuitive evolutionary responses most of the time (Croskerry et al. 2013), and thus are mostly implicit and difficult to track. Even though methodological biases can be the result of implicit or explicit attitudes, for the purposes of this paper, I am mostly interested in methodological biases introduced through implicit attitudes. Methodological biases introduced through explicit attitudes, as I mentioned before, are cases of scientific fraud, and must be judged accordingly. Let us focus then on counteracting mechanisms for implicit methodological biases.

I find the influence of implicit biases in science particularly relevant for the purpose of understanding the more likely biasing mechanisms in research, given my assumption that most scientists are good professionals, and that they are unlikely to bias their research projects deliberately (for empirical evidence to support this claim, see Fanelli 2009). Nevertheless, they are they are certainly prone to biases due to automatic cognitive mechanisms, learned social stereotypes, or practice-entrenched methodological decisions.

The central question regarding the influence of implicit biasing mechanisms in research has to do with the possibility of counteracting biases that occur in such an apparently uncontrollable fashion: Is research inevitably prone to implicit bias, or are there effective debiasing techniques that scientists can implement to avoid this problem? Neuroscientists, cognitive psychologists, and social psychologists have been exploring this question in detail in recent decades.

Contrary to previous models (e.g., Rydell et al. 2006), recent research suggests that implicit attitudes have the potential to change through both associative (implicit) and deliberative (explicit) information. In 2009, Briñol and his colleagues conducted a study to measure if rational deliberation can impact automatic evaluations, in the context of faculty hiring. Participants in the study "received a persuasive message in favor of a new policy to integrate more African American professors into the university. This message was composed of either weak or strong arguments in favor of the proposal" (2009, p. 294). By using arguments of different quality, researchers aimed to measure the influence of rational thinking in automatic responses, assuming that differentiating between weak and strong arguments requires more deliberation. In order to measure implicit racial attitudes among participants, researchers used the Implicit Association

Test (IAT).[5] The conclusion of the study states: "we expected and found argument quality to influence automatic evaluations depending on the extent of message processing. That is, under high elaboration conditions, automatic evaluations were found to be more positive toward Blacks for the strong than the weak message" (2009, p. 295). The study suggests then that the use of deliberative information prompting subjects to rational thinking has the potential of neutralizing implicit bias, at least during the timeframe of the experiment.[6]

Other studies have also led to optimistic results regarding the possibility of counteracting implicit biases. Cognitive forcing tools, such as mnemonics (O'Sullivan and Schofield 2019), as well as implementation intentions, practice-based training, and goal priming (Sheeran et al. 2013) have also shown promising effects in modifying implicit bias. O'Sullivan and Schofield (2019), for example, conducted a randomized controlled study in which they gave doctors in the treatment group a cognitive mnemonic tool called "SLOW," with the aim of slowing down for improving diagnostic accuracy. The SLOW tool was basically an acronym for a series of questions related to the diagnostic process: Sure about that?; what is Lacking?; what if the Opposite is true?; Worst case scenario?. Volunteers were given cases to diagnose, and those in the treatment group were asked to use the tool for making the diagnosis. SLOW produced "a subjectively positive impact on doctors' accuracy and thoughtfulness in clinical cases" (2019, p. 1). More generally, Croskerry (2002) has developed a catalog of biases and debiasing tools that have shown some effectiveness.

Even though debiasing mechanisms are costly, after all they require vigilance and reflection of our behavior (Croskerry 2015), they can be effective under the appropriate circumstances (Lilienfeld et al. 2009). In particular, given that the scientific environment is one of strict and rigorous controls,

5. One must notice that IATs have received important critiques in two main areas. First, from the fact that the tests use the velocity of response as a proxy to determining the agent's biases, some have argued that this proxy is not adequate (Mitchell and Tetlock 2006). Second, we have evidence that the tests are not stable in time for the same individual, i.e., factors such as the time of the day, the person's mood, or even whether the person is hungry or not, can influence the test results (Cooley and Payne 2017). Despite these problems, we also have evidence that the IATs are stable at a group level, and even for same age groups within the larger population (Payne et al. 2017).

6. Brownstein (2018, p. 170) has suggested that it might not be the logical or rational force of the argument, but perhaps the positive or negative feelings associated with the evaluation of the applicants, and in this case the bad feelings associated with the possibility of being biased towards African-American candidates, which prompts the unbiased response. In any case, there is an apparently successful debiasing mechanism in place here.

it seems especially well-adapted to implementing debiasing techniques.[7] A promising example in this respect comes from the field of medicine, and more particularly from Intensive Care Units (ICUs), where the implementation of simple check lists has proven extraordinarily successful in reducing human error that traditionally led to a high number of central line infections, cases of untreated pain, and stomach ulcers (Pronovost et al. 2003; Berenholtz et al. 2004; Pronovost et al. 2006). Taking advantage of the high levels of rigor and thoroughness expected from caregivers at ICUs, Dr. Pronovost's simple checklists have made unprecedented improvements in patient care (Gawande 2009). I consider that similar cognitive forcing tools have tremendous potential in the, also rigorous and thorough, scientific research environment. Perhaps even simple tools, such as a 5-point checklist, implemented during the design phase of research studies could prevent the introduction of at least some implicit methodological biases in the research process. The actual design and implementation of such tools, as well as their empirical evaluation is still needed.

6. Conclusions

Arguments discrediting the value-free ideal in science have left us with the question of how to distinguish desirable values from biases that compromise the reliability of research. In this paper, I argued that cognitive biases could help us understand how systematic error is introduced in research outcomes, even when research is evaluated as of good quality. Using comparator bias as an example, I showed how cognitive mechanisms might be behind the introduction of such bias in contemporary clinical studies, and how this possibility becomes crucial for figuring out ways for countering such biases. To conclude, I suggest that debiasing mechanisms, such as cognitive forcing tools, have great potential for countering implicit methodological biases in science.

7. One might argue that the scientific process has already implemented several cognitive forcing tools, and that, for example, the quality assessment tools mentioned earlier are precisely an example of how scientists work to avoid biases in their research process. Strict record keeping in laboratory notebooks could be another example of such forcing mechanisms. While I agree in general with this argument, developments in cognitive psychology and debiasing mechanisms more specifically can serve us to further develop cognitive forcing tools for the research process, especially for those biasing mechanisms that we don't handle properly yet, such as the methodological biases I have presented in this paper. I thank one anonymous reviewer for pointing this out.

References

Anderson, Elizabeth. 2004. "Uses of Value Judgments in Science: A General Argument, with Lessons from a Case Study of Feminist Research on Divorce." *Hypatia* 19: 1–24. https://doi.org/10.1111/j.1527-2001.2004.tb01266.x

Bekelman, Justin E., Yan Li, and Cary P. Gross. 2003. "Scope and Impact of Financial Conflicts of Interest in Biomedical Research: A Systematic Review." *JAMA - Journal of the American Medical Association* 289(4): 454–465. https://doi.org/10.1001/jama.289.4.454, PubMed: 12533125

Berenholtz, Sean Michael, Peter Pronovost, Pam A. Lipsett, Deborah Hobson, Karen Earsing, Jason E. Farley, Shelley Milanovich, Elizabeth Garrett-Mayer, Bradford D. Winters, Haya R. Rubin, Todd Dorman, and Trish M. Perl. 2004. "Eliminating Catheter-Related Bloodstream Infections in the Intensive Care Unit." *Critical Care Medicine* 32(10): 2014–2020. https://doi.org/10.1097/01.CCM.0000142399.70913.2F, PubMed: 15483409

Bero, Lisa, and Drummond Rennie. 1996. "Influences on the Quality of Published Drug Studies." *International Journal of Technology Assessment in Health Care* 12(2): 209–237. https://doi.org/10.1017/S0266462300009582, PubMed: 8707496

Briñol, Pablo, Richard E. Petty, and Michael J. McCaslin. 2009. "Changing Attitudes on Implicit Versus Explicit Measures. What Is the Different?" Pp. 287–326 in *Attitudes: Insight from the New Implicit Measures*. Edited by Richard E. Petty and Russell H. Fazio. New York: Psychology Press.

Brown, Matthew. 2013. "Values in Science beyond Underdetermination and Inductive Risk." *Philosophy of Science* 80: 829–839. https://doi.org/10.1086/673720

Brownstein, Michael. 2018. *The Implicit Mind: Cognitive Architecture, the Self, and Ethics*. Oxford: Oxford University Press. https://doi.org/10.1093/oso/9780190633721.001.0001

Bueter, Anke. 2022. "Bias as an Epistemic Notion." *Studies in History and Philosophy of Science* 91: 307–315. https://doi.org/10.1016/j.shpsa.2021.12.002, PubMed: 34955230

Chavalarias, David, and John P. A. Ioannidis. 2010. "Science Mapping Analysis Characterizes 235 Biases in Biomedical Research." *Journal of Clinical Epidemiology* 63(11): 1205–1215. https://doi.org/10.1016/j.jclinepi.2009.12.011, PubMed: 20400265

Clough, Sharyn. 2015. "Fact/Value Holism, Feminist Philosophy, and Nazi Cancer Research." *Feminist Philosophy Quarterly* 1(1): 1–12. https://doi.org/10.5206/fpq/2015.1.7

Collins, Harry, and Trevor Pinch. 1993. *The Golem: What You Should Know about Science*. Cambridge: Cambridge University Press.

Cooley, Erin, and B. Keith Payne. 2017. "Using Groups to Measure Intergroup Prejudice." *Personality and Social Psychology Bulletin* 43: 46–59. https://doi.org/10.1177/0146167216675331, PubMed: 28903648

Croskerry, Pat. 2002. "Achieving Quality in Clinical Decision Making: Cognitive Strategies and Detection of Bias." *Academic Emergency Medicine* 9(11): 1184–1204. https://doi.org/10.1197/aemj.9.11.1184, PubMed: 12414468

Croskerry, Pat. 2015. "When I Say … Cognitive Debiasing." *Medical Education* 49(7): 656–657. https://doi.org/10.1111/medu.12670, PubMed: 26077213

Croskerry, Pat, Geeta Singhal, and Sílvia Mamede. 2013. "Cognitive Debiasing 2: Impediments to and Strategies for Change." *BMJ Quality and Safety* 22(Suppl. 2): ii65–ii72. https://doi.org/10.1136/bmjqs-2012-001713, PubMed: 23996094

Dewey, John. 1938. *Logic: The Theory of Inquiry*. New York: Holt, Rinehart and Winston.

Douglas, Heather E. 2009. *Science, Policy, and the Value-Free Ideal*. Pittsburgh, PA: University of Pittsburgh Press. https://doi.org/10.2307/j.ctt6wrc78

Fanelli, Daniele. 2009. "How Many Scientists Fabricate and Falsify Research? A Systematic Review and Meta-Analysis of Survey Data." *PLoS ONE* 4(5): e5738. https://doi.org/10.1371/journal.pone.0005738, PubMed: 19478950

Fazio, Russell H. 2007. "Attitudes as Object-Evaluation Associations of Varying Strength." *Social Cognition* 25: 603–637. https://doi.org/10.1521/soco.2007.25.5.603, PubMed: 19424447

Fernández Pinto, Manuela. 2019. "Scientific Ignorance: Probing the Limits of Scientific Research and Knowledge Production." *THEORIA. An International Journal for Theory, History and Foundations of Science* 34(2): 195–211. https://doi.org/10.1387/theoria.19329

Fugelsang, Jonathan A., Courtney B. Stein, Adam E. Green, and Kevin N. Dunbar. 2004. "Theory and Data Interactions of the Scientific Mind: Evidence from the Molecular and the Cognitive Laboratory." *Canadian Journal of Experimental Psychology* 58(2): 86–95. https://doi.org/10.1037/h0085799, PubMed: 15285598

Gawande, Atul. 2009. *The Checklist Manifesto*. London: Profile Books.

Gendler, Tamar. 2008. "Alief in Action (and Reaction)." *Mind and Language* 23: 552–585. https://doi.org/10.1111/j.1468-0017.2008.00352.x

Gluud, Lise Lotte. 2006. "Bias in Clinical Intervention Research." *American Journal of Epidemiology* 163(6): 493–501. https://doi.org/10.1093/aje/kwj069, PubMed: 16443796
Goldberg, Nikolas H., Sebastian Schneeweiss, Mary K. Kowal, and Joshua J. Gagne. 2011. "Availability of Comparative Efficacy Data at the Time of Drug Approval in the United States." *JAMA - Journal of the American Medical Association* 305(17): 1786–1789. https://doi.org/10.1001/jama.2011.539, PubMed: 21540422
Greenland, Sander, and Timothy Lash. 2008. "Bias Analysis." Pp. 345–380 in *Modern Epidemiology*, 3rd ed. Edited by Kenneth J. Rothman, Sander Greenland, and Timothy L. Lash. Philadelphia, PA: Lippincott Williams & Wilkins.
Greenwald, Anthony G., and Mahzarin R. Banaji. 1995. "Implicit Social Cognition: Attitudes, Self-esteem, and Stereotypes." *Psychological Review* 102: 4–27. https://doi.org/10.1037/0033-295X.102.1.4, PubMed: 7878162
Harding, Sandra. 1986. *The Science Question in Feminism*. Cornell, NY: Cornell University Press.
Harding, Sandra. 1998. *Is Science Multicultural? Postcolonialisms, Feminisms, and Epistemologies*. Bloomington and Indianapolis: Indiana University Press.
Holman, Bennet, and Torsten Wilholt. 2022. "The New Demarcation Problem." *Studies in History and Philosophy of Science* 91: 211–220. https://doi.org/10.1016/j.shpsa.2021.11.011, PubMed: 34968804
Ioannidis, John P. A. 2005. "Why Most Published Research Findings are False." *PLoS Medicine* 2: e124. https://doi.org/10.1371/journal.pmed.0020124, PubMed: 16060722
Ioannidis, John P. A. 2017. "Statistical Biases in Science Communication: What We Know about Them and How They Can Be Addressed." Pp. 103–110 in *The Oxford Handbook of the Science of Science Communication*. Edited by Kathleen Hall Jamieson, Dan M. Kahan, and Dietram Scheufele. Oxford: Oxford University Press. https://doi.org/10.1093/oxfordhb/9780190497620.013.11
Jeng, Monwhea. 2006. "A Selected History of Expectation Bias in Physics." *American Journal of Physics* 74(7): 578–583. https://doi.org/10.1119/1.2186333
Kourany, Janet A. 2003. "A Philosophy of Science for the Twenty-First Century." *Philosophy of Science* 70: 1–14. https://doi.org/10.1086/367864
Kourany, Janet A. 2010. *Philosophy of Science After Feminism*. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199732623.001.0001
Lash, Timothy L., Matthew P. Fox, Richard F. Maclehose, George Maldonado, Lawrence C. McCandless, and Sander Greenland. 2014. "Good Practices for Quantitative Bias Analysis." *International Journal*

of Epidemiology 43(6): 1969–1985. https://doi.org/10.1093/ije/dyu149, PubMed: 25080530

Lexchin, Joel, Lisa Bero, Benjamin Djulbegovic, and Otavio Clark. 2003. "Pharmaceutical Industry Sponsorship and Research Outcome and Quality: Systematic Review." *British Medical Journal* 326: 1167–1170. https://doi.org/10.1136/bmj.326.7400.1167, PubMed: 12775614

Lilienfeld, Scott O., Rachel Ammirati, and Kristin Landfield. 2009. "Giving Debiasing Away: Can Psychological Research on Correcting Cognitive Errors Promote Human Welfare." *Perspectives on Psychological Science* 4(4): 390–398. https://doi.org/10.1111/j.1745-6924.2009.01144.x, PubMed: 26158987

Longino, Helen. 1990. *Science as Social Knowledge: Values and Objectivity in Scientific Inquiry*. Princeton, NJ: Princeton University Press. https://doi.org/10.1515/9780691209753

Longino, Helen. 1995. "Gender, Politics, and the Theoretical Virtues." *Synthese* 104(3): 383–397. https://doi.org/10.1007/BF01064506

Longino, Helen. 2002. *The Fate of Knowledge*. Princeton, NJ: Princeton University Press. https://doi.org/10.1515/9780691187013

Lundh, Andreas, Joel Lexchin, Barbara Mintzes, Jeppe B. Schroll, and Lisa Bero. 2017. "Industry Sponsorship and Research Outcome." *Cochrane Database of Systematic Reviews* 2(2): MR000033. https://doi.org/10.1002/14651858.MR000033.pub3, PubMed: 28207928

Lundh, Andreas, Sergio Sismondo, Joel Lexchin, Octavian A. Busuioc, and Lisa Bero. 2013. "Industry Sponsorship and Research Outcome." *Cochrane Database of Systematic Reviews* 12. https://doi.org/10.1002/14651858.MR000033.pub2

Mann, Howard, and Benjamin Djulbegovic. 2013. "Comparator Bias: Why Comparisons Must Address Genuine Uncertainties." *Journal of the Royal Society of Medicine, Supplement* 106(1): 30–33. https://doi.org/10.1177/0141076812474779, PubMed: 23358277

Marsh, David M., and Teresa J. Hanlon. 2007. "Seeing What We Want to See: Confirmation Bias in Animal Behavior Research." *Ethology* 113(11): 1089–1098. https://doi.org/10.1111/j.1439-0310.2007.01406.x

Mercier, Hugo, and Dan Sperber. 2017. *The Enigma of Reason*. Cambridge, Mass.: Harvard University Press. https://doi.org/10.4159/9780674977860

Mitchell, Gregory, and Philip E. Tetlock. 2006. "Antidiscrimination Law and the Perils of Mindreading." *Ohio State Law Journal* 67: 1023–1121.

Nelson, Lynn H. 1990. *Who Knows: From Quine to Feminist Empiricism*. Philadelphia: Temple University Press.

Nickerson, Raymond S. 1998. "Confirmation Bias: A Ubiquitous Phenomenon in Many Guises." *Review of General Psychology* 2(2): 175–220. https://doi.org/10.1037/1089-2680.2.2.175

Nuzzo, Regina. 2015. "How Scientists Fool Themselves—And How They Can Stop." *Nature* 526: 182–185. https://doi.org/10.1038/526182a, PubMed: 26450039

O'Sullivan, Eoin D., and Susie J. Schofield. 2019. "A Cognitive Forcing Tool to Mitigate Cognitive Bias—A Randomised Control Trial." *BMC Medical Education* 19(1): 1–8. https://doi.org/10.1186/s12909-018-1444-3, PubMed: 30621679

Payne, B. Keith, Heidi A. Vuletich, and Kristjen B. Lundberg. 2017. "The Bias of Crowds: How Implicit Bias Bridges Personal and Systemic Prejudice." *Psychological Inquiry* 28: 233–248. https://doi.org/10.1080/1047840X.2017.1335568

Payne, B. Keith, and Bertram Gawronski. 2010. "A History of Implicit Social Cognition. Where is it Coming From? Where is it Now? Where is it Going?" Pp. 1–18 in *Handbook of Implicit Social Cognition*. Edited by Bertram Gawronski and B. Keith Payne. New York: The Guilford Press.

Popper, Karl. 1963. *Conjectures and Refutations: The Growth of Scientific Knowledge*. London: Routledge and Kegan Paul.

Pronovost, Peter, Sean Berenholtz, Todd Dorman, Pam A. Lipsett, Terri Simmonds, and Carol Haraden. 2003. "Improving Communication in the ICU Using Daily Goals." *Journal of Critical Care* 18(2): 71–75. https://doi.org/10.1053/jcrc.2003.50008, PubMed: 12800116

Pronovost, Peter, Dale Needham, Sean Berenholtz, David Sinopoli, Haitao Chu, Sara Cosgrove, Bryan Sexton, Robert Hyzy, Robert Welsh, Gary Roth, Joseph Bander, John Kepros, and Christine Goeschel. 2006. "An Intervention to Decrease Catheter-Related Bloodstream Infections in the ICU." *The New England Journal of Medicine* 355(26): 2725–2732. https://doi.org/10.1056/NEJMoa061115, PubMed: 17192537

Resnik, David B. 2000. "Financial Interest and Research Bias." *Perspectives on Science* 8(3): 255–285. https://doi.org/10.1162/106361400750340497

Resnik, David B. 2014. "Bias, Misconduct, and Integrity in Scientific Research." Pp. 162–172 in *The Routledge Companion to Bioethics*. Edited by John Arras, Elizabeth Fenton, and Quill Kukla. New York: Routledge.

Rudner, Richard. 1953. "The Scientist *Qua* Scientist Makes Value Judgments." *Philosophy of Science* 20: 1–6. https://doi.org/10.1086/287231

Rydell, Robert J., Allen R. McConnell, Diane M. Mackie, and Laura M. Strain. 2006. "Of Two Minds: Forming and Changing Valence-Inconsistent Implicit and Explicit Attitudes." *Psychological Science* 17(11): 954–958. https://doi.org/10.1111/j.1467-9280.2006.01811.x, PubMed: 17176426

Sackett, David L. 1979. "Bias in Analytic Research." *Journal of Chronic Diseases* 32(1–2): 51–63. https://doi.org/10.1016/0021-9681(79)90012-2, PubMed: 447779

Sheeran, Paschal, Peter M. Gollwitzer, and John A. Bargh. 2013. "Nonconscious Processes and Health." *Health Psychology* 32(5): 460–473. https://doi.org/10.1037/a0029203, PubMed: 22888816

Sismondo, Sergio. 2008. "How Pharmaceutical Industry Funding Affects Trial Outcomes: Causal Structures and Responses." *Social Science and Medicine* 66(9): 1909–1914. https://doi.org/10.1016/j.socscimed.2008.01.010, PubMed: 18299169

Smith, Richard. 2005. "Medical Journals Are an Extension of the Marketing Arm of Pharmaceutical Companies." *PLoS Medicine* 2(5): e138. https://doi.org/10.1371/journal.pmed.0020138, PubMed: 15916457

Solomon, Miriam. 2001. *Social Empiricism*. Cambridge, Mass.: MIT Press. https://doi.org/10.7551/mitpress/6296.001.0001

Tversky, Amos, and Daniel Kahneman. 1974. "Judgement Under Uncertainty: Heuristics and Biases." *Science* 185(4157): 1124–1131. https://doi.org/10.1126/science.185.4157.1124, PubMed: 17835457

Vineis, Paolo. 2002. "History of Bias." *Sozial- und Praventivmedizin* 47: 156–161. https://doi.org/10.1007/BF01591887, PubMed: 12238297

Wilholt, Torsten. 2009. "Bias and Values in Scientific Research." *Studies in History and Philosophy of Science Part A* 40(1): 92–101. https://doi.org/10.1016/j.shpsa.2008.12.005

World Medical Association. 1964. "WMA Declaration of Helsinki—Ethical Principles for Medical Research Involving Human Subjects." https://www.wma.net/policies-post/wma-declaration-of-helsinki-ethical-principles-for-medical-research-involving-human-subjects/