Theoria, 2022, 37(2), 181-207
https://doi.org/10.1387/theoria.22795

Ὁ Θεὸς ἀριθμετίζει
THEORIA

ISSN 0495-4548 – eISSN 2171-679X

Inductive risk: does it really refute value-freedom?

(Riesgo inductivo: ¿realmente refuta el ideal de la ciencia libre de valores?)

Markus Dressel*
Leibniz University Hannover and University of Hamburg

ABSTRACT: The *argument from inductive risk* is considered to be one of the strongest challenges for value-free science. A great part of its appeal lies in the idea that even an ideal epistemic agent—the "perfect scientist" or "scientist *qua* scientist"—cannot escape inductive risk. In this paper, I scrutinize this ambition by stipulating an idealized Bayesian decision setting. I argue that inductive risk does not show that the "perfect scientist" must, descriptively speaking, make non-epistemic value-judgements, at least not in a way that undermines the value-free ideal. However, the argument is more successful in showing that there are cases where the "perfect scientist" should, normatively speaking, use non-epistemic values. I also show that this is possible without creating problems of illegitimate prescription and wishful thinking. Thus, while inductive risk does not refute value-freedom completely, it still represents a powerful critique of value-free science.

KEYWORDS: inductive risk; value-free ideal; scientist *qua* scientist; Bayesianism; wishful thinking; prescription.

RESUMEN: *El* argumento del riesgo inductivo *se considera uno de los retos más fuertes a la ciencia libre de valores. Gran parte de su atractivo reside en la idea de que incluso un agente epistémico ideal—el "científico perfecto" o el "científico qua científico"—no puede evitar el riesgo inductivo. En este trabajo, examino esta idea estipulando un marco de decisión bayesiano idealizado. Argumento que el riesgo inductivo no muestra que el "científico perfecto" deba, en un sentido descriptivo, hacer juicios evaluativos no epistémicos, al menos no de manera que socave el ideal de la ciencia libre de valores. Sin embargo, el argumento tiene más éxito en mostrar que hay casos en los que el "científico perfecto" debería, en un sentido normativo, usar valores no epistémicos. También muestro que esto sin posible sin crear problemas de prescripciones ilegítimas y pensamiento desiderativo. Así, aunque el riesgo epistémico no refute completamente el ideal de la ciencia libre de valores, sí representa una poderosa crítica a este ideal.*

PALABRAS CLAVE: *riesgo inductivo; ciencia libre de valores; científico qua científico; bayesianismo; pensamiento desiderativo; prescripción.*

* ***Correspondence to:*** Markus Dressel. University of Hamburg, Research Unit Sustainability & Climate Risks, Research Group Sustainability & Global Change, Grindelberg 5 (20144 Hamburg-Germany) – markus.dressel@uni-hamburg.de – https://orcid.org/0000-0002-4789-5249

How to cite: Dressel, Markus (2022). «Inductive risk: does it really refute value-freedom?»; *Theoria. An International Journal for Theory, History and Foundations of Science,* 37(2), 181-207. (https://doi.org/10.1387/theoria.22795).

Received: 2021-05-08; Final version: 2022-05-04.

1. Introduction

For a long time, philosophers maintained that science can and should be value-free. In recent years, however, this *value-free ideal* (VFI) dramatically lost support (Biddle, 2013; Douglas, 2016; Elliott & Steel, 2017; Holman & Wilholt, 2022). Many specialists in science and values today reject value-freedom, either as a possibility (Wilholt, 2009; Biddle & Winsberg, 2010; Winsberg, 2012) or as an ideal (Douglas, 2009; Steel, 2016a; Elliott, 2011). One of the most salient arguments in this regard is the *argument from inductive risk* (AIR). From AIR's perspective, science is a sequence of decisions under epistemic risk. When the consequences of these decisions reach beyond science, AIR claims, scientists cannot or should not remain value-free. What is more, AIR promises to refute value-freedom *in principle*. As inductive risk is part of "the scientific method as such" (Rudner, 1953, p. 2), the idea goes, VFI fails even for the "perfect scientists—the scientist *qua* scientist" *(ibid.)*.

In this paper, I scrutinize exactly this ambition: does AIR refute value-freedom even under the "VFI-friendly" assumption of a "perfect scientist"? I begin by introducing VFI and its conceptual restrictions (sect. 2.1). I also present two reasons why value-freedom may be appealing in the first place, the *argument from prescription* (APr) and the *argument from wishful thinking* (AWT) (sect. 2.2). In the third section, I introduce AIR (sect. 3.1) and develop a Bayesian decision setting in which an ideal agent—the "perfect scientist"—maximizes the expected utility of a given scientific decision (sect. 3.2). In the fourth section, I use the idealized setting to scrutinize whether the "perfect scientist" is forced, descriptively speaking, to make non-epistemic value-judgements (sect. 4.1). I conclude that this is not the case, at least not in a way that refutes VFI. However, AIR is more successful in showing that the "perfect scientist" should, normatively speaking, sometimes use non-epistemic values (sect. 4.2). In the fifth section, I discuss how this is possible while avoiding prescriptiveness (sect. 5.1) and wishful thinking (sect. 5.2). I argue that, while APr and AWT rightfully warn against problematic ways of using values, these concerns can be countered by taking certain measures. In the conclusion, I discuss the objection that my idealized setting is practically irrelevant (sect. 5). I argue that this impression is false and that, quite on the contrary, my idealized approach helps to elaborate something that may be called *epistemic legitimacy*: a set of rules that should govern the use of values not only in an idealized setting, but also in actual science.

2. What is value-freedom and why would we want it?

2.1. Value-freedom: definition and restrictions

The version of the value-free ideal (VFI for short) that I discuss in this paper comprises a normative and a descriptive claim, as well as four restrictions that clarify the scope of these claims. The first claim, VFI_{norm}, describes value-freedom as a *regulative ideal*:

> VFI_{norm} Scientists *should*, as much as possible, avoid significant non-epistemic value-judgments when making genuinely scientific choices.

While some authors have focused on this normative part (e.g. Dorato, 2004; Douglas, 2009, chap. 3; Betz, 2013; Bueter, 2015), others have treated VFI as partially descriptive.

For these authors, VFI not only claims that scientists *should* be value-free, but also that they *can* be value-free (e.g. Biddle & Winsberg, 2010; Biddle, 2013; Reiss & Sprenger, 2020). I here follow this twofold interpretation, mostly because the normative and the descriptive part are connected via ought-implies-can (Kitcher, 2011, p. 31; Biddle, 2013, p. 131). I refer to the descriptive part as VFI$_{\text{desc}}$:

VFI$_{\text{desc}}$ Scientists *can*, at least in principle, avoid significant non-epistemic value-judgments when making genuinely scientific choices.

As we shall see later, inductive risk challenges the two claims in different ways. To show this, we need to consider VFI's restrictions (see e.g. Weber, 1949; Rudner, 1953; Reichenbach, 1961; Kuhn, 1977; McMullin, 1982; Lacey, 1999; Dorato, 2004; Douglas, 2009; Reiss & Sprenger, 2020):

VFI-R$_1$ VFI applies only to *non-epistemic* values.
VFI-R$_2$ VFI applies only to *genuinely scientific* decisions.
VFI-R$_3$ VFI applies only to *significant* value-judgments.
VFI-R$_4$ VFI applies only to value-freedom *in principle*.

Critics have argued that these restrictions, particularly VFI-R$_1$ and VFI-R$_2$, are analytically implausible or practically infeasible (e.g. Longino, 1996; Machamer & Douglas, 1999; Putnam, 2002; Dupré, 2007; Bueter, 2015; De Melo-Martin & Intemann, 2016). For the scope of this paper, however, I accept these restrictions. I do so not because I think that critiques of VFI-R$_1$–VFI-R$_4$ are irrelevant, but because I think that *inductive risk is philosophically interesting exactly because it challenges value-freedom even under assumptions that are favorable for value-freedom*. That is, if AIR defeats VFI even if we stipulate "VFI-friendly" (and perhaps counter-factual) conditions, then this would emphasize the gravity of inductive risk. Hence, while I do not engage in debates about VFI-R$_1$–VFI-R$_4$ here, I do contend that these restrictions are useful to study the strengths—and the limits—of inductive risk.

Let us therefore look into these restrictions. VFI-R$_1$ limits VFI to *non-epistemic* (e.g. ethical) values, but permits *epistemic* values such as explanatory power, scientific fruitfulness, or other values that may foster the attainment of truth (Kuhn, 1977; Steel, 2010). Later in this paper, I will refer to such values as *scientific utilities* (Hempel, 1981). A scientific choice will, e.g., have a high scientific utility if it unifies a research field or enables new lines of study, and a low scientific utility if it leads scientists to accept falsehoods or miss important truths (Steel, 2016a; Wilholt, 2016). The next restriction, VFI-R$_2$, limits VFI to the "heart of the research process" (Douglas, 2009, p. 45), i.e. to those activities that justify a research finding *qua* truth claim (Weber, 1949; Reichenbach, 1961; Hoyningen-Huene, 2006). This includes activities such as hypothesis assessment, model choice or data collection, but excludes obviously value-laden parts such as agenda-setting, real-world applications and ethical boundary conditions (Douglas, 2009; Biddle, 2013; Reiss & Sprenger, 2020). VFI-R$_3$ limits VFI to decisions that *significantly* impact final results, e.g. by turning a hypothesis acceptance into a rejection. This makes sure that, if we are to refute VFI, we need an argument that shows how values make an actual difference (rather than an argument that "exaggerates the influence of social values", Parker, 2014, p. 27).

VFI-R_4 is crucial for my analysis of inductive risk. This restriction implies that VFI_{desc} cannot be refuted by showing that scientists *de facto* fail to be value-free, but only by showing that they cannot be value-free *in principle*. I assume VFI-R_4 for three reasons. First, it played a crucial role when inductive risk was originally introduced (Rudner, 1953) (sect. 3.1). Second, VFI-R_4 mirrors parts of the recent debate: Reiss & Sprenger (2020) describe VFI_{desc} as the claim that scientists can "at least in principle" *(ibid.)* refrain from making value judgements; Kitcher (2011) takes proponents of value-freedom to claim that scientists should "in principle" (*ibid.,* p. 33) report only the evidence; and Ruphy (2006) holds that philosophically interesting critiques of value-freedom are "not only about what doesn't happen *in practice* in science, [but] about what cannot happen even *in principle*" (*ibid.,* p. 192, orig. italics). Third, the *de facto* value-ladenness of science has long been conceded by defenders of VFI. The point is that "[p]roponents of [VFI] may grant that perfectly value-neutral results are never or very rarely obtained in the actual development of science, for all that, value-neutrality remains the aim" (Ruphy 2006, p. 192) (see also Weber, 1949, p. 9; Popper, 1976, p. 97; Koertge, 2000, p. S53). Yet this classic defense presupposes that value-freedom can be achieved at least under idealized assumptions. The question, then, is whether inductive risk really shows that value-freedom fails in principle.

2.2. Value-freedom: underlying motivation

But why should we be interested in value-freedom in the first place? In this context, Holman & Wilholt (2022) use the metaphor of "Weber's fence". They argue that champions of value-freedom such as Max Weber had relevant reasons to set up a rule, or a "fence", that shields science from inacceptable value influences, and that we should not tear down this "fence" without considering the concerns behind it (see also Proctor, 1991). De Melo-Martin & Intemann (2016) discuss two such motivations: the political concern that "the use of contextual values in scientific reasoning allows scientists to impose their personal value judgments on others" (*ibid.,* p. 503); and the epistemic concern that value-laden science may imply that scientists "accept theories about 'how they wished the world to be' rather than 'how the world really is'" (*ibid.,* p. 502). I refer to the first concern as *argument from prescription* (or APr) and the second as the *argument from wishful thinking* (or AWT).

Versions of APr have been discussed by Weber (1958), Du Bois (1935), and more recently by Bright (2018), Betz (2013), De Melo-Martin & Intemann (2016), Intemann (2015), or John (2015; 2019). One way to reconstruct APr goes like this: real-world decision-making often relies on science, be it on an individual level, e.g. regarding a person's medical choices, or on a collective level, say in climate policy or substance regulation. The worry is that, if science would depend on non-epistemic values, scientists would effectively prescribe these decisions. This may violate democratic norms: "to the extent that scientists make value judgments, there are concerns that their values will be undemocratically privileged" (Intemann, 2015, p. 218). The concern can also be expressed as a matter of subjective freedom. Here, the worry is that individuals should be able to freely pursue their version of the good life, and that, to the degree that they depend on science to do so, "personal autonomy would be jeopardized if scientific findings [...] were soaked with moral assumptions" (Betz, 2013, p. 207).

Elaborating this a bit further, I suggest that APr is the claim that value-laden science places an illegitimate constraint on the decision space of extra-scientific agents. I suggest that such a constraint is illegitimate if and only if the constraint is *relevant* (eliminating decision options that the agents may actually take interest in), *external* (lacking the agents' explicit or implicit consent) and *normative* (resulting from scientists' non-epistemic values). The underlying principle is best described as the norm of autonomy, which I take to include both democratic and personal autonomy. APr then reads:

1. Extra-scientific agents often rely on science to determine their action plans.
2. Value-laden science constrains the decision space of those agents in a *relevant*, *external* and *normative* way.
3. Such constraints violate the principle of autonomy.
4. Observing VFI_{norm} is the only (or at least the best) available way to avoid this violation.
5. Therefore, VFI_{norm} is valid.

Proceeding to AWT, precursors of this concern can be traced back to Bacon's *Novum Organum* (book I, § 39-46) or Hume's *Treatise of Human Nature* (book III, part I, sect. I). In the recent debate, AWT has been addressed both by critics (e.g. Douglas, 2009, chap. 3; Brown, 2013; De Melo-Martin & Intemann, 2016) and defenders (e.g. Koertge, 2000; Haack, 2003) of value-freedom. AWT claims that "propositions about what states of affairs are *desirable* or *deplorable* [cannot] be evidence that things *are*, or *are not*, so" (Haack, 2003, p. 13, orig. italics). As science is to inform us about actual rather than desirable states of affairs, it seems to follow that science should be value-free. This reasoning stems from a principle called *no-is-from-ought*, the "mirror image" (Jones, 1999, p. 894) of Hume's famous law. Interestingly, however, current discussions of AWT rarely take up insights from meta-ethics (Schurz, 1997; Pidgen, 2016). Closing this gap, I suggest that AWT claims that an *ought-is* inference is logically invalid if and only if its descriptive conclusion is *direct* (derived exclusively from normative premises), *non-vacuous* (a substantial implication of the normative premises) and *not semantically entailed* (not hidden the semantics of the normative premises). AWT then reads:

1. Scientific choices should be truth-conducive.
2. Making a scientific choice dependent on non-epistemic values is a *direct*, *non-vacuous* and *not semantically entailed* inference from an ought-claim to an is-claim.
3. Inferences of such kind violate no-is-from-ought.
4. Observing VFI_{norm} is the only (or at least the best) available way to avoid this violation.
5. Therefore, VFI_{norm} is valid.

Now, critics of value-freedom can either reject APr and AWT straightforwardly, or they can accept these arguments in general, but claim that they apply only to *some* rather than *all* ways of using values in science. It is my impression that the second strategy reflects the standard view regarding APr, whereas critics are split regarding AWT. Some seem to claim that this concern is somewhat exaggerated (e.g. Brown, 2013; De Melo-Martin & Intemann, 2016), whereas others hold that wishful thinking is a real problem if not properly addressed (e.g. Douglas, 2009; Wilholt, 2009). In this paper, I will only consider the latter strategy. I do so because, for one, I believe that AWT has a strong *prima facie* plausibility and, for another, those who reject AWT tend to rely on arguments other than AIR (De Melo-Martin & Intemann, 2016). However, as my aim is to scrutinize inductive risk rather than to find alternative ways to attack VFI, I will not engage with these debates.

3. *Inductive risk challenges value-freedom*

3.1. The argument from inductive risk

Let us now discuss the *argument from inductive risk* (AIR for short). Apart from early precursors in scholasticism (Schuessler, 2019) or Blaise Pascal (see also James, 1912), AIR emerged in the middle of the past century (Churchman, 1948; Rudner, 1953; Hempel, 1965) and was later reintroduced by Heather Douglas (2000; 2009). The argument's *locus classicus* is a short article by Richard Rudner (1953). Rudner started by expressing his discontent with popular critiques of value-freedom. These critiques, he claimed, either argue that truth is itself a value, or that values are needed in scientific agenda-setting, or that scientists are imperfect human beings (*ibid.,* p. 1). We can easily see that, unless one rejects VFI's restrictions (sect. 2.1), none of this refutes value-freedom. In particular, the fact that science is *de facto* value-laden does not imply that this must be so *in principle* (VFI-R_4). Rudner found this unsatisfying, for as long as values "have not been shown to be involved in the scientific method as such" (*ibid.,* p. 2), it still stands that "[t]he perfect scientist—the scientist *qua* scientist—does not allow this kind of value judgment" *(ibid.).* Rudner acknowledged that such a "perfect scientist" was nowhere to be found in reality (*ibid.,* p. 4); yet he believed that this idealization was the adequate touchstone for his argument. Hence, Rudner's question (which is also the starting point of my own approach, sect. 3.2), was essentially counter-factual: *if there were such a thing as an ideal epistemic agent, would this agent make non-epistemic judgements when making genuinely scientific decisions?*

Rudner believed that the answer to this question is "yes". He argued that the scientific method intrinsically involves that "the scientist as scientist accepts or rejects hypotheses" (*ibid.,* p. 2). However, "no scientific hypothesis is ever completely verified" *(ibid.).* As there is no certainty in science, and as there is no science without hypothesis evaluation, "the scientist must make the decision that the evidence is *sufficiently* strong" *(ibid.)* before accepting a hypothesis. Rudner argued that the only way to determine "how strong is 'strong enough'" *(ibid.)* is to weigh the potential consequences of error. When the consequences concern extra-scientific goods, say public health, values are needed to determine how much evidence is needed. Therefore, Rudner argued, hypothesis evaluation is "a function of the *importance*, in the typically ethical sense, of making a mistake" (*ibid.,* orig. italics).

Rudner's reasoning has later been refined and generalized in several ways. First, recent contributions tend to differentiate between a normative version of AIR, i.e. one that attacks $\mathrm{VFI}_{\mathrm{norm}}$, and a descriptive version, i.e. one that attacks $\mathrm{VFI}_{\mathrm{desc}}$ (Betz, 2013; Steel, 2016a)[1]. Second, today's versions typically address not only hypothesis assessment, but also other genuinely scientific activities such as model choice or data interpretation (Douglas, 2000; Elliott, 2011; Wilholt, 2013; Biddle & Kukla, 2017). And third, some current inter-

[1] De Melo-Martin & Intemann (2016) suggest that AIR may claim that non-epistemic values are necessary in a *logical*, *epistemic*, *pragmatic*, or *ethical* sense. As far as I can see, the first two readings are instances of the descriptive reading, while the latter two are instances of the normative reading.

pretations of AIR address not only the consequences of erroneous scientific decision, but also of correct or suspended decisions (Wilholt, 2009; Wilholt, 2013; Steel, 2016a; Steel, 2016b). Taking up these developments, I interpret AIR in such a way that it includes both a normative and descriptive branch; also, I take AIR to address any genuinely scientific decision that significantly impacts the final results of a scientific study (similar to Biddle & Kukla, 2017)[2]; finally, I take AIR to address not only the consequences of error, but also those of truth, missed truth and averted error (similar to Wilholt, 2009) (see sect. 3.2).

AIR, then, is the claim that scientific decisions made by an ideal epistemic agent must, or should, include non-epistemic values if and only if these decisions are *underdetermined* (involving relevant uncertainty), *unavoidable* (forced upon the agent) and *momentous* (having potential consequences for ethically relevant extra-scientific goods). AIR thus reads:

1. An ideal epistemic agent *cannot avoid* making *underdetermined* scientific choices.
2. To make these choices, the agent must specify *evidential thresholds*.
3. If the choice is *momentous* via its potential consequences for ethically relevant extra-scientific goods, the agent *cannot* or *should not* (or both) specify the evidential threshold without considering non-epistemic values.
4. Therefore, if underdetermination, unavoidability and extra-scientific momentousness are given, $\mathrm{VFI}_{\mathrm{desc}}$ or $\mathrm{VFI}_{\mathrm{norm}}$ (or both) is (or are) invalid.

3.2. Inductive risk in an idealized setting

Before discussing whether AIR really defeats value-freedom and, if so, whether AIR can avoid prescriptiveness and wishful thinking, I want to suggest an idealized decision-theoretical approach to inductive risk[3]. My motivation is twofold: On the one hand, I take seriously Rudner's claim that AIR is not merely about actual scientists, but about an idealization—the "perfect scientist" (or "scientist *qua* scientist")[4]. As Katie Steele has pointed out, such counter-factual assumptions would strengthen AIR, "because it is more surprising in the ideal setting that scientists must make value judgments" (2012, p. 895). Also, an idealized approach provides an *in-principle* perspective on value-freedom (as demanded by $\mathrm{VFI\text{-}R}_4$). On the other hand, the approach sheds light on decision problems in *actual* science. Inductive risk has often been described as a balancing problem between exactly two

[2] Biddle & Kukla (2017) suggest to substitute the term "inductive" risk with "epistemic" risk. I agree that this terminology has virtues. However, I stick to the traditional term because it is commonly used in the debate, and because scientific choices that occur previous to hypotheses assessment (model choice, data collection, test calibration etc.) have basically one purpose: to make possible an inductive step from the evidence to a hypothesis acceptance/rejection (or suspension).

[3] The following part is inspired by Wilholt (2009; 2013) and was significantly improved in discussions with Benjamin Blanz and Hermann Held.

[4] This is sometimes overlooked. Rudner explicitly says that his considerations "do not have as their import that an empirical description of every present day scientist [...] would include the statement that he made a value judgment" (1953, p. 4). Rudner's point was rather that a "rational reconstruction of the method of science" *(ibid.)* would be incomplete if it did not address inductive risk. While Rudner noted that scientists are not "coldblooded, emotionless, impersonal" (*ibid.,* p. 6), he came to this conclusion not by considering actual science, but by analyzing an *impersonal* scientist qua scientist.

risks, where one risk is clearly preferable to the other (e.g. consumer versus producer risks, Biddle & Leuschner, 2015). As shown below, however, the decision problem is much more complex.

To represent Rudner's "perfect scientist", I stipulate an ideal agent with the following properties:

— *Preferences*: the agent prioritizes the advancement of science over extra-scientific aims.
— *Evidence*: the agent possesses perfect knowledge of the available evidence.
— *Rationality*: the agent makes decisions in a rule-based and unbiased manner.

Before this background, the agent considers a scientific decision D, where D may be any methodological choice that significantly impacts the final results of the study that D is a part of. As an example, imagine that the agent contemplates whether or not to use a certain model. The agent's decision space comprises two options: *perform D* (use the model) and *not perform D* (not use the model)[5]. A central assumption of AIR is that agent cannot be certain whether performing D (using the model) would lead to true study results[6]. The agent must therefore determine a threshold t above which the probability p that the results will be true, given that D is performed (the model is used), is sufficiently high. The agent would then perform D when p exceeds t, and not perform D when p falls short of t:

— Perform D iff $p > t$.
— Not perform D iff $p < t$.

The question, then, is how the agent determines the evidential threshold t. The classic answer (Rudner, 1953; Churchman, 1948) is that the threshold depends on how bad the consequences of error would be. However, this leaves open many critical issues: How exactly do D's consequences determine t? How should scientific and extra-scientific consequences be balanced? How should outcomes other than error influence t? What is the relation between the probability that D leads to an error and the probability that the error causes the assumed consequences? Another issue is that the classic approach interprets inductive risk in a frequentist manner. In frequentist statistics, p is an objective measure

[5] For reasons of simplicity, I focus on: (a) individual decisions rather than decision sequences; (b) binary decisions (e.g. use versus not use a model) rather than decisions with three options (e.g. accept, reject or suspend a hypothesis); and (c) decisions on single methodological items (e.g. a model) rather than contrastive decisions between different items (e.g. several competing models). Note, however, that my approach could in principle accommodate these types of decisions.

[6] The extent to which scientists can avoid uncertainty is contested (Betz, 2013; Parker, 2014; Steel, 2016a; Douglas, 2017). Note, however, that AIR need not assume that each and every scientific choice is fundamentally uncertain. In fact, I believe that this radical interpretation of AIR is either trivial or false. There must be a difference between the trivial uncertainty attached to, say, the assumption that radiative forcing is a relevant factor in the climate system, and the non-trivial uncertainty attached to, say, cloud parametrizations in a given climate model. However, for inductive risk to be relevant it suffices that non-trivial uncertainty is a *typical* feature of science, and that *in a relevant number of cases* this uncertainty cannot be avoided without sacrificing science's ability to produce meaningful results. This more modest reading of AIR accounts for the possibility of uncertainty hedging (Betz, 2013), while still reserving a crucial role for inductive risk.

for the likelihood with which a property that has been found in a number of observations O_1, ..., O_n will also be found in an observation O_{n+1} (Rudner 1953, p. 3). Our agent, however, is in a different epistemic situation: the available evidence may be too limited or inconsistent to determine an objective probability; the evidence may be incommensurate, e.g. because it includes data from heterogeneous sources; and the evidence does not, by definition, account for unknown unknowns. It is therefore more plausible to interpret p in a Bayesian manner, such that p represents the agent's *probabilistic beliefs* (see also Steele, 2012).

By adopting a Bayesian perspective, I also contend that the agent can be understood as a *utility maximizer*[7]. That is, the agent will choose the option that promises the highest relative benefit, given her preferences regarding the consequences and the probability that she assumes for these consequences to occur (Wilholt, 2009; Wilholt, 2013). Apart from addressing the above issues, this sheds light on the old problem (Kuhn, 1977) that epistemic values such as precision and scope can contradict each other. From a Bayesian perspective, it is irrelevant whether, say, a model's strengths in precision are countered by its weaknesses in scope, as this simply reduces the model's overall utility. The most crucial advantage, however, is that the Bayesian approach gives us a straightforward interpretation of the evidential threshold t, where *t is the point in the probability space at which the total expected utility* EU_{total} *of both decision options, i.e. perform* D *(PerfD) and not perform* D *(¬PerfD), converge*:

$$t = EU_{\text{total}}(PerfD) = EU_{\text{total}}(\neg PerfD)$$

To determine t, the agent must thus determine both options' total expected utilities. I here suggest to differentiate between *first-order* and *second-order* outcomes (see fig. 1). Drawing on Wilholt (2009; 2013), first-order outcomes include *truth* and *error* for performing D, and *missed truth* and *averted error* for not performing D. For instance, the agent may correctly decide to use a model that leads to valid study results (truth); erroneously decide to use a model that leads to false study results (error); erroneously decide not to use a model that would have led to valid study results (missed truth); or correctly decide not to use a model that would have led to false study results (averted error). Second-order outcomes are dependent on first-order outcomes, i.e. they may occur as a causal effect of truth, error, missed truth, or averted error. Second-order outcomes include all normatively relevant consequences that D may have for both scientific and extra-scientific goods. If the agent, e.g., uses a model that turns out to imply true study results (first-order outcome), this may enable new lines of study (scientific second-order outcome), while also supporting real-world decision-making in, say, climate policy (extra-scientific second-order outcome).

[7] Note that I do not claim that utility maximization is the only plausible candidate for a rational decision rule. What I do contend, however, is that the Bayesian perspective is superior to both the frequentist approach and the simplistic decision rule "the worse the error consequences, the higher the evidential threshold" (for reasons outlined above). Note furthermore that the Bayesian approach does not contradict the fact that scientists typically put special emphasis on error avoidance (Wilholt, 2009), as this can easily represented by asymmetrically decreasing the utility of error consequences. Finally, the Bayesian approach can be reconciled with the deontologist axiom that some decisions are *intrinsically* inacceptable. As I argue later (sect. 5.2), however, such cases should be interpreted as ethical rather than genuinely scientific choices.

The agent must therefore assess how good or bad each second-order outcome would be *if* it occurred, and how likely it is *that* it occurs, given the respective first-order outcome[8].

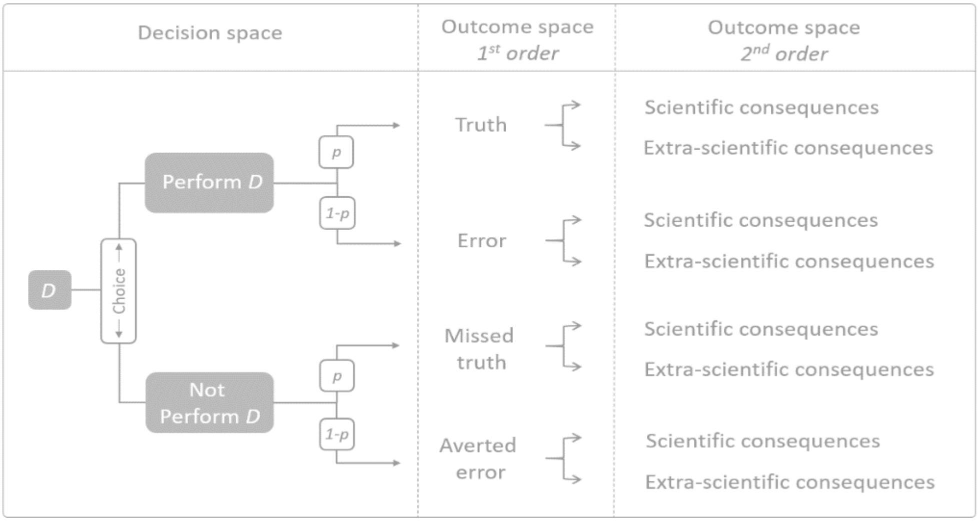

Figure 1. The structure of scientific decisions under inductive risk.

The next step is crucial: the agent must determine the right balance between scientific and extra-scientific utilities. To represent this choice, I suggest to introduce a trade-off parameter T. The agent uses T such that she weighs scientific utilities with T and extra-scientific utilities with $1 - T$. If VFI is valid, the agent will thus use $T = 1$; however, if AIR's attack on value-freedom is successful, the agent must or should (or both) use $T < 1$. The introduction of T gives us a more detailed perspective on AIR. Inductive risk debates have long focused either on *whether or not* non-epistemic values do or should influence scientific decisions, or on the *specific way* in which they should do so. While these questions are indeed crucial, they do not account for the *relative weight* that non-epistemic values should have, as compared to the epistemic ones. Scientific and extra-scientific utilities are different criteria, and merely knowing how high or low a decision option scores in one criterion does not tell us how important the criterion itself is (e.g. how important a scientifically valuable result is in comparison to improved real-world decisions). T represents this balancing problem in a more fine-grained way than, e.g., the notion of one type of value "trumping" (Elliott & McKaughan, 2014) the other.

[8] Probabilities in the second-order outcome space are *subjective* (i.e. they represent the agent's probabilistic beliefs) and *dependent* (i.e. they are estimated *given* the respective first-order outcome). Extra-scientific outcomes are generally uncertain, as the agent cannot know whether the study will actually influence real-world contexts. Scientific outcomes are uncertain if they refer to future research (e.g. a result's fruitfulness); however, the agent can be certain about some types of scientific outcomes, such as a result's scope or precision, as compared to existing results.

Table 1. The agent's decision algorithm.

Step	Description	Formalization
(1)	Determine the probability P of all second-order outcomes (2ndOrd), given the respective first-order outcome (1stOrd). First-order outcomes include truth, error, averted error, and missed truth; second-order outcomes include scientific (2ndOrdSci) and extra-scientific (2ndOrdExt) outcomes. Each first-order outcome may cause several scientific and several extra-scientific second-order outcomes. Extra-scientific second-order outcomes are generally uncertain; scientific second-order outcomes can either be certain or uncertain (see fn. 7).	$1 \geq P(2ndOrdSci_i \mid 1stOrd_k) \geq 0$ $1 > P(2ndOrdExt_i \mid 1stOrd_k) > 0$ $i \in \{1, ..., n\}$ $j \in \{1, ..., m\}$ $k \in \{\text{Truth, Error, AverteError, MissedTruth}\}$
(2)	Determine the expected utility EU_{ind} of all individual second-order outcomes, given the respective first-order outcome, based on the second-order outcomes' dependent probability P and the second-order outcomes' utility U.	$EU_{\text{ind}}(2ndOrdSci_i \mid 1stOrd_k) = U(2ndOrdSci_i \cdot P(2ndOrdSci_i \mid 1stOrd_k) + U(\neg 2ndOrdSci_i) \cdot P(2ndOrdSci_i \mid 1stOrd_k)$ $EU_{\text{ind}}(2ndOrdExt_j \mid 1stOrd_k) = U(2ndOrdExt_j \cdot P(2ndOrdExt_i \mid 1stOrd_k) + U(\neg 2ndOrdExt_j) \cdot P(\neg 2ndOrdExt_j \mid 1stOrd_k)$
(3)	Aggregate the individual expected utilities into the aggregated expected utility EU_{agg} of the scientific and extra-scientific second-order outcomes for each first-order outcome.	$EU_{\text{agg}}\left(2ndOrdSci \mid 1stOrd_k\right) = \sum_{i=1}^{n} EU_{\text{ind}}\left(2ndOrdSci_i \mid 1stOrd_k\right)$ $EU_{\text{agg}}\left(2ndOrdExt \mid 1stOrd_k\right) = \sum_{j=1}^{m} EU_{\text{ind}}\left(2ndOrdExt_j \mid 1stOrd_k\right)$
(4)	Determine a setting for the trade-off parameter T.	Use only scientific utilities iff $T = 1$ Use only extrascientific utilities iff $T = 0$ Use both scientific and extrascientific utilities iff $1 > T > 0$
(5)	Determine the T-weighted expected utility EU_{T} of each first-order outcome.	$EU_{\text{T}}(1stOrd_k) = EU_{\text{agg}}(2ndOrdSci \mid 1stOrd_k \cdot T + EU_{\text{agg}}(2ndOrdExt \mid 1stOrd_k) \cdot (1 - T)$
(6)	Determine the probability p that performing D (PerfD) will imply true results.	$1 > p > 0$ $p = P(Truth \mid PerfD) = P(MissedTruth \mid \neg PerfD)$ $1 - p = P(Error \mid PerfD) = P(AvertedError \mid \neg PerfD)$
(7)	Determine the total expected utility EU_{total} of the decision options perform D (PerfD) and not perform D ($\neg$PerfD), given p.	$EU_{\text{total}}(PerfD) = EU_{\text{T}}(Truth) \cdot p + EU_{\text{T}}(Error) \cdot (1 - p)$ $EU_{\text{total}}(\neg PerfD) = EU_{\text{T}}(MissedTruth) \cdot p + EU_{\text{T}}(AvertedError) \cdot (1 - p)$
(8)	Determine the evidential threshold t.	$t = EU_{\text{total}}(PerfD) = EU_{\text{total}}(\neg PerfD)$ $t = \frac{EU_T(AvertedError) - EU_T(Error)}{EU_T(AvertedError) - EU_T(Error) + EU_T(Truth) - EU_T(MissedTruth)}$

4. *Challenging the challenger: Does inductive risk really refute value-freedom?*

4.1. Does AIR refute $\mathrm{VFI}_{\mathrm{desc}}$?

In this section, I use the technical basis developed above to discuss AIR's challenge of value-freedom. I first discuss whether the agent *can* come to a conclusion about D without considering extra-scientific utilities, and then discuss whether, and under what circumstances, the agent *should* use extra-scientific utilities. Starting with the first question, there seems to be an obvious problem with the inductive risk story: if the extent to which extra-scientific utilities influence D is regulated by T, and if $T = 1$ is a possible setting, then the claim that value-ladenness is inevitable seems to be trivially false. However, there may still be ways in which D could be value-laden even under $T = 1$:

(i) Non-epistemic values could be hidden in the scientific utilities.
(ii) Non-epistemic values could be hidden in p.
(iii) Non-epistemic values could be hidden in T.

As it turns out, all three claims may be true, but not in a sense that threatens $\mathrm{VFI}_{\mathrm{desc}}$ in our decision setting. Regarding (i) and (ii), a strategic and a substantial point should be considered. The strategic point is that proponents of AIR are not well advised to focus on (i) or (ii), as this would undermine much of AIR's appeal. If it is true that AIR is philosophically interesting because it challenges VFI even under "VFI-friendly" assumptions, supporters of AIR should not claim that values are basically *everywhere*, but rather focus on the specific way in which values influence D via the evidential threshold t. The trouble is that inductive risk is irrelevant in (i) and (ii), as even a decision that is highly certain and non-risky for extra-scientific goods —e.g. accepting the statement that cold and salty water sinks to deeper ocean layers— will be value-laden if non-epistemic values were hidden in p or in scientific utilities such as explanatory power. This does not go well with the inductive risk narrative, which fundamentally depends on *relevant* uncertainty (Betz, 2013) and on *identifiable* causal effects on extra-scientific goods. Also, (i) is effectively a rejection of $\mathrm{VFI}\text{-}\mathrm{R}_1$ (the restriction that VFI only refers to non-epistemic values). Of course, this does not mean that (i) or (ii) are false; rather, it means that the aspiration to refute VFI even under conditions that are favorable for value-freedom cannot be maintained. From a strategic point of view, those who believe that AIR constitutes one of the strongest challenges of value-freedom should therefore not attempt to capitalize on (i) and (ii).

But there is a more substantial case against (i) and (ii). With respect to (i), we should clarify what exactly the agent maximizes under $T = 1$: D's *conditional* utility for a given understanding of science, or D's *unconditional* utility for some transcendent idea of "science as such"? Clearly, it is the former. Understandings of science change over time, and even at a given point in time there may be more than one definition of good science (just think of the current debate among data scientists and statisticians about whether models should rather be interpretable or accurate, Hassani *et al.,* 2021). As "science as such" is an ill-defined term, the agent can only maximize D's utility for a *given* version of science. D will thus have a different utility in a version of science that favors, say, simplicity, than in one that favors heterogeneity (Longino, 2008). But this can be said about *any* decision. For instance, a person who aims to maximize private wealth will make different choices if "wealth" refers only to financial resources than when it also includes resources such as time.

And surely, determining whether time counts as wealth is a value-judgement. However, once the general goal is sufficiently defined, judgements about a choice's utility are rather instrumental than genuinely normative. We may thus grant that the *general* aims of science involve non-epistemic judgements (Kitcher, 2001; Bueter, 2015) and still maintain that the agent need not make such judgements in a *specific* decision.

The same is true for (ii). We may grant that p depends on the kinds of truths that science is supposed to find; but this does not mean that, once these expectations are set, the agent needs to consider non-epistemic values in any specific decision. For instance, if p were to represent the probability with which a climate model will produce trustworthy estimations of climate impacts, then p depends *inter alia* on what counts as an "impact" (De Melo-Martín & Intemann, 2016). This, in turn, depends on "judgments about what goods are worth protecting" (*ibid.,* p. 514). However, once this has been determined, the agent need not make any further value-judgement when determining p. Hence, even if (i) and (ii) were true, they would not refute VFI_{desc} in the decision setting under consideration.

So what about (iii)? We may here think of an argument that shows that $T = 1$ is itself a value-judgement. One promising basis for such an argument could be Jürgen Habermas' (1979; 1998) speech act theory. Every communicative act, Habermas argued, presupposes certain validity claims, one of which he called *moral rightness*. For instance, if we order a non-vegetarian meal in a restaurant, we implicitly claim that eating animals is permissible, irrespective of whether we considered this ethical claim in our actual decision process. Yet if a speech act such as "I would like to order the steak" entails "it is morally acceptable to eat animals", then the act of performing or not performing D under $T = 1$ seems to entail "it is morally acceptable to disregard D's effects on extra-scientific goods". As this seems to be a non-epistemic value-judgement, D seems to be value-laden even under $T = 1$.

Now, whether or not we accept this argument depends on what we think the propositional content of the entailed normative claim is. One might argue that the entailed claim is that it is acceptable for scientists to cause D's specific extra-scientific consequences. For instance, if a model choice could affect the regulation of a given toxicant, then $T = 1$ seems to entail that it is permissible to cause exactly the effects that this regulation may have—say, an increase in cancer rates. But this interpretation is misleading. First, $T = 1$ clearly does not presuppose that it is permissible to *cause* such consequences, but rather that it is permissible to *ignore* them. There is a difference between ignoring something and bringing something about. Funding agencies, for instance, may rightfully ignore whether a rejection hampers an applicant's career, but they may not intentionally cause such harm. Second, the entailed claim is *unspecific*, i.e. it refers to *any* extra-scientific consequence. The consequences could therefore change without changing D. This is very different to the restaurant example, where varied consequences can actually change the decision (imagine the choice harmed not cattle but, e.g., dogs). Third, and most importantly, the entailed claim is *redundant*: that it is acceptable to ignore D's extra-scientific consequences simply means that it is acceptable to make value-free scientific choices. Contrary to the restaurant example, this is not an independent proposition, but rather a trivial implication of VFI_{norm} —any norm presupposes that it is permissible to observe the norm ("ought" trivially implies "may"). As the claim entailed by $T = 1$ does not contain anything that was not already obvious, this type of value-judgement is uninteresting in our context. We must thus conclude that our agent may indeed presuppose non-epistemic judgements under $T = 1$, but not in a sense that undermines VFI_{desc}.

4.2. Does AIR refute $\mathrm{VFI}_{\mathrm{norm}}$?

The above conclusion is consistent with other authors (Betz, 2013; De Melo-Martin & Intemann, 2016; Steel, 2016a) who, with different arguments, have also claimed that AIR does not refute $\mathrm{VFI}_{\mathrm{desc}}$. Let us therefore see whether AIR is more successful in showing that the ideal agent *should* use non-epistemic values. Regarding this question, we need to consider two cases: one case where $T = 1$ implies equal (or at least very similar) expected scientific utilities for both decision options (case A); and another case where $T = 1$ implies a relevant difference between the two options, such that either performing or not performing D scores higher in expected scientific utilities (case B):

Case A $EU_{\mathrm{total}}(PerfD, T = 1) = EU_{\mathrm{total}}(\neg PerfD, T = 1)$
Case B $EU_{\mathrm{total}}(PerfD, T = 1) \neq EU_{\mathrm{total}}(\neg PerfD, T = 1)$

It turns out that case A provides much stronger grounds for attacking $\mathrm{VFI}_{\mathrm{norm}}$ than case B. Case A describes a state of *epistemic indifference*, i.e. a situation in which both decision options are equally promising regarding their desired scientific effects. The agent can therefore pursue her primary goal —the advancement of science—equally well by performing or by not performing D. In order to resolve the indifference, the agent has two options at her disposal: either she leaves $T = 1$ unchanged and "simply rolls a die" (Betz, 2013, p. 210); or she decreases the T-setting ($T < 1$) to a level where one decision option scores higher in total expected utilities than the other. In such a situation, it seems obvious that the agent should not randomize the choice, but rather decrease T. The striking reason is that *the surplus in expected extra-scientific utilities does not come at the expense of the expected scientific utilities*. If the agent can maximize both types of utilities at the same time, it is highly implausible that she should jeopardize the extra benefit by rolling a dice. Not only is it *irrational* to reject the raise in total expected utilities, it is also *blameworthy*, as failing to do good when it comes without costs is inappropriate even for an ideal epistemic agent. After all, the agent's commitment to scientific aims does not justify moral indifference, as long as the moral aims are compatible with the primary aim. Situations of epistemic indifference hence constitute a strong case against $\mathrm{VFI}_{\mathrm{norm}}$.

The idea that non-epistemic values should work as "tie breakers" to resolve epistemic indifference has been proposed by others (Steel, 2010; Steel & Whyte, 2012; Winsberg, 2012). Yet, it is important to see what exactly this means. "Tie breaker" situations have sometimes been described as "cases where hypotheses score equally well with respect to the evidence" (Magnus, 2018, p. 415, see also Intemann, 2005 p. 1007; Brown, 2013, p. 832). From a decision-theoretical perspective, however, this is only half true. Evidential support, i.e. p, is only *one* parameter that influences the expected scientific utility of a decision option; besides p, the agent must also consider U (the utility of D's consequences) and P (the dependent probability that these consequences actually occur). For instance, if two options are equally well supported by the evidence, but one option scores higher in U and P (e.g. because it will very likely have very positive impacts on future research), then the expected scientific utilities of the two options will diverge. The agent can therefore have a strong preference despite an identical p (Wilholt, 2009). Hence, contrary to some interpretations of the "tie breaker" thesis, equal evidential support alone does not constitute epistemic indifference. Irrespective of the interpretation, however, the "tie breaker" thesis expresses a valid idea: that even the "perfect scientist" should consider non-epistemic values if she can do so without compromising her scientific preferences.

Some authors, however, argue that non-epistemic values should also be considered in case B, i.e. in a scenario where the agent has a clear epistemic preference (Brown, 2013; Elliott & McKaughan, 2014; Intemann, 2015; De Melo-Martin & Intemann, 2016). While I agree that this may (at least sometimes) be plausible in actual science, I disagree that such an argument can be made for Rudner's "perfect scientist". The problem is that, contrary to case A, adopting $T < 1$ in case B *can* be scientifically detrimental. This can occur when the expected scientific and extra-scientific utilities pull into opposing directions. Imagine a situation where the introduction of a new model may be highly beneficial for the future development of a given research area, e.g. because it eliminates existing inconsistencies or enables new types of questions; yet this model may also make the research field less applicable to real-world problems, e.g. because the model's practical implications are ambiguous or because it generates data that are irrelevant for real-world decisions. It is hard to see why, in such a situation, the "perfect scientist" should disregard the scientific benefits and favor the extra-scientific benefits instead. After all, a crucual part of what it means to be a "perfect scientist" is exactly this: to prioritize the advancement of science. Choosing an option that may be scientifically detrimental is clearly incompatible with this preference. Hence, while AIR is strong in case A, it fails in case B.

Let me now discuss three questions that immediately emerge from the above considerations:

1. I have argued that AIR succeeds only in case A, i.e. in a scenario where the expected scientific utilities of performing and not performing D are identical. However, this scenario seems to be rather untypical. We thus have to ask how relevant AIR's success against VFI$_{\text{norm}}$ really is.
2. I have argued that AIR does not succeed in case B, as the "perfect scientist" cannot favor extras-scientific over scientific benefits. Yet, this seems to presuppose that scientific and extra-scientific utilities imply opposing decisions. This raises the question how the agent should act when both types of utilities pull into the same (rather than the opposite) direction.
3. I have argued that the agent cannot jeopardize her scientific preferences without ceasing to be a "perfect scientist". At the same time, I have said that this may not necessarily be so in actual science. The question is thus how relevant the above reasoning is for actual science.

I discuss the first two questions here and consider the third question in the conclusion. Regarding the first question, I concede that an exact convergence of expected scientific utilities (case A) may seem untypical, thus creating an impression of irrelevance. Yet, this impression is false. First, even if *exact* convergences were untypical, utilities may well be *approximately* equal. Which option the agent choses would then be rather unimportant for science. Given this lack of significance, we can plausibly treat approximate and exact epistemic indifference analogously, which broadens the set of scenarios covered by case A. Second, there are contexts where epistemic indifference is not uncommon at all, namely when a research field is still young. In avant-garde science, it is often unclear which option will yield higher scientific benefits, as the field's future development is highly uncertain. Third, the impression that epistemic indifference is untypical rests on the assumption that p represents a point prediction. However, as Wendy Parker (2014) has argued, "one must know a lot to be a position to say with justification that the probability (degree of belief) that should be assigned to a hypothesis is 0.38 rather than 0.37 or 0.39" (*ibid.,* p. 27). Whenever the evidence is scarce, inconsistent, or ambiguous, *p* will plausibly be expressed as an interval, say [0.3, 0.4] rather than 0.38. Note that this holds even for the "perfect scientist", who is just as confined to the currently available evidence as actual scientists are. Yet, as soon as *p* comes as an interval, epistemic indifference is more likely. Case A, and hence AIR's refutation of VFI_{norm}, is thus more relevant than it may seem at first sight.

Regarding the second question, note that case B comprises two different scenarios: one where *D* is expected to be beneficial for science, but detrimental for extra-scientific goods; and one where *D* promises scientific and extra-scientific benefits at the same time. Critics of value-freedom tend to focus on the first scenario, where there is a trade-off between scientific and extra-scientific considerations (Douglas, 2000; Douglas, 2009; Elliott & McKaughan, 2014). As noted by Steel (2016b), however, epistemic and non-epistemic values need not necessarily pull into opposite directions. For instance, scientific simplicity can be good for both extra-scientific decision-making (by providing quick results) and for science (by reducing complexity) *(ibid.).* Interestingly, this non-trade-off scenario is irrelevant and relevant at the same time. It is irrelevant as extra-scientific utilities do not change *D* if they merely reconfirm an already existing preference. Also, remember that the version of VFI under consideration is restricted to judgements that actually change a decision, e.g. from using to not using a model (VFI-R_3). Unless one rejects VFI-R_3, it thus follows that the agent is permitted, although not obliged, to consider non-epistemic values in a non-trade-off scenario. Of course, whether or not she does so is effectively irrelevant, at least from a consequentialist perspective (AIR is obviously an instance of consequentialist ethics). Yet, the non-trade-off scenario is relevant in a different sense. Inductive risk narratives can create the impression that there is an intrinsic conflict between doing what is good for science and doing what is good from an ethical perspective. While such conflicts exist, they are clearly contextual, i.e. they may occur or not. The relevance of the non-trade-off scenario is thus that it shows that $T = 1$ need not necessarily imply ethically undesirable results.

5. *Can AIR avoid prescription and wishful thinking?*

5.1. APr's charge of prescriptiveness

The previous section has argued that even the "perfect scientist" should sometimes use non-epistemic values. However, I have also argued that Holman & Wilholt (2022) and others are right to claim that we should not tear down "Weber's fence" *(ibid.)* without addressing VFI's concerns. In this section, I will thus discuss how *D* can be value-laden, yet not prescriptive and logically fallacious. I start by discussing the argument from prescription (APr)[9]. APr's main claim reads (see *sect. 2.2*):

[9] This part benefited from discussions with members of the Consortium for Science, Policy and Outcomes at Arizona State University and the Mercator Research Institute on Global Commons and Climate Change, particularly Martin Kowarsch.

APr (2) If *D* is value-laden, it constitutes a *relevant*, *external* and *normative* constraint of extra-scientific agents' decision space (which violates their autonomy).

As said before, critics of VFI can either reject this claim, e.g. by arguing that *D* does not really constrain extra-scientific agents[10] or by arguing that such constraints are actually legitimate[11]. Alternatively, they can accept APr in general, but argue that—if the right measures are taken—*D* does not fulfill at least one of APr's conditions (relevance, externality, normativity). As previously said, I only discuss the latter strategy. Two conditions are promising for this strategy: relevance and externality. A constraint is *relevant* if it removes options from an agent's decision space that the agent may actually take interest in; a constraint is *external* if the agent did not, explicitly or implicitly, consent to the constraint. The third condition, normativity, refers to *D*'s value-ladenness. This condition makes sure that those scientific choices that are not value-laden in AIR's sense (e.g. accepting the statement "coal burns", Betz, 2013, p. 21) cannot qualify as prescriptive. However, as we are here interested in cases were *D* includes extra-scientific utilities, the normativity condition is obviously fulfilled.

So what about relevance? To illustrate this condition, consider Rudner's example of the Manhattan Project. Before conducting their detonation experiments, the involved scientists had to accept "the hypothesis that no uncontrollable pervasive chain reaction would occur" (1953, p. 2-3). Assuming that they considered extra-scientific utilities, we can take it that *U* was high for preventing the nuclear accident, and low for causing it. Was this judgement prescriptive? Obviously not. As none of the potentially affected stakeholders can have preferred the accident, the judgement did not restrict anyone's freedom of choice[12].

[10] Critics of the so-called "linear model of expertise" (Jasanoff & Wynne, 1998) have argued that "the influence of science on policy is [not] strong and deterministic" (Beck 2011, p. 298). In their view, actual science-policy processes show that "[i]t would be an exaggeration to state that science [is] driving this process" (Grundmann & Rödder 2019, p. 4). This may undermine APr's claim that science constrains real-world decisions. But this reasoning is implausible. While both APr and AIR assume that *D* influences extra-scientific agents, neither of them presupposes determinism. The false impression stems from confusing *D*'s first- and second-order outcomes. It then seems that, if *D* implies a true or a false result, certain extra-scientific effects must occur. However, since extra-scientific effects are mediated by various factors (individual reflection, public debate, political compromise etc.), this is clearly false. The notion of "decision constraint" should hence be interpreted probabilistically (via *P*), such that *D* makes it more or less *likely* that extra-scientific agents make certain decisions.

[11] APr presupposes some commitment to liberal democracy. However, some authors argue that liberal freedom is less important than other goods such as the prevention of environmental disasters. James Lovelock, e.g., has famously argued that climate change may make it necessary "to put democracy on hold for a while" (*The Guardian*, March 29th, 2010) (see also Shearman & Smith, 2007; Beeson, 2010). Supporters of this reasoning could hence argue that the principle of autonomy is too weak to sustain APr. However, such claims typically presuppose some argument from emergency. Even if such considerations were successful, they would thus undermine APr only in exceptional cases.

[12] Stephen John (2019) has recently suggested a notion of "value-aptness" that seems to point into a similar direction (although John refers to the *communication* of scientific findings, not the making of the scientific decision as such). John argues that value-laden communication by scientists does not violate the audience's autonomy if the underlying values are compatible with the values held by the audience. An implication of John's "value-apt ideal" would thus be that the employed values no longer constitute a relevant decision constraint. As discussed below, however, avoiding the relevance condition is only one way to avoid illegitimate prescription.

One way to avoid prescriptiveness is thus to use only uncontroversial values. Yet, this approach has limits. More often than not, there will be no consensus on extra-scientific utilities. Even in Rudner's example, the consensus comprises only the consequences of error and averted error, while the extra-scientific effects of truth (building the atomic bomb) and missed truth (not building the atomic bomb) are clearly controversial. Furthermore, scientists may assume a consensus where there is none. This problem can be mitigated, e.g. by conducting stakeholder surveys and by using scenario approaches (Edenhofer & Kowarsch, 2015) that include "solution pathways for any of the major attitudes that can be found in society" (Held 2011, p. 115). Note, however, that this will not always be possible. While, e.g., climate researchers need not commit themselves to only one climate projection, the scientists in Rudner's example could either conduct or not conduct the experiment, but not both. Also, surveys and scenario approaches are again subject to inductive risk (choice of sample sizes, definition of scenarios etc.), thus repeating the prescriptiveness issue on a higher level. Attempting to use only uncontroversial values may thus not always be successful.

Let us therefore consider APr's externality condition. Critics of VFI have offered two strategies to avoid externality, the *transparency* and the *democratic* approach. In the former, scientists determine extra-scientific utilities by themselves, but communicate their choices transparently (Rudner, 1953, p. 6; Douglas, 2009, chap. 4; Elliott & McKaughan, 2014). Extra-scientific agents can then scrutinize these choices and, should they disagree, simply ignore the respective study. This protects their freedom of choice. The problem with this approach is that it views autonomy as an *ex post* capacity, i.e. as the right to reject or accept a choice that has already been made. Call this autonomy *qua* recipient. Moreover, it seems implausible that extra-scientific agents can easily "backtrack" value-judgments, as Elliott & McKaughan (2014) have argued, "and adopt their own alternative assessments and conclusions" (*ibid.,* p. 16). For this to be possible, the implications of these judgements must be deducible just by extrapolation. In most cases, however, extra-scientific agents will only have rough clues what a study would have looked like if, say, a different model would have been used. Thus, while the transparency approach has the virtue of practicality, it promotes only a weak form of autonomy.

In contrast, the democratic approach promotes an *ex ante* notion of autonomy, where stakeholders are consulted before the respective judgements are made. Call this autonomy *qua* author. Clearly, being the author of a value-judgement allows for more autonomy than being its recipient. Such authorship may be realized in various ways. The most ambitious forms are iterative (steady consultations rather than one-time interactions), direct (involving ordinary citizens rather than professional representatives), deliberative (consensus-oriented and rational) and inclusive (involving all affected parties) (Douglas, 2005; Douglas, 2009, chap. 8; Brown, 2009, chap. 9-10; Kitcher, 2011; Kowarsch *et al.,* 2016). Citizen panels are a good approximation to this ideal (Davies *et al.,* 2005; Tomblin *et al.,* 2017). The trouble is that such formats are slow and costly, thus diminishing resources that could be used for other scientific and social projects. Moreover, they may be suited to discuss the general normative issues of a research field, but not the numerous, highly technical decisions that must be made in an individual study. These problems can be mitigated, e.g. by using less iterative or less direct forms of participation. Participation could also be restricted to a higher institutional level, such that extra-scientific agents contribute to the production of general guidelines, but not to their application in specific studies (Steel, 2016a). But this does not come without downsides either. The less stakeholders participate in making the actual judgement, the less can they be seen as its authors; also, the "downscaling" of general guidelines to specific scientific choices will again be subject to inductive risk. Thus, while ambitious variants of the democratic approach create more autonomy at the expense of practicality, the less ambitious variants are more practical, but allow for less autonomy.

What does this mean for APr? I would argue that if there were only one way to address prescriptiveness, this would undermine AIR's claim that *D* should sometimes involve extra-scientific utilities. However, while none of the above strategies is satisfying on its own, in conjunction they provide a feasible set of means to legitimize non-epistemic judgements in science. In some cases, it will be possible to circumvent the relevance condition by making uncontroversial value-judgements, or by using scenario sets that represent the spectrum of existing value commitments (Edenhofer & Kowarsch, 2015). In the remaining cases, there are ways to avoid the externality condition. We may here think of a multi-layer system (Steel, 2016a), where stakeholders contribute intensely to those studies that are closely entangled with extra-scientific decisions, e.g. advisory reports or technology assessments (Sclove, 2011; Garard & Kowarsch, 2017), but contribute to everyday science only on a higher level (e.g. via general guidelines, see Steel, 2016a). If additional judgements are needed in a concrete study, e.g. to interpret the general guidelines or to make choices that are not covered by the guidelines, scientists can use the transparency approach to create some autonomy *qua* recipient. Therefore, while APr is right to emphasize the danger of prescriptiveness, this danger can be countered. Suitable measures against prescriptiveness exist, and as long as these are taken, APr does not refute AIR.

5.2. AWT's charge of wishful thinking

The second concern behind VFI, the argument from wishful thinking (AWT), holds that *D* is logically fallacious if it includes extra-scientific utilities. The main claim reads (see sect. 2.2):

> AWT (2) If *D* is value-laden, *D* represents a *direct*, *non-vacuous* and *not semantically entailed* inference from an ought-claim to an is-claim (which violates no-is-from-ought).

Similar to APr, AWT hinges on three conditions. I restrict my discussion to these conditions, thus presuming that, if they are jointly fulfilled, *D* is indeed fallacious. Taking up meta-ethical work (Schurz, 1997; D'Arms & Jacobson, 2000; Pidgen, 2010; Pidgen, 2016), I understand AWT's conditions as follows: an ought-is inference is *direct* if the propositional source of the descriptive conclusion lies exclusively in a set of normative premises (e.g. "x should be the case, therefore x is the case"); an ought-is inference is *non-vacuous* if the descriptive conclusion is a substantial or non-arbitrary implication of a set of normative premises (an example of a vacuous inference is "x should be the case, therefore x should be the case or x is the case") (Prior, 1960; Pidgen, 2010); an ought-is inference is *not semantically entailed* if the descriptive conclusion is not hidden in the set of normative premises (an example of a semantically entailed inference is "x should be done, therefore x can be done") (Searle, 1964; Pidgen, 2016).

To see whether *D* fulfills these conditions, imagine that the agent concludes that it is justified to use a certain model, and that this conclusion is value-laden in the previously discussed sense ($T < 1$). The agent's set of premises would then comprise four types of elements:

D_1 A descriptive premise that specifies the probability *p* that using the model leads to truth.

D_2 A set of descriptive premises that specify the model's potential scientific and extra-scientific consequences, as well as these consequences' dependent probability *P*.

N_1 A set of normative premises that specify the utility *U* of the model's scientific consequences.

N_2 A set of normative premises that specify the utility *U* of the model's extra-scientific consequences.

It may be objected that the model choice cannot be subject to no-is-from-ought, as "it is justified to use the model" is not a descriptive conclusion. However, from the perspective of the "perfect scientist" this is just short for "before the background of the available evidence and the expected consequences, using the model promises more or at least equal benefits for science than not using the model". If we now accept AWT's premise that science should aspire truth (AWT-1), the conclusion commits the agent to the descriptive claim that the model helps to find scientific truths. Note that this is compatible with the idea that different versions of science may aspire different kinds of truth. In some contexts it may, e.g., be rational to prefer a less precise over a more precise model (Elliott & McKaughan, 2014). But this neither means that such a choice does not benefit science (Steel, 2016b), nor that the less precise model is not supposed to be truth-conducive. Rather, the model is supposed to help find exactly the kind of scientific truths that are deemed relevant in a given context.

So is the model choice fallacious in AWT's sense? As it turns out, two of AWT's conditions are fulfilled: the inference is non-vacuous because the descriptive proposition is not arbitrarily attached to the conclusion (as in "the model is ethically good, therefore the model is ethically good or the model is truth-conducive"); the inference is also not semantically entailed, i.e. not derived from an implicit descriptive content of a normative premises (as in "the model should be used, therefore the model can be used"). The third condition, *directness*, is more ambiguous. We may argue that directness is not given because the conclusion's descriptive content originates not from N_2 (the set of premises that represent the non-epistemic value-judgement), but from D_1 (the premise that characterizes the evidence). Douglas (2000; 2008; 2009) has argued into this direction. The "virtue of truth-seeking" (2008, p. 10), she holds, precludes non-epistemic values from acting "as reasons in themselves to accept a claim" (2009, p. 96); rather, their role is to "weigh the importance of uncertainty" *(ibid.)*. On this account, the inference is not an instance of "the model is ethically good, therefore the model is truth-conducive" because the model's truth-conduciveness is inferred from a descriptive rather than from a normative premise (Kevin Elliott has called this the "logical interpretation" of Douglas' approach, see Elliott, 2013, p. 377).

I see two problems with this defense against AWT. First, it remains possible that non-epistemic values only "weigh the importance of uncertainty" (Douglas, 2009, p. 96) and yet dominate the evidence. This can occur when the expected extra-scientific utilities are distributed very unevenly, e.g. when the expected error damages are very high, while the expected truth benefits are very low (or vice versa). In such a scenario, utility distributions are conceivable where the agent *never* (or *always*) uses the model, irrespective of how well (or poorly) the model is supported by the evidence. Metaphorically speaking: if only the scale that weighs the evidence is sensitive or insensitive enough, any amount of evidence will be sufficiently "heavy" or "light" to justify a choice. Yet, if an evidential threshold is never (or always) met, D_1 is obviously irrelevant. As the only plausible source for the conclusion's descriptive content would then be N_2, AWT's directness condition can be fulfilled even if non-epistemic values only "weigh the importance of uncertainty" (ibid). Secondly, Douglas' approach allows for *inverse preference orders*, i.e. scenarios where the agent prefers error over truth, and missed truth over averted error. This is because not all truths are extra-scientifically good, while not all errors are extra-scientifically bad. Placebo drugs, e.g., can have positive effects not although, but exactly because they are used on the basis of a false belief. However, if the agent prefers a convenient error over an inconvenient truth, she will use the model when *p* is low and dismiss it when *p* is high. Such a paradox notion of the evidence would undermine the claim that the conclusion stems from D_1. The possibility of inverse preference orders thus provides further support for AWT.

It could be objected that both issues, polarized utility distributions and inverse preference orders, are untypical. However, the claim is not that these issues occur often, but that, if they occur, they cannot be prevented by restricting non-epistemic values to determining evidential thresholds. Furthermore, the first issue can be relevant even if an evidential threshold is not *conceptually* impossible to meet or miss; it can suffice that a threshold is *practically* never met or missed in a given area of study. We may call such choices *material* rather than formal ought-is fallacies, as D_1 is still present in the set of premises, but practically irrelevant. One might then object that such choices can still be acceptable, as science has legitimate goals besides truth (Elliott, 2013; Elliott & McKaughan, 2014). As argued before, however, goals such as applicability or timeliness are not unrelated to truth, but qualify the kinds of truths that are aspired in a given context. Yet, I agree with the objection in one respect: some scientific choices may be ethically impermissible, irrespective of how unlikely an error is. But this does not undermine my point that the mentioned issues are problematic. For one, this reasoning seems inapplicable to the other scenarios (e.g. truths that are so desirable that a choice is always made, or truths that are so undesirable that they are valued less than errors). For another, if a choice is morally impermissible, it should be seen as a boundary condition, similar to ethical norms in human trials. However, such norms are not subject to AWT or VFI in the first place (via VFI-R_2). Thus, while ethics may indeed sometimes trump truth-seeking, this does not help against AWT.

Another way to address AWT is to return to the previously discussed cases (*sect. 4.2*). I have argued that the "perfect scientist" should consider non-epistemic values only to resolve epistemic indifference (case A). Now, if "epistemic indifference" means that both options promise equal scientific benefits, and if "scientific benefit" means that an option helps to find scientific truths, then the conclusion's descriptive content is clearly not derived from N_2. As non-epistemic values only decide the choice between options whose truth-conduciveness has already been established, AWT's directness condition is not fulfilled in case A. Additionally, it seems implausible that scientific utilities could force evidentially unsupported choices. It is hard to see how an unattainable or unmissable evidential threshold could benefit science, as this would either add falsehoods to the body of scientific beliefs or make it impossible to find scientific truths. The same holds true for inverted preference orders. Even if some errors may be scientifically fruitful, the *deliberate* adoption of a false belief seems incompatible with the truth-seeking nature of science (note that we are not taking about simplifications or counterfactual assumptions here, as long as these are used to find scientific truths; inverted preferences, on the other hand, mean that a choice is made to purposely generate scientific falsehoods). As the directness condition is not fulfilled in case A, AWT does not succeed in this case.

So what about scenarios where the agent has a clear epistemic preference (case B)? I have argued that, at least for the "perfect scientist", $\mathrm{VFI}_{\mathrm{norm}}$ remains valid in case B. However, this is because the "perfect scientist" cannot trade scientific for extra-scientific benefits, not because the use of non-epistemic values is *necessarily* fallacious in case B. I have already discussed that scientific and extra-scientific expected utilities may well pull into the same direction. Similar to case A, I would argue that such non-trade-off scenarios do not constitute ought-is fallacies, as non-epistemic values merely confirm an independent epistemic preference. AWT's directness conditions is hence not fulfilled in these scenarios. Nevertheless, the condition can indeed be fulfilled in the other scenario of case B, namely when scientific and extra-scientific utilities pull into opposite directions. If non-epistemic values are used in such trade-off scenarios, and if they change a decision from, say, using a model to not using a model, then the conclusion is clearly derived from N_2. Note again that such ethics-driven decisions may sometimes be acceptable in actual science. As said before, however, they should then be treated as ethical boundary conditions. Hence, while AWT would indeed succeed if ethics-driven decisions are interpreted as descriptive conclusions, AWT's charge of wishful thinking can be averted by seeing them as what they are: moral rather than genuinely scientific choices.

6. *Conclusion: idealized* versus *actual science*

Inductive risk is widely recognized as "[o]ne of the most important reasons for thinking that non-epistemic values can play a legitimate role in scientific reasoning" (Elliott & Steel, 2017, p. 6). A great part of AIR's appeal lies in its promise to refute VFI even under idealized assumptions. As Rudner has put it, AIR claims that VFI fails even for the "perfect scientist" or "the scientist *qua* scientist" (1953, p. 2). Not only is it "more surprising in the ideal setting that scientists must make value judgments" (Steele, 2012, p. 895), such idealizations also capture the *in principle* nature of VFI (Weber, 1949; Popper, 1976; Koertge, 2000; Ruphy, 2006; Kitcher, 2011; Reiss & Sprenger, 2020).

Taking up this challenge, I have proposed a Bayesian framework that accounts for subjective probabilities, outcomes other than error (Wilholt, 2009; Wilholt, 2013), and the difference between first- and second-order outcomes. The approach also gives us a clearer decision rule than more classic takes on AIR and eases the old problem (Kuhn, 1977) that epistemic values can stand in tension with each other. Finally, the trade-off parameter T represents the balancing problem of epistemic versus non-epistemic values in a more fine-grained way than, e.g., the notion of values "trumping" (Elliott & McKaughan, 2014) each other. Using the idealized setting as a testing ground, I have argued that AIR does not refute $\mathrm{VFI}_{\mathrm{desc}}$. Regarding $\mathrm{VFI}_{\mathrm{norm}}$, I have argued that AIR fails whenever the agent has a clear epistemic preference (case B), but succeeds whenever the expected scientific utilities of the decision options converge (case A). I have argued that the notion of utility convergence goes beyond common versions of the "tie breaker" thesis (Intemann, 2005; Brown, 2013; Magnus, 2018), and that utility convergence is more typical than it may seem at first. Hence, while AIR's refutation of VFI is not *complete*, it still represents a powerful critique of value-freedom. This is further supported by the fact that two of the main concerns behind VFI, APr and AWT, can be countered by avoiding these arguments' conditions.

It may be argued that my decision setting is unrealistic and, hence, practically irrelevant. I disagree. In fact, many of its aspects are surprisingly realistic. First, inductive risk involves more than, e.g., the classic distinction between consumer versus producer risks (e.g. Carrier, 2011; Biddle & Leuschner, 2015). In reality, any given scientific choice can have many consequences; these will have different (dependent) probabilities and will occur not only in case of error (e.g. missing a truth is not an error, but can be ethically relevant). Second, the classic inductive risk heuristic "worse consequences = higher evidential thresholds" is not more, but less practical than the Bayesian reconstruction, as it neither contains a point of reference nor an idea of how exactly these thresholds should be determined. Third, it is quite realistic to interpret epistemic indifference as utility convergence, as a decision's scientific worth will *practically* also depend on the likelihood and the desirability of its scientific consequences. Fourth, it is very realistic to assume that probabilities are typically subjective in a Bayesian sense; also, probabilities will often be imprecise (Parker, 2014), which makes epistemic indifference, and hence my arguments regarding case A, highly relevant in actual science. This is also why the objection that scientific utilities may be unclear or disputed in practice does not speak against my reconstruction; this simply means that utilities can come as intervals as well, which in turn makes case A even more relevant. Finally, the balancing problem represented by T represents a quite practical issue, as scientists cannot make inductive risk decisions if the relative weight of the epistemic and non-epistemic values remains unclear.

Thinking this a bit further, I would argue that the idealized setting outlines something that may be called *epistemic legitimacy*—a set of rules that should govern the use of non-epistemic values in actual science. Similar to Steel (2010)[13], I contend that it should be the standard approach for scientists to use non-epistemic values only in cases of epistemic indifference. As we can see from the discussion of AWT, scientists should always favor truth over error and averted error over missed truth, irrespective of how desirable a decision's second-order outcomes are. Furthermore, scientists must carefully avoid illegitimate prescription, typically by applying a combination of the approaches presented in the discussion of APr. While I believe that scientists should normally not use non-epistemic values if they have a clear epistemic preference, I also concede that ethical concerns may *sometimes* outweigh scientific considerations (e.g. when the extra-scientific consequences are both very bad and very likely). In such cases, scientists should scrutinize whether there is a real trade-off, i.e. whether epistemic and non-epistemic values actually pull into different directions. If there is a real trade-off, scientists should use non-epistemic values only if the expected extra-scientific benefit is significantly higher than the expected scientific loss.

[13] Note that Steel (2010) interprets epistemic indifference in a different way, namely as a balance between epistemic values. As said before, however, epistemic indifference is better captured as a convergence of expected scientific utilities.

I would also argue that the relative weight that non-epistemic values can have in a trade-off scenario, i.e. *T*, should not be determined by individual scientists or research groups, but by codified guidelines (issued by, e.g., national academies or research associations). Most importantly, if scientists make decisions against an epistemic preference, they must communicate this as an *ethical* rather than a scientific choice.

Acknowledgements

I would like to thank Torsten Wilholt, Dietmar Hübner, Roel Visser, Hermann Held, and the members of the research training group GRK 2073 “Integrating Ethics and Epistemology of Scientific Research” for their support and helpful comments. I particularly wish to thank Benjamin Blanz for his generous support on the decision-theoretical part. This paper also benefited from discussions with members of the Consortium for Science, Policy and Outcomes at Arizona State University and the Mercator Research Institute on Global Commons and Climate Change, particularly Martin Kowarsch. Funding was provided by the German Research Foundation (DFG) (254954344/GRK2073) and the German Academic Exchange Service (DAAD).

REFERENCES

Beck, S. (2011). Moving beyond the linear model of expertise? IPCC and the test of adaptation. *Regional Environmental Change*, 11(2), 297-306.

Beeson, M. (2010). The coming of environmental authoritarianism. *Environmental Politics*, 19(2), 276-294.

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

Biddle, J. (2013). State of the field: Transient underdetermination and values in science. *Studies in History and Philosophy of Science*, 44, 124-133.

Biddle, J., & Kukla, R. (2017). The geography of epistemic risk. In K. Elliott, & T. Richards (Eds.). *Exploring Inductive Risk: Case Studies of Values in Science* (pp. 215-237). New York: Oxford University Press.

Biddle, J. & Leuschner, A. (2015). Climate skepticism and the manufacture of doubt: can dissent in science be epistemically detrimental? *European Journal for Philosophy of Science*, 5, 261-278.

Biddle, J., & Winsberg, E. (2010). Value judgements and the estimation of uncertainty in climate modelling. In P.D. Magnus, & J. Busch (Eds.). *New Waves in Philosophy of Science* (pp. 172-197). Basingstoke: Palgrave Macmillan.

Bright, L. (2018). Du Bois’ democratic defence of the value free ideal. *Synthese*, 95(5), 2227-2245.

Brown, M. B. (2009). *Science in democracy. Expertise, institutions, and representation*. Cambridge: MIT Press.

Brown, M. J. (2013). Values in science beyond underdetermination and inductive risk. *Philosophy of Science*, 80, 829-839.

Bueter, A. (2015). The irreducibility of value-freedom to theory assessment. *Studies in History and Philosophy of Science*, 49, 18-26.

Carrier, M. (2011), Knowledge, politics, and commerce: science under the pressure of practice. In M. Carrier & A. Nordmann (Eds.) *Science in the Context of Application* (pp. 11-30). Dordrecht: Springer.

Churchman, C. W. (1948). *Theory of Experimental Inference*. New York: Macmillan.

D’Arms, J. & Jacobson, D. (2000). The moralistic fallacy: on the ‘appropriateness’ of emotions. *Philosophy and Phenomenological Research*, 61(1), 65-90.

Davies, C., Wetherell, M., Barnett, E., & Seymour-Smith, S. (2005). *Opening the box. Evaluating the citizens council of NICE*. The Open University.

De Melo-Martín, I., & Intemann, K. (2016). The risk of using inductive risk to challenge the value-free Ideal. *Philosophy of Science*, 83, 500-520.

Dorato, M. (2004). Epistemic and nonepistemic values in science. In P. Machamer & G. Wolters (Eds.). *Science, Values, and Objectivity* (pp. 52-77). Pittsburgh: University of Pittsburgh Press.

Douglas, H. (2000). Inductive risk and values in science. *Philosophy of Science*, 67(4), 559-579.

Douglas, H. (2005). Inserting the public into science. In S. Maasen & P. Weingart (2005) (Eds.): *Democratization of Expertise? Exploring Novel Forms of Scientific Advice in Political Decision-Making* (pp. 153-169). Dordrecht: Springer.

Douglas, H. (2008). The role of values in expert reasoning. *Public Affairs Quarterly*, 22(1), 1-18.

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

Douglas, H. (2016). Values in science. In P. Humphreys (Ed.), *The Oxford Handbook of Philosophy of Science* (pp. 609-630). New York: Oxford University Press.

Douglas, H. (2017). Why inductive risk requires values in science. In K. Elliott & D. Steel (Eds.). *Current Controversies in Values and Science* (pp. 81-93). New York: Routledge.

Du Bois, W. E. B. (1935). *Black reconstruction in America*. New York: The Free Press.

Dupré, J. (2007). Fact and value. In H. Kincaid, J. Dupré & A. Wylie (Eds.). *Value-Free Science? Ideals and Illusions* (pp. 27-41). Oxford: Oxford University Press.

Edenhofer, O., & Kowarsch, M. (2015). Cartography of pathways: A new model for environmental policy assessments. *Environmental Science & Policy*, 51, 56-64.

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

Elliott, K. (2013). Douglas on values: From indirect roles to multiple goals. *Studies in History and Philosophy of Science,* 44, 375-383.

Elliott, K., & McKaughan, D. (2014). Nonepistemic values and the multiple goals of science. *Philosophy of Science*, 81, 1-21.

Elliott, K. & Steel, D. (2017). Introduction: Values and science: Current controversies. In K. Elliott & D. Steel (Eds.): *Current controversies in values and science* (pp. 1-11). New York: Routledge.

Garard, J. & Kowarsch, M. (2017). Objectives for stakeholder engagement in global environmental assessments. *Sustainability*, 9, 1571.

Grundmann, R. & Rödder, S. (2019). Sociological perspectives on Earth System Modeling. *Journal of Advances in Modeling Earth Systems*, 11, 3878-3892.

Haack, S. (2003). Knowledge and propaganda. Reflections of an old feminist. In C. Pinnick, N. Koertge & R. Almeder (Eds.). *Scrutinizing feminist epistemology. An examination of gender in science* (pp. 7-19). New Brunswick: Rutgers University Press.

Habermas, J. (1979). What is universal pragmatics? In *Communication and the Evolution of Society* (pp. 1-68). Boston: Beacon Press.

Habermas, J. (1998). Some further clarifications of the concept of communicative rationality. In *On the pragmatics of communication* (pp. 307-342). Cambridge: Polity Press.

Held, H. (2011). Dealing with uncertainty – From climate research to integrated assessment of policy options. In G. Gramelsberger & J. Feichter (Eds.). *Climate change and policy. The calculability of climate change and the challenge of uncertainty* (pp. 113-126). Berlin: Springer.

Hempel, C. (1965). Science and human values. In *Aspects of scientific explanation and other essays in the philosophy of science* (pp. 81-96). New York: The Free Press.

Hempel, C. (1981). Turns in the evolution of the problem of induction. *Synthese* 46, 389-404.

Holman, B. & Wilholt, T. (2022). The new demarcation problem. *Studies in History and Philosophy of Science*, 91, 211-220.

Hassani, H., Beneki, C., Silva, E. S., Vandeput, N., & Madsen, D. Ø. (2021). The science of statistics versus data science: What is the future?, *Technological Forecasting and Social Change*, 173, 121111.

Hoyningen-Huene, P. (2006). Context of discovery versus context of justification and Thomas Kuhn. In J. Schickore & F. Steinle (Eds.). *Revisiting discovery and justification* (pp. 119-131). Dordrecht: Springer.

Intemann, K. (2005). Feminism, underdetermination, and values in science. *Philosophy of science*, 72(5), 1001-1012.

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

James, W. (1912). The will to believe. In *The Will to Believe and other essays in popular philosophy* (pp. 1-31). London: Longmans, Green & Co.

Jasanoff, S., & Wynne, B. (1998). Science and decisionmaking. In S. Rayner & E. Malone (Eds.). *Human choice and climate change. Vol 1: The societal framework* (pp. 1-87). Ohio: Battelle Press.

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

John, S. (2019). Science, truth and dictatorship: Wishful thinking or wishful speaking?. *Studies in History and Philosophy of Science*, 78, 64-72.

Jones, O. (1999). Sex, culture, and the biology of rape: Toward explanation and prevention. *California Law Review*, 87(4), 827-941.

Kitcher, P. (2001). *Science, truth, and democracy*. New York: Oxford University Press.

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

Koertge, N. (2000). Science, values, and the values of science. *Philosophy of Science* (Supplement) 67, 45-57.

Kowarsch, M., Garard, J., Riousset, P., Lenzi, D., Dorsch, M., Knopf, B., Harrs, J., & Edenhofer, O. (2016). Scientific assessments to facilitate deliberative policy learning. *Palgrave Communications*, 2(1), 1-20.

Kuhn, T. (1977). Objectivity, value judgement, and theory choice. In *The essential tension. Selected studies in scientific tradition and change* (pp. 320-339). Chicago: University of Chicago Press.

Kuhn, T. (1962). *The structure of scientific revolutions*. Chicago: University of Chicago Press.

Lacey, H. (1999). *Is science value free? Values and scientific understanding*. London: Routledge.

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

Longino, H. (2008). Values, heuristics, and the politics of knowledge. In M. Carrier, D. Howard & J. Kourany (Eds.). *The challenge of the social and the pressure of the practice* (pp. 68-86). Pittsburgh: University of Pittsburgh Press.

Machamer, P., & Douglas, H. (1999). Cognitive and social values. *Science & Education*, 8(1), 45-54.

Magnus, P. D. (2018). Science, values, and the priority of evidence. *Logos & Episteme*, 9(4), 413-431.

McMullin, E. (1982). Values In science. *Proceedings of the Biennial Meeting of the Philosophy of Science Association*. Vol. Two: Symposia and Invited Papers, 3-28.

Parker, W. (2014). Values and uncertainties in climate prediction, revisited. *Studies in History and Philosophy of Science*, 46, 24-30.

Pidgen, C. (2010). On the triviality of Hume's law: a reply to Gerhard Schurz. In C. Pidgen (Ed.): *Hume on is and ought* (pp. 212-236). Hampshire: Palgrave Macmillan.

Pidgen, C. (2016). Hume on is and ought. In P. Russel (Ed.). *The Oxford Handbook of Hume* (pp. 401-415). Oxford: Oxford University Press.

Popper, K. (1976). The logic of the social sciences. In T. Adorno, H. Albert, R. Dahrendorf, J. Habermas, H. Pilot & K. Popper (Eds.). *The positivist dispute in German sociology* (pp. 87-104). London: Heinemann.

Prior, A. (1960). The autonomy of ethics. *Australasian Journal of Philosophy*, 38 (3), 199-206.

Proctor, R. (1991). *Value-free science? Purity and power in modern knowledge*. Cambridge: Harvard University Press.

Putnam, H. (2002). *The collapse of the fact/value dichotomy and other essays*. Cambridge: Harvard University Press.

Reichenbach, H. (1961 [1938]). *Experience and prediction*. Chicago: University of Chicago Press.

Reiss, J. & Sprenger, J. (2020). Scientific objectivity. In E. N. Zalta (Ed.): *The Stanford Encyclopedia of Philosophy*. https://plato.stanford.edu/entries/scientific-objectivity/.

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

Ruphy, S. (2006). Empiricism all the way down: a defense of the value-neutrality of science in response to Helen Longino's contextual empiricism. *Perspectives on Science*, 14 (2), 189-214.

Schurz, G. (1997). *The is-ought problem*. Dordrecht: Springer.

Schuessler (2019). *The debate on probable opinions in the scholastic tradition*. Leiden: Brill.

Sclove, R. (2010). Reinventing technology assessment. *Issues in Science and Technology*, 27(1), 34-38.

Searle, J. (1964). How to derive "ought" from "is". *The Philosophical Review*, 73(1), 43-58.

Shearman, D. J. C. & Smith, J. W. (2007). *The climate change challenge and the failure of democracy*. Westport: Praeger Publishers.

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

Steel, D. (2016a). Climate change and second-order uncertainty: Defending a generalized, normative, and structural argument from inductive risk. *Perspectives on Science*, 24(6), 696-721.

Steel, D. (2016b). Accepting an epistemically inferior alternative? A comment on Elliott and McKaughan. *Philosophy of Science*, 83, 606-612.

Steele, K. (2012). The scientist qua policy advisor makes value judgments. *Philosophy of Science*, 79(5), 893-904.

Steel, D. & Whyte, K. P. (2012). Environmental Justice, Values, and Scientific Expertise. *Kennedy Institute of Ethics Journal*, 22(2), 163-182.

Tomblin, D., Pirtle, Z., Farooque, M., Sittenfeld, D., Mahoney, E., Worthington, R., Gano, G., Gates, M., Bennett, I., Kessler, J., Kaminski, A., Lloyd, J., & Guston, D. (2017). Integrating public deliberation into engineering systems: Participatory technology assessment of NASA's Asteroid Redirect Mission. *Astropolitics*, 15(2), 141-166.

Weber, M. (1949 [1904]). *On the methodology of the social sciences*. Glencoe: The Free Press.

Weber, M. (1958 [1919]). Science as a vocation. *Daedalus*, 87(1), 111-134.

Wilholt, T. (2009). Bias and values in scientific research. *Studies in History and Philosophy of Science*, 40, 92-101.

Wilholt, T. (2013). Epistemic trust in science. *British Journal for the Philosophy of Science*, 64, 233-253.

Wilholt, T. (2016). Collaborative research, scientific communities, and the social diffusion of trustworthiness. M. Brady & M. Fricker (Eds.), *The Epistemic Life of Groups Essays in the Epistemology of Collectives* (pp. 218-249). Oxford: Oxford University Press.

Winsberg, E. (2012). Values and uncertainties in the prediction of global climate models. *Kennedy Institute of Ethics Journal*, 22(2), 111-137.

Markus Dressel is a research associate at the Research Unit Sustainability and Climate Risks at University of Hamburg, with previous occupations at Humboldt University Berlin, Leibniz University Hannover, and University of Graz. He has worked on a range of topics such as values in science, climate science and policy, social theory, or transdisciplinary research. His main focus is on normative issues of the science-society relation.

Address: University of Hamburg, Research Unit Sustainability & Climate Risks, Research Group Sustainability & Global Change, Grindelberg 5 (20144 Hamburg-Germany).
E-mail: markus.dressel@uni-hamburg.de. ORCID: 0000-0002-4789-5249