Department of Philosophy, Logic and Scientific Method

Programme: MSc in Philosophy and Public Policy

Academic Year: 2024/25

Course: PH458 Evidence and Policy

Essay question: What is the value free ideal and should scientists strive to uphold it?

Author: Maksim Karliuk

Assessment: Distinction

This essay was written as part of the requirements for the course Evidence and Policy within the programme of postgraduate study MSc in Philosophy and Public Policy

What is the value free ideal and should scientists strive to uphold it?

There are two main strands of arguments regarding the value-free ideal (VFI): desirability and achievability (Reiss and Sprenger 2020). In this essay, I will argue for what I will call a compatibilist account of upholding the VFI focusing on its desirability even if the VFI is unachievable. First, I will explain what the VFI is. Second, I will show that striving to uphold the VFI (desirability) is compatible with the rejection of its achievability. Third, I will demonstrate that the main arguments against the VFI do not refute its desirability. Finally, I will provide arguments on why it is desirable to strive to uphold the VFI even if the VFI is unachievable and show what role it can play in scientific inquiry.

There is no single definition of the VFI, yet the most common way to interpret it is that non-epistemic values ought not to influence scientific reasoning (Brown 2024, 2). Non-epistemic values are understood as certain ethical, social, cultural or political considerations. Therefore, it is the role of epistemic values, such as accuracy, consistency, empirical adequacy and simplicity, to be part of and to ensure proper scientific reasoning.

There seems to be a near consensus among philosophers of science rejecting the VFI (Holman and Wilholt 2022). If that is true, one of the consequences thereof is that upholding the VFI prima facie appears to violate the *ought implies can* principle, which stipulates that if a theory requires someone to do something, they are able to do that. Thus, the requirement to uphold the VFI would assume that one can in fact gather evidence and accept scientific theories based only on epistemic values, while that would not be the case.

Not resorting to non-epistemic values in scientific reasoning is sometimes referred to as the value-neutrality thesis (VNT) (Reiss and Sprenger 2020). Thus, VNT is a descriptive statement and is related to the 'can' version of the argument from inductive risk (AIR) as formulated by Zina B. Ward (2021, 16): non-epistemic values *can* enter into choices about hypothesis acceptance that run inductive risk (i.e. the risk of potential consequences of erring in accepting a false claim or rejecting a true one). This formulation presupposes VNT since it is in the form of permission, where it is at the moral discretion of scientists to make use of non-epistemic values or not, and if the choice is negative, no such values allegedly would enter such choices.

The VFI, a normative statement stemming from the fact that it is an ideal, especially if framed in terms of striving to uphold it, on its own does not imply the VNT. This is similar to other ideals: the fact that it is virtually impossible not to lie at every conceivable instance does not mean that we should not strive to uphold the ideal 'not to lie'. Moreover, it can be necessary or morally required to deviate from the ideal, provided specific circumstances. Although the analogy is imperfect as it is not epistemological as is the

case in a scientific context, the parallel is in the function as a guiding principle shaping the actions and minimizing deviations.

The most prominent argument against the VFI is the aforementioned AIR (Rudner 1953). However, even the strongest, 'must' form of AIR (AIR_m) as per Ward's classification, which states that "values *must* (in a non-moral sense) play a role in choices about hypothesis acceptance that run inductive risk" (Ward 2021, 16–17) does not necessitate abandoning the VFI. If AIR_m is true, it necessitates rejecting the VNT as AIR_m makes it impossible for scientists to get around non-epistemic values. Nevertheless, it remains conceivable to employ non-epistemic values in science where AIR_m requires it, e.g. for the purposes of operationalizing and trading off epistemic values, while continuing to strive for the VFI. In this regard, Mathew J. Brown's requirement that "any defence of the VFI must show either that [AIR] form is invalid, or that one of the premises is false" (Brown 2024) is unclear since it does not seem to be applicable to the desirability defence.

A prominent critique of the VFI's desirability stems from the feminist philosophy of science. The claim is that the VFI usually disguises "the unexamined dominance of a narrow range of hegemonic values" (Menon and Stegenga 2023, 7). Therefore, instead of striving to uphold the VFI, one should enrich scientific reasoning with more non-epistemic values, the difference being that such values should be more diverse and representative (Longino 2004). Moreover, it is claimed that the move away from androcentrism in primatology – the field which was dominated by men until the 1970s, resulting in biased and erroneous theories about female reproductive strategies and social hierarchies – has not been a move to the VFI but a replacement of one set of values with feminist values (Longino 1992).

The claim about the predominance of certain non-epistemic values has merit, contributes to the understanding of biases in scientific inquiry and is supported by other historical examples. Nevertheless, there is an issue with the argument against the VFI itself, which is two-fold. First, the claim that the VFI is a disguise for a particular set of non-epistemic values seems, in fact, to be a claim about the dominance of the VNT. Thus, in the primatology example, scientists appear to have been under the impression that they were pursuing their inquiry free of non-epistemic values, unaware of the fact that, in reality, it was not the case. However, on its own, the claim does not require the rejection of the VFI; it rather requires explicitly rejecting the VNT and reinforcing the need for scientists to consciously examine and manage the influences of non-epistemic values.

Second, intentionally introducing non-epistemic values in the scientific and related practice, particularly without proper constraints in place, can produce outcomes incompatible with expected standards, e.g. those of inference. This seems to have been the case in the development of some aspects of AI technologies. On the one hand, non-recognition that technology is not value-neutral (Miller 2021) by developers and other actors involved at different stages of the AI system life cycle has led to multiple problems

with the outputs of such systems, including widespread biases. On the other hand, the solution employed, at least in some instances, appears to have been aligned with the one proposed by Helen Longino discussed above – enriching the methods with other non-epistemic values – and has had negative results of its own. For example, in the case of the Gemini large language model, it has led to overcompensations, which manifested in the production of historically inaccurate images such as women popes (Samuel 2024).

The worry is relevant not only for the cases of deliberate value-enrichment but also if arguments such as AIR are to be understood not as a rejection of the VNT but as a rejection of the VFI as a guiding ideal. The potential outcome of such approaches can transform scientific inquiry into just another space for an ideological battle, losing one of the unique selling points of science that it has – being closer to the discovery of truths than other ways of inquiry. Abandoning the VFI and relying on other considerations can introduce a level of permissiveness that might be detrimental to the core function of science. As the latter enjoys certain authority in public discourse, this could also undermine the credibility of science (Bright 2018).

Instead, the VFI can help in avoiding these outcomes even if the VNT is to be rejected. This can be achieved through VFI's original function as guidance for scientific research by reframing the ideal as a constraint in resorting to non-epistemic values and choosing among values. VFI's operationalisation in this capacity can take various forms, a promising one of which could amount to a proportionality test in the sense of *suitability*, *necessity* and *balancing* (Alexy 2017) modified for the purposes of scientific inquiry.

Thus, Heather Douglas's concern regarding the potential arbitrariness of choosing whether there is sufficient evidence to accept a claim, which is used as a reason for resorting to non-epistemic values (Douglas 2017, 83), might satisfy the *suitability* condition – causally interconnecting the aim pursued and the means used – as pursuing a legitimate aim. It would further require the determination of whether employing certain values goes beyond what is *necessary* to achieve such an aim and whether there are alternatives of doing so. Thus, Gregor Betz suggests that scientific statements that cannot be established beyond reasonable doubt can be hedged, i.e. "[a]llegedly arbitrary and value-laden decisions can be systematically avoided … by making uncertainties explicit and articulating findings carefully" (Betz 2013, 209). It is probably not likely that hedging can be successful as systematically as Betz claims, given the response to his proposal from a number of philosophers of science (see in Brown 2024, 112). However, he is particularly focused on the context of policy-making with the involvement of scientific advisors. If Betz is right about hedging at least in some instances of scientific policy advice informing a decision under uncertainty, employing non-epistemic values might go beyond what is necessary in such contexts. Further, Zina B. Ward and Kathleen A. Creel, although denying Betz's attempt to defuse the methodological critique of the VFI as such, suggest that his approach can be applicable for hedging public-facing claims

in science by reducing "the dependence of scientific claims on evaluative standards that are not shared" (Ward and Creel 2024, 1001).

If the two above tests are passed, finally, employing non-epistemic values has to be *balanced* in terms of not having an excessive impact on the inquiry as compared to the epistemic values within the pursued aim. The balancing exercise does not necessarily have to be overly restrictive: in principle, it can even accommodate the denial of epistemic priority – the idea that evidence and epistemic values are to be prioritised over considerations of non-epistemic values. In circumstances where the use of non-epistemic values is suitable and necessary but where epistemic priority might have far-reaching detrimental consequences, as shown by Brown (2017), it can potentially become a good reason for balancing in favour of non-epistemic values in some contexts.

In any event, it is evident that proportionality as a means-end testing or other methods of operationalisation are value-laden themselves. However, they offer a structured approach to deciding on non-epistemic values if and when they are warranted to be employed so that the main aim of scientific enquiry – the discovery of truths – is not compromised. In this framing, the VFI can be a solution to what Brown calls "the value-management question" (Brown 2024, 112), i.e. how ought the role of non-epistemic values be managed in science, as well as can potentially address his concern regarding Tarun Menon's and Jacob Stegenga's apparent assumption that non-epistemic values are never truth-apt (Menon and Stegenga 2023, 13).

Therefore, the compatibilist account of the VFI is that if arguments against the VFI hold, it is necessary and sufficient to reject the VNT, while retaining the VFI as a desideratum to strive to uphold. The account bridges the gap between the ideal of value-free science and the realities of scientific inquiry shaped by inductive risk. It will help ensure that scientists are acutely aware of the role of non-epistemic values in scientific inquiry and of the necessity of their management through a structured method. In this case, the VFI can play the role of a first-order approximation to the science in its original role and can be further operationalised through structured frameworks like proportionality testing to address the impact of non-epistemic values on the inquiry.

References

Alexy, Robert. 2017. ‘Proportionality and Rationality’. In *Proportionality: New Frontiers, New Challenges*, edited by Mark Tushnet and Vicki C. Jackson, 11–12. Comparative Constitutional Law and Policy. Cambridge: Cambridge University Press. https://doi.org/10.1017/9781316691724.002.

Betz, Gregor. 2013. ‘In Defence of the Value Free Ideal’. *European Journal for Philosophy of Science* 3 (2): 207–20. https://doi.org/10.1007/s13194-012-0062-x.

Bright, Liam Kofi. 2018. ‘Du Bois’ Democratic Defence of the Value Free Ideal’. *Synthese* 195 (5): 2227–45. https://doi.org/10.1007/s11229-017-1333-z.

Brown, Matthew J. 2017. ‘Values in Science: Against Epistemic Priority’. In *Current Controversies in Values and Science*. Routledge.

———. 2024. ‘For Values in Science: Assessing Recent Arguments for the Ideal of Value-Free Science’. *Synthese* 204 (4): 112. https://doi.org/10.1007/s11229-024-04762-1.

Douglas, Heather. 2017. ‘Why Inductive Risk Requires Values in Science’. In *Current Controversies in Values and Science*. Routledge.

Holman, Bennett, and Torsten Wilholt. 2022. ‘The New Demarcation Problem’. *Studies in History and Philosophy of Science* 91 (February):211–20. https://doi.org/10.1016/j.shpsa.2021.11.011.

Longino, Helen E. 1992. ‘Subjects, Power, and Knowledge: Description and Prescription in Feminist Philosophies of Science’. In *Feminist Epistemologies*. Routledge.

———. 2004. ‘How Values Can Be Good for Science’. In *Science, Values, and Objectivity*, edited by Peter K. Machamer and Gereon Wolters, 127--142. University of Pittsburgh Press.

Menon, Tarun, and Jacob Stegenga. 2023. ‘Sisyphean Science: Why Value Freedom Is Worth Pursuing’. *European Journal for Philosophy of Science* 13 (4): 48. https://doi.org/10.1007/s13194-023-00552-7.

Miller, Boaz. 2021. ‘Is Technology Value-Neutral?’ *Science, Technology, & Human Values* 46 (1): 53–80. https://doi.org/10.1177/0162243919900965.

Reiss, Julian, and Jan Sprenger. 2020. ‘Scientific Objectivity’. In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Winter 2020. Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/win2020/entries/scientific-objectivity/.

Rudner, Richard. 1953. ‘The Scientist Qua Scientist Makes Value Judgments’. *Philosophy of Science* 20 (1): 1–6.

Samuel, Sigal. 2024. ‘Black Nazis? A Woman Pope? That’s Just the Start of Google’s AI Problem.’ Vox. 28 February 2024. https://www.vox.com/future-perfect/2024/2/28/24083814/google-gemini-ai-bias-ethics.

Ward, Zina B. 2021. ‘On Value-Laden Science’. *Studies in History and Philosophy of Science Part A* 85 (February):54–62. https://doi.org/10.1016/j.shpsa.2020.09.006.

Ward, Zina B., and Kathleen A. Creel. 2024. ‘To Hedge or Not to Hedge: Scientific Claims and Public Justification’. *Philosophy of Science* 91 (4): 990–1008. https://doi.org/10.1017/psa.2024.17.