Values, bias and replicability

Michał Sikorski[1]

Abstract

The Value-free ideal of science (VFI) is a view that claims that scientists should not use non-epistemic values when they are justifying their hypotheses, and is widely considered to be obsolete in the philosophy of science. I will defend the ideal by demonstrating that acceptance of non-epistemic values, prohibited by VFI, necessitates legitimizing certain problematic scientific practices. Such practices, including biased methodological decisions or Questionable Research Practices (QRP), significantly contribute to the Replication Crisis. I will argue that the realizability of VFI is not a necessary condition for its validity. Then, I will show how some of the prominent proposals of value-laden science legitimize problematic scientific practices, provide real-world examples, and generalize the argument. Finally, I will show how value-laden methodological decisions contribute to the Replicability Crisis and discuss two strategies for realizing VFI.

1 Introduction

The value-free ideal of science (VFI) is a view that claims that scientists should not use non-epistemic values when they are justifying their hypotheses. The view has enjoyed varying degrees of popularity. It was popular in the high days of neopositivism. Then it became controversial. It was both criticized (e.g., by Rudner, 1953) and defended (e.g., by Jeffrey, 1956). In the last thirty years, it was generally regarded as obsolete, for example, by Longino (1990), Douglas (2009), Elliott (2011), or Tsui (2016).[1] Recently, researchers started, once again, to look at it more favorably, for example, Koertge (2004), Hudson (2016), Betz (2013), John (2015a), Lacey (2018), Djørup et al. (2019), or Menon and Stegenga (2023). My paper fits into the second group. I will defend the VFI by showing that if we accept the use of non-epistemic values prohibited

---

[1] See Holman and Wilholt (2021) or Hicks (2014) for discussion.

✉ Michał Sikorski
  michalpsikorski@gmail.com

[1] Center for Philosophy, Science, and Policy, Marche Polytechnic University, Ancona, Italy

by it, we are forced to accept as legitimate scientific conduct some of the disturbing phenomena of present-day science like founder bias or questionable research practices. Then, I will show that value-laden science contributes to the replication crisis.

In the Sect. 2, I will discuss the VFI in more detail. In the Sect. 3, I will present three views that were proposed as alternatives to the VFI: Douglas' value-laden science, Ludwig's proposal concerning ontological choices, and the proposal developed by Janet Kourany. In the Sect. 4, I will present problems caused by the rejection of the VFI. Then I will show how value-laden science contributes to the replication crisis in the Sect. 5. Finally in Sect. 6, I will discuss two strategies (from Betz, 2013 and Levi, 1960) for realizing the VFI.

2 The value-free ideal: motivation and controversy

Generally speaking, the VFI is a view that claims that the non-epistemic values of the scientists should not influence reasoning in science. Epistemic values are typically understood as values that promote the attainment of truths; all other values are non-epistemic (e.g., Steel, 2010). Examples of non-epistemic values include fairness and equality, whereas examples of epistemic values encompass internal consistency and parsimony. In various scenarios, values classified as non-epistemic may indeed promote the truth. For instance, in some cases, hypotheses that promote equality are more likely to be true when compared to naive pre-scientific stereotypes. However, there are also cases where equality may not promote the truth. In contrast, the connection between epistemic values and truth is more systematic.[2] For instance, a hypothesis cannot be true without being internally consistent. Similarly, the connection between other epistemic values and truth is equally systematic, though not necessarily equally strong. For example, it has been argued that parsimonious hypotheses are more likely to be true (e.g., Sober, 2015). If it is so, parsimony is an epistemic value. Consequently, I will define epistemic values as those that *systematically* promote the attainment of truths, while non-epistemic values are accidentally connected to truth or not connected at all.[3] For the purposes of this paper, I will assume the validity of this distinction.

The second assumption of my paper will be that we can distinguish between stages of discovery and justification in a scientific process. The stage of discovery, also called theory construction, includes all tasks a scientist performs to develop and present a hypothesis (or theory). The stage of justification begins when the hypothesis or theory is ready, and scientists proceed to test it empirically. This stage includes designing and conducting studies and analyzing and interpreting data. For the purposes of this paper, we can consider the stage of justification to conclude when collected data are analyzed and interpreted, and responsible scientists have assessed whether the results support the tested hypothesis.

The initial formulation of VFI is too strong; some of the uses of these values in science are not problematic. For example, there is nothing wrong with scientists choosing topics for their studies based on their interests. I will defend the VFI restricted

---

[2] I am grateful to an anonymous reviewer for directing me toward including the requirement of systematicity.

[3] A similar definition was presented in Steel (2017).

to the context of justification. It claims that scientists should not use non-epistemic values when they are justifying their claims.[4]

It is important to remember that the VFI is an ideal. A few sometimes overlooked consequences follow from this. First of all, no descriptive claim follows from the VFI. It may be the case that no one complies with a given norm, but it is still a valid norm. Secondly, it seems that even if it is not possible to satisfy an ideal, this does not make it invalid. It may still be the case that by approaching an unattainable ideal, we will become better off. This is consistent with how the role of ideals is typically understood in the literature, e.g., "Ideals, on the other hand, should articulate aspirational goals. They should help us define what we want to be, if we had all the time and resources in the world." (Douglas, 2014 p. 9). Additionally, it seems that some ideals widely believed to be valid are unrealizable. For example, impartiality is typically understood in the following way:

> The word "impartial" connotes absence of bias, actual or perceived. (UNODC, 2019, p. 28)

Each member of the jury, the whole jury, and the justice system overall are required to be impartial. At the same time, in light of the psychological evidence showing the prevalence of biases, it seems that there was never and never will be a perfectly impartial jury or a perfectly impartial judicial system. This does not make the ideal of impartiality any less valid or attractive, and possible improvements in impartiality any less beneficial (see e.g., Gobert, 1988 or Cammack, 1994).

This issue is connected to the discussion of the *Ought imply Can* principle. The principle requires that an obligation or ideal needs to be possible to satisfy in order to be valid. The principle was popular among philosophers (see e.g., Feldman, 1986; Zimmerman, 1996, or Vranas, 2007) and used as a premise in many arguments (see e.g., Graham, 2011). On the other hand, more recently, there has been a shift away from this principle, with counterexamples being proposed against it (see, for example, Graham, 2011; Mizrahi, 2009, or Hughes, 2018). The standard counterexamples involve cases of promises, for example:

(1) I promised that I will pick up my spouse from the airport. On my way there, my car broke down, and because of that, it was not possible for me to be there on time. It makes sense for me to say in such a situation: "I ought to be at the airport in 5 min." even if that is no longer possible.[5]

Those and similar arguments presented in different contexts (see e.g., Mizrahi, Mizrahi (2012) for arguments in an epistemic context) and growing empirical evidence showing that the principle is unintuitive (see e.g., Buckwalter & Turri, 2015; Chituc et al., 2016, or Kissinger-Knox et al., 2018) strongly suggest that the principle is false and obligations or ideals do not have to be realizable to be valid.[6]

As we have already seen, the VFI is controversial. Arguments both for and against it were presented. Following Betz (2013), we can divide arguments against it into two

---

[4] Some authors claim that the use of non-epistemic values is harmful even in the context of discovery (e.g., Okruhlik, 1994).

[5] This example is based on examples from Hughes (2018).

[6] A different argument for the same conclusion was presented in Menon and Stegenga (2023).

groups. The semantic arguments show that the VFI is not a sound proposal. For example, Dupré (2007) argues that the distinction between epistemic and non-epistemic values on which the VFI is based is not well-defined. The methodological arguments show that the VFI is unrealizable. Elliott (2011) distinguishes two groups of such arguments: error and gap arguments.[7] According to the gap arguments, no evidence can ever fully support a general hypothesis. Therefore, a scientist needs to use background assumptions that fill this gap. When it comes to practically relevant hypotheses, the non-epistemic values of scientists inevitably play a role in selecting which of the background assumptions are applicable in a given case. Therefore, scientists need to use their non-epistemic values when they accept their hypotheses. Versions of gap arguments were presented in Longino (1990), Howard (2006), and Kourany (2003). Error arguments start from the realization that when a scientist draws a conclusion based on the collected evidence, she risks committing an error. She may accept a false hypothesis (type I error) or reject a true one (type II error). If the tested hypothesis is practically relevant, both kinds of errors have practical consequences. Therefore, while deciding on what standard of evidence to use in a given case (which translates to a specific trade-off of both kinds of errors), the scientist has to consider the practical consequences of error. Her non-epistemic values play a crucial role in such considerations. Versions of this argument were proposed in Churchman and Acirc (1948), Rudner (1953), and Douglas (2009). Finally, some methodological arguments fit neither of the groups. For example, Ludwig (2015) argues that scientists have to use non-epistemic values to make ontological choices on which their hypotheses are based.

The methodological arguments rely on the assumption that an ideal has to be realizable to be valid. As I argued above, this assumption seems to be problematic; at the very least, it should be explicitly defended. As far as I know, it was not. The second assumption is that non-epistemic values are necessary to fulfill the role identified in the arguments, be it to fill a gap between the evidence and theory or justify the chosen standard of evidence and ontological choices. This assumption was undermined, for example, in Levi (1960), Betz (2013), or John (2015a). If the arguments present in those papers are sound, the VFI may be realizable after all. I will discuss this in the Sect. 6.

Authors who criticize the VFI usually propose alternative ideals of scientific conduct. Not surprisingly, the core of all these proposals is a claim that scientists have to or even should use non-epistemic values when they are justifying their hypotheses. In the next section, I will sketch three of such proposals.

3 Value-laden science

In this section, I will present three counterproposals to the VFI.

Douglas (2009) presents the value-laden ideal of science. She claims that the influence of both epistemic and non-epistemic values should be restricted. Values can be used in two ways. Firstly, a scientist can use them in a direct way when they serve as reasons and evidence. Secondly, she can use them in an indirect way. According to

---

[7] For the discussion concerning the connection between two types of arguments see: ChoGlueck (2018).

Douglas, values can be used in a direct way just in the early phases of the scientific process. For example, scientists can decide based on their values which problem they will work on. During justification, the part of the scientific process interesting for our purposes, scientists should not use values in a direct way. Values should not be treated as reasons. On the other hand, in opposition to the VFI, she claims that values, including non-epistemic ones, should play an indirect role in the context of justification. What exactly is meant by the *indirect role*? Scientists have to be responsible for the theories they accept. Therefore, according to Douglas, to make the final judgment concerning the acceptance of a given hypothesis, a scientist has to take into consideration the non-epistemic consequences of her possible mistake. The essential part of such considerations is played by non-epistemic values. They are necessary to assess how costly is a possible error and therefore what amount of evidence is necessary to accept the hypothesis in question. It is important for our purposes to add that, according to Douglas, making the final decision to accept a hypothesis is not the only part of the justification where scientists should use non-epistemic values:

> Choices regarding which empirical claims to make arise at several points in scientific study. From selecting standards for statistical significance in one's methodology to choices in the characterization of evidence during a study to the interpretation of that evidence at the end of the study to the decision of whether to accept or reject a theory based on the evidence, a scientist decides which empirical claims to make about the world. (Douglas, 2009, p. 103)

As far as I understand, she claims that they should use values (including non-epistemic ones) to make all decisions that are not determined by the evidence.

The role of values discussed in Ludwig (2015) is consistent with Douglas' theory. He claims that scientists have to use them when they make ontological choices. For example, choices concerning the meanings of theoretical terms used by their theories, such as species or intelligence. According to the author, such choices cannot be avoided or made without the use of non-epistemic values, and therefore scientists have to use them. Again, it seems to be less of an argument against the VFI than against its realizability.

Both Douglas' and Ludwig's proposals define the acceptable use of values in terms of their function in the scientific process. Other proposals look to define a set of values that can be safely used in scientific practice. An example of such a theory is the proposal developed by Janet Kourany. The most recent version of the theory was proposed in Kourany (2020). Kourany claims that scientific effort is supported by society because science contributes to human flourishing. This determines which values should be used in science: "values that support the kind of research that promotes human flourishing are the right ones, values that don't are not." (see Kourany, 2020, p. 11).

4 Problematic consequences of value laden science

In the next two sections, I will build an argument in favor of VFI. In this section, I aim to demonstrate that rejecting the VFI legitimizes unacceptable scientific practices. To achieve this, I will first present a counterexample in the form of problematic scientific

practices for each of the theories presented in the second section. Following this, I will briefly discuss other theories of value-laden science and outline the general conditions in which the influences of non-epistemic values become problematic. In the subsequent section, I show how the influences on methodological choices contribute, through those unacceptable practices, to the replicability crisis.

Consider scientists who test a hypothesis, for example, a null hypothesis that some new medical product has no negative side effects. The result of the experiment will impact the society in which scientists live. If it is a (false) negative, policymakers are likely to ban the product. A lucrative industry and many job opportunities will not be created. On the other hand, if it is a (false) positive, the potentially harmful product will be used in clinical practice. According to Douglas, a scientist should consider the negative impact of both false-negative and false-positive results, and based on that adjust the amount of evidence necessary for accepting the hypothesis. Let us assume that two scientists conduct two experiments intending to test the hypothesis. Both of them choose an established method and, luckily, they get the results that perfectly match the distribution of the property in question in the population. At the same time, the results are not conclusive, the revealed correlation is not very strong. Both scientists are aware of the positive and negative effects of the adoption of the drug. Because of different personal histories, they believe that different potential effects are the most important. The first of the scientists was raised in a small town with high unemployment and saw the new product mainly as a great opportunity to reduce unemployment in poorer regions of the country. Consequently, she sets a high threshold of evidence in order to lower the chance of false positives. She concludes that the new product is not dangerous. The second scientist has a history of serious diseases in her family and therefore she thinks that her priority should be to shield the population against possible harm caused by the new product. Therefore, she sets a low standard for rejecting the null hypothesis and concludes, on the basis of the collected evidence, that the product is dangerous.

According to Douglas' proposal, both scientists perform perfectly well. Neither of them used non-epistemic values in a direct way. They have used acceptable non-epistemic values and consequently based their methodological decisions on reasonable assessments of possible consequences of errors. Consequently, we have no reason to suppose that neither did anything wrong. At the same time, they draw different conclusions from the same data, and it seems that something went wrong. The final conclusions were not determined by data but by the personal histories of both scientists. The example shows that even an indirect influence of non-epistemic values can be a decisive factor. Nothing hangs on the fact that the scientists are using the non-epistemic values to fix the sufficient level of evidence rather than to make any other methodological choice.

Douglas does not specify to which degree one can lower or raise a sufficient level of evidence. She just points out that values should not weigh stronger than the evidence. If we can raise the required level of evidence arbitrarily high or low, it is hard to see how that can be a consequence of her view rather than an additional postulate.

The problem is not just theoretical. Time and again, scientists used their non-epistemic values in the way permitted by Douglas to deliver results that fit their non-epistemic interests. An example of that is the founder bias (see e.g., Stelfox, 1998),

which consists of the fact that empirical studies are significantly more likely to support a result preferred by researchers or their employers. Biased studies are not frauds; scientists do not fabricate their results or use other clearly unacceptable strategies. At the same time, there seems to be something wrong with these experiments. To see that, let us consider an example from Wilholt (2008). Industry-funded studies that tested the toxicity of a chemical substance called Bisphenol A tend to use a special strain of rats that are less susceptible to the substance. Consequently, none of the industry-funded experiments showed a carcinogenic effect of the substance in opposition to 90% of government-funded studies. The non-epistemic values were not used in a direct way. On the other hand, it seems that the scientists used their non-epistemic values, perhaps self-interest while making methodological choices concerning which strain of rats to use. According to Douglas' story, the industry-funded scientists did nothing wrong, assuming that the crucial methodological decision was consistent with her values. On the other hand, it is now widely recognized in the methodology of medical science that such influences of industry on scientific results are prevalent (see, for example, Oostrom, 2023). It is also recognized that the preference of the founder likely influences the results through the interest and preference of the scientists as reconstructed above (see e.g., Angell 2004; Friedberg et al., 1999; Krimsky, 2013; or May, 2020). Finally, it is widely believed that those biasing influences and resulting unreliable results are responsible for enormous human suffering and waste of resources (see e.g., Abramson & Starfield, 2005; Angell, 2004; or Purdy et al., 2017 for literature review).

It is similar to the case of Ludwig's proposal. As we remember, he claims that scientists should make their ontological choices on the basis of their non-epistemic values. At the same time, the author rightly claims that the results of experiments depend on these choices. Therefore, the results of the experiments depend on non-epistemic values. The way the non-epistemic values influence the results (through ontological choices or through the adjustment of the required level of evidence) does not make much of a difference.

Once again, the problem is not just theoretical. An example from Oreskes and Conway (2010) shows how tampering with a notion of causality can misrepresent a scientific result. I will assume, for the purpose of the example, that a choice concerning the conceptualization of causality is an ontological choice.[8] Causality is not explicitly mentioned by Ludwig among ontological notions but it seems to play a similar role in a scientific experiment:

> When asked if a three-pack-a-day habit might be a contributory factor to the lung cancer of someone who'd smoked for twenty years, Cline again answered no, you "could not say [that] with certainty. . . I can envision many scenarios where it [smoking] had nothing to do with it." When asked if he was paid for the research he did on behalf of the tobacco industry, he acknowledged that the tobacco industry had supplied $ 300,000 per year over ten years - $3 million but it wasn't "pay," it was a "gift." (Oreskes & Conway, 2010, p. 31)

---

[8] See also Lexchin (2006).

The specialist in question, Martin I. Cline, seems to be using an implausibly demanding version of the regularity theory of causation to present scientific results misleadingly. The theory he implicitly endorses claims that a necessary condition for z to be a contributory factor of y is that in all (possible?) cases in which z is present, y is. That would explain his reluctance to admit that a "three-pack-a-day habit" is a possible contributory cause of lung cancer. This theory of causality is too demanding. The questioned claim is true, in light of scientifically informed common sense, as soon as we accept any less demanding theory of causality. The quote is taken from a transcript of a trial in which Cline served as an expert. It is hard to imagine a more responsible function for a scientist. Given that he received money from a tobacco company, a strong case can be made that the misleading ontological choice was caused by one of his non-epistemic values, once again self-interest.

Another example of an ontological choice being problematic because of the influence of non-epistemic values was described in Bishop (1990). Dorothy Bishop discusses studies investigating an association between handedness and developmental disorders. The field is filled with inconsistent results:

> ...developmental dyslexia has been linked to mixed hand preference (Harris, 1957), strong left-handedness (Geschwipd & Behan, 1982), inconsistent right-handedness (Schachter et al., 1987) and both left- and strong right-handedness (Apnett & Kilsbaw, 1984). There are also many studies that report no association (see review by Bishop, 1983). (Bishop, 1990, p. 812)

She shows through simulation that if a scientist is free to choose the way she conceptualizes handedness after she already obtains the experimental data, she is almost guaranteed to get a positive result. According to the author, the fact that scientists make their choices motivated by a desire to find a significant result at least partially explains the inconsistent results.

In the case of Kourany's proposal, finding a similar problematic case is a bit harder. Acceptable values in this proposal are determined by the society that supports a given scientific practice. All scientists within the society should use only non-epistemic values deemed conducive to the flourishing of that community and therefore, problematic influences of values of individual scientists are excluded. However, even at the societal level, the influence of non-epistemic values can be problematic. For instance, Asatryan (2020) presents empirical results strongly suggesting that the nationality of scientists has a biasing influence on estimations of fiscal multipliers, comparable to the influence of founder preferences. An extreme illustration of the influence of societally determined values can be found in Julian Huxley's early account of Soviet genetic research (Huxley, 1949). Huxley describes how political and ideological preferences compromised Soviet genetic research, ultimately leading to the emergence and short-lived dominance of the discipline of Lysenkoism. To what extent the values that influenced reasoning in these cases supported research promoting human flourishing remains unclear. Consequently, it is uncertain whether these cases constitute counterexamples to Kourany's proposal. Nevertheless, these examples demonstrate that the influence of societal values can be just as corrupting as the influence of the values of founders or individual scientists.

In all the cases described above, problems arise from the influence of scientists' non-epistemic values on their methodological choices. I am not suggesting that all applications of non-epistemic values lead to false or misleading results. However, any influence of non-epistemic values introduces another potential source of error. Furthermore, in the Sect. 5, I will argue that if different scientists use different values, such influences can lead to inconsistent results.

In light of the cases of problematic influence of non-epistemic values, it is now widely acknowledged that a proposal of a value-laden ideal needs to be able to exclude problematic influences of non-epistemic values (see e.g., Anderson, 2004; Hicks, 2014, or Holman & Wilholt, 2021). Literature devoted to the role of values in science is full of such proposals, many of them are subtle and ingenious. Because of that, discussing in detail how susceptible each of them is to my argument goes beyond the scope of the paper. On the other hand, I would like to present two arguments suggesting that it may be unlikely that any of such proposals will succeed in excluding all of the problematic cases.

In my first point, I will rely on the work of Holman and Wilholt (2021). In their article, the authors collected some of the demarcation proposals and categorized them into four groups depending on the strategy used to demarcate between problematic and acceptable uses of non-epistemic values. Axiological proposals focus on the kind of non-epistemic values whose uses are not problematic. Already discussed theory from Kourany (2020) is an example of such theory. Another prominent proposal from this group was the ideal of well-ordered science developed by Philip Kitcher (see e.g., Kitcher, 2011). The proposal claims that the correct answers to the methodological value-laden decisions, such as those central to the gap and error arguments, are determined by a decision reached by well-informed disputants in an ideal conversational situation in which all perspectives were respected. The functionalist proposals focus on the function played by the non-epistemic values in question. The theory, presented in Anderson (2004), claims that an influence of non-epistemic values is acceptable as long as they do not "operate to drive inquiry to a predetermined conclusion." Another notable theory of this kind is the proposal developed by Heather Douglas which I discussed in the second section. Consequential strategies highlight the consequences of the methodological choices influenced by given non-epistemic values. For example, a theory presented in Intemann (2015) claims that "it is legitimate for scientists to appeal to non-epistemic values insofar as doing so will promote democratically endorsed epistemological and social aims of research" (p. 218). The last two groups of proposals explicitly focus on social rather than an individual perspective. Coordinate strategies utilize the interplay between scientific practice and the expectations of the general public. For example, a proposal from Wilholt (2008) claimed that scientific conventions (both implicit and explicit) regulate the way scientists can use their non-epistemic values while making methodological decisions. Finally, critical-contextual theories rely on social institutions to counterbalance the possible negative consequences of scientists using their non-epistemic values. The most prominent of such proposals is contextual empiricism developed by Helen Longino (see e.g., Longino, 1990). Longino's proposal can be interpreted as claiming that a given influence of value is acceptable as long as a given experiment is embedded in a broader epistemic

context that has venues for criticism, is responsive to this criticism, has a standard of discussion, and promotes equality of intellectual authority.

How do those theories square with the presented problematic cases? Anderson's theory will clearly rule out cases of biased experiments; the methodological decisions in those experiments were made in a way that makes a favored outcome far more likely. On the other hand, the proposal will not exclude the practices described in Bishop (1990) and similar ones. In those cases, the decisions targeted a positive result rather than any specific result. The situation is likely similar in the case of other proposals that put the responsibility to decide if a given use of non-epistemic values is legitimate on individual scientists. Plausibly, different scientists consider different results to promote "democratically endorsed epistemological and social aims of research." If so, the proposal from Intemann (2015) will not rule out all problematic cases. In light of that, it seems that the proposals that relegate the responsibility of demarcation to the communal level are more promising (e.g., proposals from Kitcher, 2011 or Wilholt, 2008). On the other hand, the results of Asatryan (2020) suggested that the inconsistency in results on the level of communities exists and may be equally problematic.

There is a more general reason to expect that neither of the demarcation proposals will be successful. If the only feature common to all of the problematic cases is that they involve non-epistemic values motivating the crucial methodological decisions, a value-laden proposal that would be successful in excluding all of the problematic cases would have to exclude all the uses of non-epistemic values and therefore transform it into a VFI-like proposal. Is it plausible that there are no features present only in cases of problematic use of non-epistemic values? To answer this question, we have to first introduce researchers' degrees of freedom (see e.g., Simmons et al., 2011 or Wicherts et al., 2016). During designing and conducting a scientific experiment, a scientist has to make a lot of decisions concerning the exact shape of the experiment. For example, she has to decide: How to collect the data? When to stop collecting the data? Should some observations be excluded? How to understand theoretical terms? Each of these issues can be approached in many equally correct and some incorrect ways, and this constitutes researchers' degrees of freedom.

Researchers' degrees of freedom can be misused in many different ways. First of all, a decision may be biased, that is, it can be made with a specific result in mind as in the case described by Wilholt (2008). Biases come in many different flavors (see e.g., MacCoun, 1998 for a detailed discussion). Some of them are conscious, others unconscious; some are caused by the preferences of the scientist (as the one we discussed), and others by their prior beliefs (confirmation bias). Secondly, after the experiment has been performed, a scientist can consider which combination of possible choices will likely generate a positive result and report just these choices. This and similar practices are called questionable research practices (see e.g., Simmons et al., 2011). Other examples of questionable research practices are: using underpowered statistical designs or optional stopping, which I will discuss in the next section. These phenomena (questionable research practices and biases) are not considered to belong to one homogeneous group and were not analyzed as such.

In all of the mentioned cases, non-epistemic values motivate crucial decisions. However, these values vary, potentially encompassing material self-interests, national values, or the value attached to a successful publication. Any other type of non-

epistemic value can motivate a problematic decision. Even the most noble and uncontroversial non-epistemic values do not inherently promote truth. Consequently, they can similarly motivate misuses of researchers' degrees of freedom and steer a study toward a false result. An interesting example of noble values compromising scientific reliability was uncovered in the recent trial of Elizabeth Holmes. In 2022, Holmes was found guilty of defrauding investors of her company *Theranos* (see USAO NDCA, 2022). The company was developing a revolutionary method of blood testing requiring only a small amount of blood to run a wide variety of tests. During the trial, it was revealed that many of the claims concerning the reliability of the method used by Holmes to attract investors were not warranted. It was also demonstrated that those claims were supported by the results of unreliable experiments. Responsible scientists were urged by Holmes to treat data points that went against the preferred result (high diagnosticity and reliability of developed blood tests) as outliers and disregard them (Randazzo, 2021). On the other hand, the majority of jurors remain convinced that Holmes's main motivation for her questionable and illegal actions was noble. They were convinced that Holmes genuinely believed that the methods and medical devices developed by her company had the potential to improve the reliability and availability of medical diagnosis (Cohen, 2022). Holmes's noble and indirectly used non-epistemic values did not make the practices she commissioned any less problematic.

In light of the above examples, we can attempt to sketch the general conditions in which influences of non-epistemic values will be problematic in a way analogous to the discussed problematic methodological practices. Let us assume that we do not know if the tested hypothesis is true or not. In such a situation, every methodological decision that has the potential to change the outcome of the study and is not motivated purely by empirical values is problematic. Such a decision will steer the study toward results in line with the underlying non-epistemic value and therefore is a potential source of bias (see e.g., May, 2020 or Hudson, 2022). Decisions motivated by epistemic value, by definition, orient the results of the study toward the truth. This is not guaranteed in the case of any other values. As often pointed out by proponents of value-laden science, it may be the case that a non-epistemic value will accidentally skew the scientific study toward the right result. Does this validate the uses of such values? No, it does not. Questionable research practices are not legitimized by rare cases in which they lead to true results. They were shown to, on average, inflate the rate of false positive results and therefore are not considered to be acceptable scientific strategies. It is the same in the case influences of non-epistemic values.

This idea can also be expressed in terms of logical dependence. A scientific result that relies solely on collected data and epistemic values, which promote truth, may have a high chance of being true. When dependence on other factors is introduced, it introduces the risk of compromising the result. In a given case, it is uncertain whether these additional factors are guiding the result toward truth. It is known that, in general, they do not promote truth. This way of articulating the conclusion clearly highlights that the effectiveness of reforms aimed at reducing the influence of the non-epistemic values does not hinge on the ultimate reliability of the VFI. Each potential source of bias that is eliminated improves the reliability of the procedure in question and therefore enhances the reliability of the results, regardless of whether all such sources can ultimately be eliminated or not.

5 Value-laden science and the replication crisis

In this section, I will show how the influence of non-epistemic values on methodological decisions in the presence of conflicting values contributes to the replication crisis. I do not intend to claim that it is the sole or primary cause of the crisis, but I will argue that it constitutes a significant and plausibly sufficient contributing factor.

Somebody may still think that there is nothing wrong with using non-epistemic values, for example, in the way proposed by Douglas (2009). She may think that there is nothing wrong with two scientists reaching different conclusions from the same observation because of their different sympathies or interests. I do not think that this response is plausible or, in other words, that the bullet is worth biting. To see that, let us consider the *Replication Crisis* (for discussion see: Open Science Collaboration, 2015 or Romero, 2017). The crisis consists of the fact that the results of many scientific experiments are not replicable. This means that an experiment with a similar or identical design conducted by different scientists (or even the second time by the same researchers) delivers different results. The exact percentage of replicable studies is unknown. Some approximation is provided by Open Science Collaboration (2015). The authors attempted to replicate the results of one hundred experiments from papers published in prestigious psychological journals. Less than half of the attempts were successful.[9] This is a disappointing result. In light of this, it is hard to have any confidence in the truth of a psychological hypothesis based on a single experiment. The crisis is generally perceived as something very disturbing, as the name itself suggests (see e.g., Pashler & Wagenmakers, 2012; Anwari & Daniël, 2019 or Simine, 2018). Many philosophers and methodologists explore different ways to deal with it (e.g., Ioannidis, 2005). As we have seen, a similar attitude toward inconsistent scientific results is present in the literature devoted to funding bias in biomedical research. A recent study, Oostrom (2023) compared the results of randomized control trials (RCT) funded by the drug manufacturer with the results of other RCTs testing the same drug. The study found significant differences in the results; experiments financed by the manufacturer were more likely to deliver results demonstrating the effectiveness of the drug. The presence of these effects in pharmacology and resulting inconsistencies are widely believed to be problematic by scientists.[10]

At the same time, according to Douglas' proposal, such situations are expected and natural. If different scientists share different non-epistemic values, they make different methodological decisions. As a result of that, they sometimes, or maybe even often, draw different conclusions from the same data. If this is the case, the fact that the scientists responsible for the original study have different non-epistemic values than the scientists responsible for its replication may be a reason for the failure of the replication attempt. To see that, let us go back to our example (p. 8). Suppose the first scientist tests the hypothesis concerning the harmfulness of *x*, and the second one

---

[9] "Thirty-six percent of replications had significant results; 47% of original effect sizes were in the 95% confidence interval of the replication effect size; 39% of effects were subjectively rated to have replicated the original result..." (Open Science Collaboration, 2015, p. 943)

[10] It has also been frequently noted that funding can influence a scientist's interests or non-epistemic values, which, in turn, can impact her methodological decisions and final results (see e.g., Friedberg, 1999 or Angell, 2004).

tries to replicate the result, given the inconclusiveness of the data. In that case, the replication will fail no matter how similar they will be. Given the differences in the statistical design, the second experiment is an attempt to conceptually replicate the result. The conceptual replication, unlike the direct replication, does not impose using an identical statistical design in both experiments (see e.g., LeBel et al., 2018). In the case of a direct replication, a scientist who attempts to directly replicate an experiment cannot use her non-epistemic values to adjust the sufficient level of evidence; she has to follow the design of the original study. Even in the case of direct replication, there are some decisions that scientists can use to tamper with the outcome of the experiment. For example, it is typically not required for replication to use the same population as the original study. As we have seen, these choices can be used to change the results of the experiment.

Once again, the examples from Wilholt (2008) and Bishop (1990), as well as the results of Friedberg et al. (1999) or Oostrom (2023), show that this way of influencing the results is present in science. In the case of industry-biased studies, it is unlikely that the result of a biased study will be successfully replicated by an unbiased experiment. Similarly, in the case of studies of handedness, it is not highly probable that the result of any of the experiments will be repeated in a high-quality replication in light of their conflicting results and the prevalence of questionable research practices.

The connection between the use of non-epistemic values influencing methodological decisions and the replication crisis is acknowledged in methodological literature. For example, consider two factors named as the causes of questionable research practices by Simmons et al. (2011):

> (a) ambiguity in how best to make these decisions and (b) the researcher's desire to find a statistically significant result. (Simmons et al., 2011, p. 1359)

The first cause is the existence of researchers' degrees of freedom; the second is the fact that scientists value statistically significant results. Both causes are connected by an implicit assumption that these non-epistemic values motivate methodological choices. The use of questionable research practices greatly inflates the chance of a false-positive result and therefore lowers the replicability of a given study (see Simmons et al., 2011). Similar factors are predictors of false-positive results according to Ioannidis (2005):

> Corollary 4: The greater the flexibility in designs, definitions, outcomes, and analytical modes in a scientific field, the less likely the research findings are to be true. (Ioannidis, 2005 p. 698)

> Corollary 5: The greater the financial and other interests and prejudices in a scientific field, the less likely the research findings are to be true. (Ioannidis, 2005, p. 698)

Interestingly, some of the questionable research practices can be interpreted as a way of "deciding what amount of evidence is necessary," as recommended by Douglas and other defenders of value-laden science. A perfect example is *optional stopping* (see e.g., Rianne & Grünwald, 2017 or Montori et al., 2005). A scientist using this procedure, instead of deciding beforehand how many subjects she will use in her experiment, decides during the experiment. She either stops and concludes the experiment

after some subjects are tested or adds additional subjects. Unsurprisingly, scientists seem to be inclined to conclude their experiments when the correlation they hoped for is present. Similarly to other questionable research practices, the procedure increases the rate of false-positive errors. Another decision a scientist has to make while designing her experiment is how statistically powerful it will be. Statistical power is the likelihood that an experiment will detect the effect if it exists. If the effect size is small, a large number of participants is required to get a moderately powerful experiment. Because of that, scientists are sometimes inclined to conduct many cheaper, less powerful studies instead of one more powerful study with the hope to get a significant result in at least one of them by luck. The low statistical power of an experiment is also responsible for its low replicability (e.g., Vazire, 2016 or Ioannidis, 2005).

It is not clear what exactly proponents of value-laden science have in mind when they write about "deciding what amount of evidence is necessary." On the other hand, given that neither of the above procedures involves using non-epistemic values in a direct way and any non-epistemic value can motivate it, it seems that they are acceptable ways of using such values. At the same time, both contribute to the replication crisis.

According to Douglas, both scientists in our example performed perfectly well. Similarly, some cases of optional stopping or underpowered experiments are consistent with her theory. Situations described in other counterexamples from Sect. 4 have similar effects on replicability. In light of that, it seems clear that the discussed proposals failed to rule out problematic uses of non-epistemic values. If scientists are permitted to incorporate their non-epistemic values during the process of justification, these values may introduce bias into the results of their studies and, when conflicting values are at play, potentially contribute to the replication crisis. Similar concerns over plausible effects of non-epistemic values on replicability were raised by Hudson (2020). He argued that the value-ladenness of scientific practice contributes to the crisis by causing publication bias, another phenomenon commonly recognized as a factor contributing to the crisis (see e.g., Scargle, 2000). This argument together with the concerns raised above suggests that there may be many ways in which the influence of non-epistemic values compromises the replicability of scientific results. In light of the crucial role played by replicability according to many scientists or methodologists, for instance:

> Reproducibility is a defining feature of science, but the extent to which it characterizes current research is unknown. Scientific claims should not gain credence because of the status or authority of their originator but by the replicability of their supporting evidence. (OSC, 2015, p. 943)

the very idea of value-laden science seems to be problematic.

Before we proceed, it's essential to address the limitations of the replicability objection. Replicability is not compromised if all scientists adhere to the same set of non-epistemic values. In such a scenario, the influence of values would guide the results of experiments in a consistent direction, thus not jeopardizing the replicability of the findings. Therefore, if a value-laden proposal identifies a unique set of non-epistemic values that scientists can legitimately employ and excludes all other values, the constrained influence will not compromise reliability. For instance,

in the case of Kourany's proposal, all scientists within a given society should utilize the same set of values, preventing any differences in results due to influences of values. It is similar in the case of theories that rely on an ethical assessment of scientists to distinguish legitimate from illegitimate values (e.g., Douglas' proposal, see Schroeder, 2020 for discussion of ethical and political approaches) if we assume a version of moral realism, which implies a single correct solution for each situation that necessitates the use of non-epistemic values. Such a combination would imply that at least one of the scientists from the example discussed earlier in the fourth section made a mistake in her assessment of values. If both had been correct, they would have used the same values and reached the same conclusion. However, it is worth noting that the biasing influence of uniform non-epistemic values will be more problematic than the influence of the non-epistemic values of individual scientists. When the same set of potentially biasing factors is imposed on all scientists, the resulting bias becomes more prevalent. Moreover, such a systematic bias may be challenging to detect due to its universality. Replication plays a crucial role in scientific self-correction (see e.g., Sikorski & Andreoletti, 2023) and failed replications can help uncover biasing assumptions. If the problematic assumption is widespread among scientists, the chances of it being detected through replication are slim.

Steps toward implementing the VFI promote replicability by reducing the influence of non-epistemic values on methodological choices, which is a significant factor contributing to the irreplicability of results. For instance, experiments biased by the preferences of the founder often produce results inconsistent with unbiased studies. Furthermore, the manner in which non-epistemic values influence results does not differ significantly, as different scientists may reach divergent conclusions due to these influences. Thus, each step toward achieving VFI diminishes the impact of such factors and consequently enhances the replicability of results. Even if VFI is ultimately unattainable, eliminating potential causes of irreplicability improves scientific practice. Although other factors contribute to the replicability crisis, addressing the influence of non-epistemic values must be part of the solution, as argued herein and by Hudson (2020). In the subsequent section, I will outline two strategies for realizing VFI, which translate into specific recommendations aimed at promoting replicability.

6 Realizability of VFI

In light of the ways in which non-epistemic values can compromise scientific practice, it is worth reconsidering the realizability of VFI.

As we have seen, error and gap arguments convinced many philosophers that non-epistemic values are indispensable in scientific practice, and therefore the VFI is unattainable. Even if we grant that from the fact that an ideal is impossible to realize it follows that it is not a valid ideal, which is, as I argued, not obvious, at least two strategies for demonstrating its realizability were proposed.

First, Betz (2013) describes a general strategy for hedging scientific results in cases where some of the assumptions or methods are not standard or well-supported. For example, if the experimental design used is not standard, researchers can present their results in the form of a conditional statement with the assumption in the antecedent and

the original non-hedged results in the consequent. This strategy can be applied in cases of other methodological decisions which, according to methodological arguments, require the use of non-epistemic values. For instance, both scientists from our example could present their results in the following form: "Given the used threshold of evidence, the substance x is harmful/harmless". Betz presents examples of this strategy from actual scientific practice. Surprisingly, even some philosophers who oppose the VFI seem to favor a similar solution. As noted by Betz (2013), Douglas (2009) proposes a similar strategy. Similarly, Rudner (1953) points out that it is essential for a scientist to make their value decision explicit, and one way to do so is to follow Betz's strategy.

Betz's strategy was criticized by John (2015b) and ChoGlueck (2021). The strategy employed in both papers seems to be similar, as they do not argue that hedging cannot be used to avoid the use of non-epistemic values[11] but rather that it is in some way unattractive. John reanalyzes an original case study used by Betz, the work of the Intergovernmental Panel on Climate Change, and on this basis argues that a scientist, in order to avoid the use of VFI, has to not only hedge the degree to which the collected evidence support the tested hypothesis but also should hedge other methodological choices made for the purpose of the study, for example, the chosen criterion for acceptable evidence. The report hedged in this way would be, according to John, too complicated to be useful for a policymaker which is too costly a price to pay for the value-freedom. Similarly, ChoGlueck argues, using a case study of a "drug effect" of morning-after pill Plan B, that expressions used to hedge hypotheses like "may" are sometimes value-laden and therefore Betz's strategy can be in some cases unsuccessful. Both authors have a point here, it seems true that in some cases the hedging strategy, despite demonstrating that the VFI is realizable, may be impractical. Other such cases may be already discussed ontological choices. It seems to be impractical to list the results of all the ontological choices done on behalf of an experiment. Moreover, some assumptions may be just too standard or too widely shared to be listed, or listing them may make the results of the study too hard to interpret to be useful.

In such cases, a strategy of adopting conventions instead of making value-laden choices may be superior. The strategy was proposed by Levi (1960). I will present it in the context of biased science discussed in Wilholt (2008). As we have already seen, industry-funded scientists use a special strain of rats while examining the toxicity of Bisphenol A. Those rats are less susceptible to the substance, and therefore the choice makes the experiment far less likely to show its harmful effect. In analyzing the founder bias, Witholt does not blame what seems to be the main cause of the problem, namely non-epistemic values of the scientist affecting their methodological choices. Instead, he points out that in all cases of founder bias, scientists failed to comply with some important scientific conventions. For example, in order to prevent the use of insensitive strains of animals, the scientific community formed the following convention:

---

[11] See: "Proponents of the methodological critique who claim that value-freedom is impossible are, as Betz suggests, mistaken. However, there is a difference between showing that some state-of-affairs is possible and showing that it is ideal. The discussion above shows that achieving value-freedom for science comes at a significant non-epistemic cost." (John, 2015b, p. 9) and "I do not take my case study as evidence that values are ineliminable from all drug facts, all regulatory science, or all scientific knowledge. While I am sympathetic to such conclusions, I am not arguing here for the generalizability of this case study and my two-fold account of value-ladenness." (ChoGlueck, 2021, p. 69).

> Because of clear species and strain differences in sensitivity, animal model selection should be based on responsiveness to endocrine active agents of concern (i.e. responsive to positive controls), not on convenience and familiarity (National Toxicology Program, 2001, p. vii)

This analysis does not seem to be satisfactory. What about instances of similar practices that occurred before the norm was established? They inspired the norm but cannot be explained as cases of biased science due to the lack of compliance with a norm that had not yet been established. Wilholt claims that in such cases, the norms may be implicit. While this defends his analysis against obvious counterexamples, it also makes it much more speculative. Implicit norms can always be postulated. A more satisfying analysis should consider the behavior of scientists that these norms were meant to prevent in the first place.

Given the described cases, it seems plausible that these and similar norms aim to prevent methodological decisions motivated by non-epistemic values that can alter the outcome of a given experiment. This suggests the following conceptualization of founder bias:

> A study is founder-biased iff during its course:
>
> (a) a methodological decision(s), influenced by consideration concerning which outcome of the study would be preferable to the founder, was made;
> (b) this decision(s) influenced the outcome of the study.

It seems to be a more straightforward and satisfying analysis of founder bias (see also May, 2020).

At the same time, the process during which biased studies inspire new scientific norms is interesting for our purposes. In light of all the above, it seems that by constituting new conventions, the scientific community releases individual scientists from some of their responsibilities. Some of the choices for which they were responsible before are no longer required, following the proposal from Levi (1960). In response to the version of the error argument from Rudner (1953), he proposed that a standard of evidence can be conventionally fixed. The convention then determines if a given inference was successful. A scientist, instead of making value-laden decisions, just needs to follow the convention.[12] A notable example of the conventionally fixed standard of evidence is the value of formal statistical significance set to 0.05. This standard is commonly used in psychology (see e.g., McLeod, 2019), but other standards of evidence are used in other scientific disciplines; for example, a much stricter 5σ standard is used in physics (see e.g., Franklin, 2013). Recently, in response to the replication crisis, some conventions were proposed to regulate the statistical power required for an experiment to make it appropriate for publication. For example:

> If the effect being examined is likely to be in the range of typical published effects in social/personality psychology, which is an r of .21 or d of .43 (Richard et al.

---

[12] Wilholt (2008) criticizes a convention aimed at setting the acceptable standard of evidence proposed in Levi (1962), but he does not argue that a conventional strategy will not be successful. As we have seen in the third section, he himself adopts a similar solution.

2003), then researchers should aim for a sample size that will provide at least 80% power to detect such an effect or more... (Vazire, 2016, p. 4)

The convention specifies the statistical power for which scientists should aim in order to get published. If we interpret the question about the necessary level of evidence posed by proponents of value-laden science in terms of the statistical power and significance, then a scientist does not have to decide which standard of evidence to adopt. Instead, she just has to conform to the convention, which is at the same time a precondition for being published. Similarly, there are conventions in place that regulate other aspects of experimental design, like conventions regulating the use of experimental animals discussed above or following conventions regulating some of the aspects of experiments testing the efficacy of β-alanine:

> "The diversity of the methodologies used in the assessment of the efficacy of β-alanine highlights the importance of clear logical progression through the different aspects of supplementation to eventually be able to produce a clear concise set of criteria for its efficacy. Authors should also make every attempt to demonstrate: the purity of the supplement used; the double-blinding of the treatments; and the reliability of the exercise tests or measure employed." (Hobson et al., 2012, p. 34)

As we see, here the authors express a need for a conventionally fixed efficiency criterion. Then, in the next sentence, they formulate a few methodological recommendations, for example, one concerning a need to demonstrate the purity of the substance used in the experiment. With the support of the community of scientists, the advice can become a fully-fledged convention. All those and similar conventions are steps toward the realization of Levi's strategy.

It is easy to see that both strategies are similar. Conventions play a role analogous to assumptions. Neither is meant to be tested, and both impose a particular methodological decision. From the perspective of an individual scientist, it seems to make little difference if she assumes something or follows a convention to the same effect. If an assumption becomes popular and well-established enough, it can be turned by an appropriate sanctioning body into a convention. On the other side, in line with John (2015b), it may make a lot of difference for consumers of the results. Conventions that a scientist follows in a given experiment are often implicit and do not necessarily need to be stated in the presentation of the results. Therefore, Levi's strategy may provide results that are easier to interpret and utilize than heavily hedged results criticized in John (2015b).

Many reforms aimed at reducing the methodological flexibility of experiments (and therefore the need for the use of non-epistemic values) were proposed in reaction to the replication crisis. As we have seen, Ioannidis (2005) lists such flexibility as one of the main factors that increase the probability that a result of a scientific experiment is false. In response to this problem, he recommends standardizing the conduct and reporting of research designs (see also Schulz, 2010). Similar proposals include preregistration (see e.g., Nosek et al., 2018), which amounts to a requirement that the experimental design used by a scientist has to be specified and registered before the data are collected, or multiverse analysis (see e.g., Steegen et al., 2016), which consists of conducting

a statistical analysis for some/all methodological choices that can be made during data analysis. Compliance with those conventions restricts how scientists can use their non-epistemic values. In principle, it is possible that the further development of scientific methodology will lead to the construction of a system of conventions and other precautions that will eventually make value-laden choices unnecessary. If so, the VFI may be realizable after all.[13]

It is sometimes argued that scientific conventions need to be justified by considerations based on non-epistemic values, and therefore the conventional strategy is unsuccessful in realizing VFI. Scientific conventions clearly have practical consequences. For example, conventions concerning the acceptable level of statistical power translate into an expected probability of type II error, which in turn will influence political decisions. A similar example was described in Wilholt (2022). He argues that implicit conventions present in climate science promote underestimation of indicators of global warming in comparison to actual changes. Such results influence political decisions and lead to under-regulation. One might expect that those practical consequences will translate to (a need for) the pragmatic justifications of such conventions, based in part on the non-epistemic values. Interestingly, justifications presented for those and similar scientific conventions are typically epistemic. For example, experiments need to be well-powered in order to be reliable, which justifies the convention from Vazire (2016). Similarly, the rich discussion of the optimal level of statistical significance (see e.g., Benjamin et al., 2017; Fiedler et al., 2012, or Oberauer & Lewandowsky, 2019) focuses on epistemic reasons for the adoption of the stricter criterion for statistical significance (e.g., $\alpha > 0.005$ proposed in Benjamin et al., 2017). The justifications presented for all other conventions mentioned in the paper seem to be purely epistemic. Moreover, even in the case of the biased convention with practically problematic consequences described in Wilholt (2022), the primary deficiency seems to be epistemic. The results in question fail to deliver empirically adequate results (there is a discrepancy between predictions and observations) due to the politically motivated, conservative convention used during the study. Consequently, the justification for a new improved version of the convention will likely be epistemic.

In many cases, the epistemic justification will single out a unique optimal convention. Randomization or blinding are uniquely epistemically beneficial approaches at many stages of experiments, for example, during assigning participants to control experimental groups. On the other hand, in some cases, the epistemic reason will not determine which one of many possible conventions will be optimal. For example, there likely is not a good epistemic reason to instantiate a convention that will ask scientists to aim for a statistical power of 80% rather than 81% or 79%. Consequently, one may claim that non-epistemic values need to be used in order to distinguish which of the infinitely many possible values should be used. I do not find this diagnosis plausible. In many cases, the content of the convention does not need to be justified in order for conventions to fulfill their regulatory function.[14] Conventions are typically understood as solutions to coordination problems, a situation in which agents have several possible

---

[13] For papers arguing for the realizability of VFI from a different angle see Mattes (2019) or Djørup et al. (2019).

[14] This feature of convention seems to motivate Poincare's conventionalism (Poincare, 1905). General laws of nature cannot be fully confirmed by empirical observations and at some point, further tests have a

ways to coordinate their actions for mutual benefit. The content of the convention to drive on the left side of the road does not require justification and perhaps cannot be justified. The fact that the members of society agreed to drive on one side of the road is useful, no matter which side of the road was chosen. Likely there is no reason to prefer conventions prescribing either of the sides and consequently, there seems not to be much point in asking for justification for the content of the convention. Plausibly, the situation is similar in the case of equally (or close to equally) plausible scientific conventions. For instance, there may be no strong (epistemic or non-epistemic) reason to prefer 80% rather than 81% if both values fall within an epistemically plausible range.

I will not be able to demonstrate that no scientific convention was justified by non-epistemic values or has to be justified in this manner. Plausibly, there are examples of such conventions, and perhaps a hedging strategy can be employed in some of these cases. Furthermore, adherence to scientific conventions often comes with costs, and in certain situations, these costs may be prohibitive. I will not discuss these costs and the trade-off between the epistemic benefit and the cost of conventions. On the other hand, I argued that many important methodological conventions were justified based on epistemic considerations, and in some cases, exhaustive justification may not be necessary.

The existence of those strategies (hedging and convention) not only makes a strong case for the realizability of the VFI but also gives us a clear idea of what a scientist can do to approach it. She has to follow the standard scientific conventions for a given problem when justifying her claim. Whenever the conventions do not specify which methodological decision should be made, she should make the decision herself and explicitly present it as an assumption of her result. As we have seen, even if the choice is motivated by her non-epistemic values, a hedged conclusion does not depend on them. This easy-to-follow procedure restricts the flexibility of the design. Either the convention was followed and therefore the non-epistemic values of scientists were not involved or the final result was hedged and because of that, it does not depend on the non-epistemic values of the scientist. Additionally, she should preregister her study and if possible use a multiverse analysis. All of those methodological improvements reduce the flexibility of the design of the study (and the potential influence of non-epistemic values) and therefore promote replicability (see also van Dongen & Sikorski, 2021). The improvements in replicability are likely proportional to the achieved reduction in flexibility and do not depend on a given procedure being fully methodologically rigid (i.e., involving no methodological flexibility) or even the possibility that such rigidity can ever be achieved. Even if some of the choices requiring non-epistemic values are unavoidable, and as far as I know it was not convincingly shown that there are such choices, it does not make removing others any less beneficial. All this is consistent with my argument concerning the realizability of VFI not being a condition for its usefulness from the first section. In conclusion, it seems that if we combine these strategies with the above argument, the VFI becomes, once again, a viable and perhaps even attractive ideal of scientific conduct.

Footnote 14 continued

marginal utility. In such a case, we can elevate the status of the law by conventionally accepting it to be true.

7 Conclusion

In my article, I defended the VFI by showing that if we reject it, we are forced to accept, as legitimate scientific conduct, some of the problematic scientific practices like founder bias. I argued that it does not have to be realizable to be useful. I presented counterexamples to some of the main value-laden proposals. I also showed that value-laden science contributes to the replication crisis. Finally, I presented strategies for making the VFI realizable. Following Betz (2013), a scientist can hedge her final result and therefore make it independent from her (value-laden) methodological choices. Secondly, in line with Levi (1960), a scientific community can instantiate a scientific convention that recommends a particular solution for a given methodological problem and therefore makes a corresponding (value-laden) choice unnecessary.

Declarations

Conflict of interest There are no conflicts of interest.

References

Abramson, J. D., & Starfield, B. H. (2005). The effect of conflict of interest on biomedical research and clinical practice guidelines: Can we trust the evidence in evidence-based medicine? *The Journal of the American Board of Family Practice, 18*(5), 414–8.

Anderson, E. (2004). Uses of Value Judgments in Science: A general argument, with lessons from a case study of feminist research on divorce. *Hypatia, 19*(1), 1–24.

Angell, M. (2004). The truth about the drug companies: How they deceive us and what to do about it.

Annett, M., & Kilshaw, D. (1984). Lateral preference and skill in dyslexics: Implications of the right shift theory. *Journal of Child Psychology and Psychiatry, 25*(3), 357–377. https://api.semanticscholar.org/CorpusID:8172232

Anwari, F., & Lakens, D. (2019). The replicability crisis and public trust in psychological science. *Comprehensive Results in Social Psychology, 3*, 1–21. https://doi.org/10.1080/23743603.2019.1684822

Asatryan, Z., et al. (2020). Biases in fiscal multiplier estimates. *Econometric Modeling: Macroeconomics eJournal, 63*, 101861.

Benjamin, D. J., et al. (2017). Redefine statistical significance. Nature Human. *Behaviour, 2*, 6–10.

Betz, G. (2013). In defence of the value free ideal. *European Journal for Philosophy of Science, 3*(2), 207–220.

Bishop, D. V. (1983). How sinister is sinistrality? *Journal of the Royal College of Physicians of London, 17*, 161–172. https://api.semanticscholar.org/CorpusID:45454921

Bishop, D. (1990). How to increase your chances of obtaining a significant association between handedness and disorder. *Journal of Clinical and Experimental Neuropsychology, 12*(5), 812–816.

Buckwalter, W., & Turri, J. (2015). Inability and obligation in moral judgment. *PLoS ONE, 10*, 8.

Cammack, M. (1994). In search of the post-positivist jury. *Indiana Law Journal, 70*, 405.

Chituc, V., et al. (2016). Blame, not ability, impacts moral ought judgments for impossible actions: Toward an empirical refutation of ought implies can. *Cognition, 150*, 20–25.

ChoGlueck, C. (2018). The error is in the gap: Synthesizing accounts for societal values in science. *Philosophy of Science, 85*(4), 704–725.

ChoGlueck, C. (2021). Drug facts, values, and the morning-after pill. *Public Affairs Quarterly, 35*(1), 51–82.

Churchman, C. W. (1948). Statistics, pragmatics, induction. *Philosophy of Science, 15*(3), 249–268.

Cohen, M. (2022). EXCLUSIVE: Juror speaks out after convicting Elizabeth Holmes. ABC News. https://abcnews.go.com/US/exclusive-juror-speaks-convicting-elizabeth-holmes/story?id=82082789

Djørup, S., Klemens, K., & Bjørn Gunnar, H. (2019). Can we comply with the ideal of value-freedom? A reply to Miller's critique of the ideal of value-freedom in science. *Ethics, Policy and Environment, 22*(1), 90–99.

Douglas, H. (2009). *Science, policy, and the value-free ideal*. University of Pittsburgh Press.

Douglas, H. (2014). The moral terrain of science. *Erkenntnis, 79*(S5), 1–19.

Dupré, J. (2007). Fact and value, (pp. 27–41).

Elliott, K. (2011). *Is a little pollution good for you?: Incorporating societal values in environmental research*. Oxford University Press.

Feldman, F. (1986). *Doing the best we can: An essay in informal deontic logic*. D. Reidel Publishing Company.

Fiedler, K., Kutzner, F., & Krueger, J. I. (2012). The long way from α-error control to validity proper. *Perspectives on Psychological Science, 7*, 661–669.

Franklin, A. (2013). *Shifting standards: Experiments in particle physics in the twentieth century*. University of Pittsburgh Press.

Friedberg, M., Saffran, B., Stinson, T. J., Nelson, W., & Bennett, C. L. (1999). Evaluation of conflict of interest in economic analyses of new drugs used in oncology. *JAMA, 282*(15), 1453–1457. https://doi.org/10.1001/jama.282.15.1453. PMID: 10535436.

Geschwind, N., & Behan, P. (1982). Left-handedness: Association with immune disease, migraine, and developmental learning disorder. *Proceedings of the National Academy of Sciences, 79*(16), 5097–5100. https://api.semanticscholar.org/CorpusID:35512404

Gobert, J. J. (1988). In search of the impartial jury. *Journal of Criminal Law and Criminology, 79*, 269.

Graham, P. (2011). "Ought" and ability. *Philosophical Review, 120*(3), 337–382.

Harris, A. J. (1957). Lateral dominance, directional confusion, and reading disability. *The Journal of Psychology, 44*(2), 283–294. https://api.semanticscholar.org/CorpusID:144996270

Hicks, D. J. A. (2014). New direction for science and values. *Synthese, 191*(14), 3271–95.

Hobson, R., et al. (2012). Effects of β-alanine supplementation on exercise performance: A meta-analysis. *Amino Acids, 43*(1), 25–37.

Holman, B., & Wilholt, T. (2021). The new demarcation problem. *Studies in History and Philosophy of Science, 91*, 211–220.

Howard, D. (2006). Lost wanderers in the forest of knowledge: Some thoughts on the discovery-justification distinction. In J. Schickore & F. Steinle (Eds.), *Revisiting discovery and justification: Historical and philosophical perspectives on the context distinction* (pp. 3–22). Springer.

Hudson, R. (2016). Why we should not reject the value-free ideal of science. *Perspectives on Science, 24*(2), 167–191.

Hudson, R. (2020). Should we strive to make science bias-free? A philosophical assessment of the reproducibility crisis. *Journal for General Philosophy of Science/Zeitschrift f ü r Allgemeine Wissenschaftstheorie, 53*, 1–17.

Hudson, R. (2022). Rebuttal to Douglas and Elliott. *Journal for General Philosophy of Science, 53*, 211–216.

Hughes, N. (2018). Guidance, obligations and ability: A close look at the action guidance argument for ought-implies-can. *Utilitas, 30*(1), 73–85.

Huxley, J. (1949). Soviet genetics: The real issue. *Nature, 163*, 935–942.

Intemann, K. (2015). Distinguishing between legitimate and illegitimate values in climate modeling. *European Journal for Philosophy of Science, 5*(2), 217–232.

Ioannidis, J. P. (2005). Why most published research findings are false. *PLoS Medicine, 2*(8), 55–69.

Jeffrey, R. C. (1956). Valuation and acceptance of scientific hypotheses. *Philosophy of Science, 23*(3), 237–246.

John, S. (2015). Inductive risk and the contexts of communication. *Synthese, 192*(1), 79–96.

John, S. (2015). The example of the IPCC does not vindicate the value free ideal: A reply to Gregor Betz. *European Journal for Philosophy of Science, 5*(1), 1–13.

Kissinger-Knox, A., Aragon, P., & Mizrahi, M. (2018). Ought implies can, framing effects, and empirical refutations. *Philosophia, 46*(1), 165–182.

Kitcher, P. (2011). *Science in a democratic society*. Prometheus Books.

Koertge, N. (2004). How might we put gender politics into science? *Philosophy of Science, 71*(5), 868–879.

Kourany, J. A. (2003). A philosophy of science for the twentyfirst century. *Philosophy of Science, 70*(1), 1–14.

Kourany, J. A. (2020). The new worries about science. *Canadian Journal of Philosophy, 52*, 1–19.

Krimsky, S. (2013). Do financial conflicts of interest bias research? *Science, Technology, & Human Values, 38*, 566–587.

Lacey, H. (2018). Roles for values in scientific activities. *Axiomathes, 28*(6), 603–618.

LeBel, E., et al. (2018). A unified framework to quantify the credibility of scientific findings. *Advances in Methods and Practices in Psychological Science, 3*, 389–402.

Levi, I. (1960). Must the scientist make value judgments? *Journal of Philosophy, 57*(11), 345–357.

Levi, I. (1962). On the seriousness of mistakes. *Philosophy of Science, 29*(1), 47–65.

Lexchin, J. R. (2006). Bigger and better: How Pfizer redefined erectile dysfunction. *PLoS Medicine, 3*, e132.

Longino, H. E. (1990). *Science as social knowledge: Values and objectivity in scientific inquiry*. Princeton University Press.

Ludwig, D. (2015). Ontological choices and the value-free ideal. *Erkenntnis, 6*, 1–20.

MacCoun, R. (1998). Biases in the interpretation and use of research results. *Annual Review of Psychology, 49*, 259–87.

Mattes, J. (2019). Reports of the death of value-free science are greatly exaggerated. *Review of Philosophy and Psychology, 10*(4), 689–699.

May, J. (2020). Bias in science: Natural and social. *Synthese, 199*, 3345–3366.

McLeod, S. (2019). What a p-value tells you about statistical significance. Simply Psychology. https://www.simplypsychology.org/p-value.html

Menon, T., & Stegenga, J. (2023). Sisyphean science: Why value freedom is worth pursuing. *European Journal for Philosophy of Science, 13*(48), 1–24.

Mizrahi, M. (2009). Ought 'does not imply' can. *Philosophical Frontiers, 4*(1), 19–35.

Mizrahi, M. (2012). Does 'ought' imply 'can' from an epistemic point of view? *Philosophia, 40*(4), 829–840.

Montori, V. M., et al. (2005). Randomized trials stopped early for benefit: A systematic review. *JAMA, 294*(17), 2203–2209.

National Toxicology Program. (2001). *National Toxicology Program's report of the endocrine disruptors low-dose*.

Nosek, B. A., et al. (2018). The preregistration revolution. *Proceedings of the National academy of Sciences of the United States of America, 115*(11), 2600–2606.

Oberauer, K., & Lewandowsky, S. (2019). Addressing the theory crisis in psychology. *Psychonomic Bulletin & Review, 26*, 1596–1618.

Okruhlik, K. (1994). Gender and the biological sciences. *Canadian Journal of Philosophy, 24*(sup1), 21–42.

Oostrom, T. (2023). Funding of clinical trials and reported drug efficacy.

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science, 349*, 6251.

Oreskes, N., & Conway, E. M. (2010). *Merchants of doubt: How a handful of scientists obscured the truth on issues from tobacco smoke to global warming*. Bloomsbury Press.

Pashler, H., & Wagenmakers, E.-J. (2012). Editors' introduction to the special section on replicability in psychological science: A crisis of confidence? *Perspectives on Psychological Science, 7*, 528–530.

Poincaré, H. (1905). The principles of mathematical physics. *The Monist, 15*(1), 1–24.

Purdy, S., et al. (2017). Debates about conflict of interest in medicine: Deconstructing a divided discourse. *Journal of Bioethical Inquiry, 14*, 135–149.

Randazzo, S. (2021). Former Theranos lab worker testifies about data manipulation. *The Wall Street Journal*. https://www.wsj.com/livecoverage/elizabeth-holmes-trial-theranos/card/u0WLLvDzUeIQTZ98jqf9

Rianne de, H., & Grünwald, P. (2017). Why optional stopping is a problem for Bayesians.

Richard, F. D., Bond Jr, C. F., & Stokes-Zoota, J. J. (2003). One hundred years of social psychology quantitatively described. *Review of General Psychology, 7*(4), 331–363. https://api.semanticscholar.org/CorpusID:145794181

Romero, F. (2017). Novelty versus replicability: Virtues and vices in the reward system of science. *Philosophy of Science, 84*(5), 1031–1043.

Rudner, R. (1953). The scientist qua scientist makes value judgments. *Philosophy of Science, 20*(1), 1–6.

Scargle, J. (2000). Publication bias: The File-Drawer problem in scientific inference. *Journal of Scientific Exploration, 14*, 91–106.

Schachter, S. C., Ransil, B. J., & Geschwind, N. (1987). Associations of handedness with hair color and learning disabilities. *Neuropsychologia, 25*(1), 269–276. https://snjrnlpms.straive.com/utilities/pes.php https://api.semanticscholar.org/CorpusID:6656317

Schroeder, S. A. (2020). Thinking about values in science: Ethical versus political approaches. *Canadian Journal of Philosophy, 52*, 246–255.

Schulz, K. F., Altman, D. G., & Moher, D. (2010). CONSORT 2010 Statement: Updated guidelines for reporting parallel group randomised trials. *BMJ, 340*, 100–107.

Sikorski, M. l., & Andreoletti, M. (2023). Epistemic functions of replicability in experimental sciences: Defending the orthodox view. *Foundations of Science*.

Simmons, J. P., Leif, D., & N., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science, 22*(11), 1359–1366.

Sober, E. (2015). *Ockham's razors: A user's manual*. Cambridge University Press.

Steegen, S., et al. (2016). Increasing transparency through a multiverse analysis. *Perspectives on Psychological Science, 11*(5), 702–712.

Steel, D. (2010). Epistemic values and the argument from inductive risk. *Philosophy of Science, 77*(1), 14–34.

Steel, D. (2017). Qualified epistemic priority: Comparing two approaches to values in science.

Stelfox, H. T., et al. (1998). Conflict of interest in the debate over calcium-channel antagonists. *New England Journal of Medicine, 338*(2), 101–106.

Tsui, A. (2016). Reflections on the so-called value-free ideal: A call for responsible science in the business schools. *Cross Cultural & Strategic Management, 23*(1), 4–28.

United Nations Office on Drugs and Crime. (2019). *Commentary on the Bangalore principles of judicial conduct*. United Nations Office on DrugsÂ / Crime.

United States Attorney's Office for the Northern District of California. U.S. v. Elizabeth Holmes, et al. (2022). United States Attorney's Office for the Northern District of California. https://www.justice.gov/usao-ndca/us-v-elizabeth-holmes-et-al

van Dongen, N., & Sikorski, M. (2021). Objectivity for the research worker. *European Journal for Philosophy of Science, 11*, 93.

Vazire, S. (2016). *Editorial. Social Psychological and Personality, 7*(1), 3–7.

Vazire, S. (2018). Implications of the credibility revolution for productivity, creativity, and progress. *Perspectives on Psychological Science, 13*(4), 411–417.

Vranas, P. B., & Acirc, M. I. O. (2007). Therefore I can. *Philosophical Studies, 136*(2), 167–216.

Wicherts, J. M., et al. (2016). Degrees of freedom in planning, running, analyzing, and reporting psychological studies: A checklist to avoid p-hacking. *Frontiers in Psychology, 7*, 1832.

Wilholt, T. (2008). Bias and values in scientific research. *Studies in History and Philosophy of Science Part A, 40*(1), 92–101.

Wilholt, T. (2022). Epistemic interests and the objectivity of inquiry. *Studies in History and Philosophy of Science Part A, 91*(C), 86–93.

Zimmerman, M. J. (1996). *The concept of moral obligation*. Cambridge University Press.