C.R.AIPS 2

Proceedings of *the Académie Internationale de Philosophie des Sciences*

Comptes Rendus de l'Académie Internationale de Philosophie des Sciences

Tome II

Justification, Creativity, and Discoverability in Science

Éditeur
Lorenzo Magnani

Hypothetical value judgements: Reconciling value-neutrality and value-engagement in science

Gerhard Schurz

Institut für Philosophie, Heinrich Heine University Düsseldorf, Werdener Straße 4, 40227 Düsseldorf, Germany

1 Introduction: the program of value neutrality

The value judgment debate constitutes one of the most enduring discussions in the philosophy of science and social philosophy. The first value judgment debate took place in Germany between 1913 and 1917 when Max Weber (1917) defended the value neutrality of the social sciences against the so-called "*Kathedersozialisten*" (professors advocating social policy), who argued that social scientists should proclaim value judgments from the lectern with scientific authority. Weber vigorously countered that objectivity in science can only be achieved if scientists limit themselves to descriptive statements of facts and clearly separate these from their value attitudes, for the reason that value judgments are neither logically nor empirically justifiable but arise from subjective human interests and worldviews.

A second value judgment debate occurred in the German-speaking world during the 1960s and 1970s in the context of the positivism dispute (cf. Albert/Topitsch 1971). Here, proponents of "critical theory" (particularly Habermas in 1968) opposed the ideal of scientific value neutrality, arguing that science is inevitably bound by interests, and the only question is *which* interests scientists serve—a view that was fiercly contested by proponents of the empirical-analytical approach.

A third value judgment debate developed in Anglophone philosophy, starting with papers of Rudner (1953), Jeffrey (1957), and Hempel (1965), during which several novel objections to scientific value freedom were raised, leading many philosophers of science to advocate values in science (e.g., Douglas 2000, 2009, Schroeder 2021, Holman/Wilholt 2022, Elliot 2017). Nevertheless, numerous defenders of the ideal of value freedom can still be found today (e.g., Betz 2017, Lacey 2013, Henschen 2021, Parker 2024).

The advocacy for values in science can be understood in a *weak* and a *strong* sense (cf. Betz 2017, 96f.). Weak in the sense of value engagement, meaning that scientists should strive to make their knowledge useful for the values of its users; and strong in the sense that the justification of scientific knowledge depends on or should depend on non-epistemic value assumptions. My position advocates for values in the weak but not in the

Justification, Creativity, and Discoverability in Science, edited by Lorenzo Magnani.
Comptes Rendus de l'Académie Internationale de Philosophie des Sciences 2 (2024).
G. Schurz, Hypothetical value judgements, pp. 149–178.

strong sense. I will therefore refer to my preferred version as the demand for *value neutrality*, as opposed to a general understanding of "value freedom," which demands that value statements should not appear in science at all. According to the position defended in this paper, value engagement in science is desirable, but a substantive dependence of scientific knowledge on non-epistemic values can and should be avoided. I will argue that a substantive value-boundedness must have a destructive impact on the enterprise of science and its reputation in society, while the properly understood value neutrality of science constitutes an important building block of democracy.

The value neutrality thesis developed here is based on a (at least partially) novel idea of how the value-independence of scientific knowledge can be connected with scientific value engagement: the idea of *hypothetical value statements*. These are conditional value recommendations derived by means-end reasoning from hypothetically assumed values, which do not come from scientists but from science users. The idea of hypothetical value judgments forms a central component of my position, together with a justification of why value independence in science is not only possible but also desirable.

In the next section, the value-neutrality thesis will be cleared from common misunderstandings and precisely defined (Section 2). Then, the idea of hypothetical value statements will be developed and illustrated with experiences from the Covid-19 pandemic (Section 3). Referring to the tragic case of the L'Aquila earthquake, it will be shown how this same hypothetical method can also solve the problem of "inductive risk" without having to resort to categorical value judgments (Section 4). After a brief outline of the metaethical justification of the value neutrality demand (Section 5), further controversial questions will be elegantly resolved through the method of hypothetical value statements, including the "new demarcation problem" (Section 6). Ultimately, it is argued that a value-neutral yet value-engaged position represents the best way to maintain the trust of broad segments of the population in scientific expertise within a democratic society.

2 The requirement of value neutrality: clarification and explication

First, we need a clarification of some important terms. Factual or *descriptive* statements are statements that say something about the factual constitution of the world, including not only to individual facts but also lawlike relations that may be of a strict or statistical sort. Simple value statements have the form "P is (or is not) valuable," where P is a proposition describing a state of affairs or an action. That a state of affairs P is valuable *cannot* be reduced to a factual statement about the value perception of some or many people regarding P (as suggested by Anderson 2004 and Clough 2008), but possesses *normative power*. Therefore, in all ethical systems, value statements and normative statements stand in a close conceptual relation which can be roughly summarized by the formula "The good should be done, and the bad should be avoided" (see Section 3). Value and normative statements are also summarized as *ethical* (or prescriptive) statements. The first prerequisite for fulfilling the requirement of value neutrality is to distinguish between factual statements and value statements in the practice of judgment. In natural language, this is not always easy, but with good will, it is always possible; more on this in Section 5.

In the value judgment debate, a variety of arguments have been introduced that generate important insights but distract from the core question of value neutrality. We will first address four such arguments before we can precisely define the value neutrality thesis.

1.) It has been argued that the social sciences deal with people's value attitudes as the object of their research and therefore must inevitably make value judgments (cf. Strauss 1971). However, as Max Weber (1917, 499–502) already pointed out, it is perfectly possible for the social sciences to empirically investigate the actual value attitudes of individuals or societies without making value judgments themselves. The claim that a person or society holds certain value attitudes or makes value judgments is not itself a value judgment but a descriptive statement, whose confirmation is based on empirical data.

2.) Already in the second value judgment debate it has been pointed out that science cannot be absolutely value-free, because science itself is based on certain so-called science-internal or *epistemic* values—primarily, the value of the pursuit of objective knowledge (Schmidt 1971)—and the same point has been made later in the Anglophone debate (Doppelt 2007). The reference to epistemic values is undoubtedly correct, but nevertheless irrelevant, as value neutrality concerns only non-epistemic values such as wealth, power, prestige, etc.; epistemic values are not what is meant here.

3.) As mentioned at the outset, the demand for value neutrality is often oversimplified into the notion of "value freedom," according to which values "have no place" in empirical sciences. This is a gross mischaracterization, as scientists can attain a number of *derived* values (or norms) from given categorical and fundamental values (or norms) using *means-end reasoning* based on empirical knowledge, and pass these derived values on as *recommendations of means* to the knowledge *users*. These recommendations of means are understood by science in a hypothetical or conditional sense, that is, *relative* to the assumed fundamental values: *If* certain fundamental values are assumed, *then* certain recommendations of means emerge as derived values. A value statement is termed *categorical* if it takes the form "P is (or is not) valuable," and it is termed *hypothetical* or conditional if it is an implication between categorical value statements. A categorical value statement is called *fundamental* if it cannot be derived from other value statements through means-end inferences; otherwise, it is termed *derived*.

The identification of suitable *means* for given *ends* is the most important practical task of empirical-descriptive science (see Section 3). What the demand for value neutrality excludes is only that the sciences establish fundamental categorical values (or norms).

4.) The entire process of scientific research is usually divided into three phases (cf. Schurz 2014, sec. 2.1, 2.5.3): In the *context of origination* (*CO*) (also called "context of discovery"), research questions are first defined. In the *context of justification* (*CJ*), data is collected, and hypotheses are generated and tested. In the *context of application* (*CA*), well-established findings are finally applied to various purposes. Non-epistemic values assume a role in both the CD and the CA. Which research questions are considered important enough to address is partly influenced by non-epistemic interests, and this is even more true for the choice of ends scientific knowledge is applied to. What the demand for value neutrality addresses alludes exclusively to the CJ.

Summarizing we can explicate our thesis as follows (see Figure 1):

> *Explication of the Value Neutrality Thesis:* The justification of scientific knowledge should be independent of fundamental non-epistemic value assumptions.

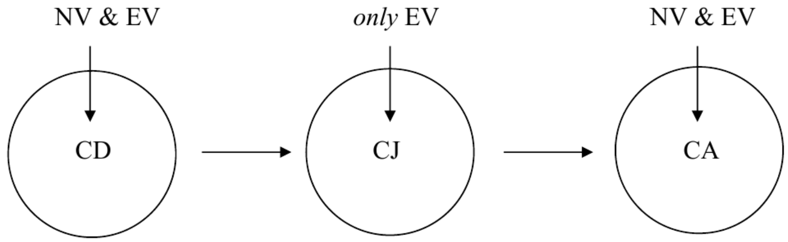

FIGURE 1. Schematic representation of the value neutrality requirement. *Legend:* CD – context of discovery, CJ – context of justification, CA – context of application; NV – non-epistemic values, EV – epistemic values.

The above value neutrality thesis is a *demand*, not a matter of fact, as not all scientists adhere to this standard. The content of the value neutrality demand can, alternatively, also be expressed as the *thesis of value independence*: the *justificatory status* of scientific claims is independent of non-epistemic values. This is a statement about the normative-epistemic status of scientific statements and not about scientists' practices.

Douglas (2000) has emphasized, too, that non-epistemic values influence *several* stages of the scientific research process. However, she refers specifically to the *context of justification*, such as the choice of scientific method. This, of course, is no longer compatible with the value neutrality thesis defended here. Supporters of Douglas (2000) often argue as follows: on the one hand, non-epistemic interests are decisive for the formulation of the *research question* in the CO, and on the other hand, the research question influences the *choice of methodology* in the CJ; this would prove that non-epistemic interests *indirectly* affect the choice of methodology.

While the argument of indirect influence is indeed correct, it nevertheless leads to confusion, because no direct influence is present: *the determination of the research question completely screens off extra-epistemic interests from the choice of the epistemically optimal method.* In other words, extra-epistemic interests may determine *which* research question is investigated, but once the research question is precisely defined, the choice of the optimal method for finding a true or most probable answer is a purely epistemic matter. The concept of *screening-off* goes back to Reichenbach (1956) and is defined within the Markov theory of causality (Spirtes et al. 2000, Sec. 3.4.1–2; Schurz and Gebharter 2016): direct causes screen-off indirect causes from the effects, meaning that if the direct cause of an effect is fixed, knowledge of the indirect cause does not exert any additional influence on the probability of the effect.

Our thesis that the precise formulation of the research question screens-off non-epistemic value influences from the epistemic question of method choice aligns well with Parker's (2024) *theory of epistemic projection.* According to this theory, non-epistemic values are projected onto a research problem, which is then addressed based solely on epistemic criteria, with the answer being more or less useful to the original non-epistemic interests depending on the outcome (ibid., 19, 33). The theory of screening-off explains why answering the research question is a purely epistemic matter, although non-epistemic interests were initially at play. We illustrate this with an example. A research project on efficient treatment methods for *depression* may be driven by the non-epistemic interests of the pharmaceutical industry, which are focused on *pharmaceutical* therapy options: which antidepressants have the best effects (operationalized in terms of efficiency, sustainability, etc.)? This is a precise question, and the choice of the optimal method for this question (data collection, statistical procedures, etc.) *no longer* depends on non-epistemic interests—provided our solution to the "inductive risk" issue in Section 4 is accepted. This holds true even if it turns out that for certain types of depression, *psychological* therapies are more effective than antidepressants. In that case, a rigorously conducted study would yield a largely negative result, meaning antidepressants would have only limited success. Consequently, in the context of justification, it would become clear that restricting research to pharmaceuticals is *counterproductive* for exploring efficient depression treatments. In other words, *even if* non-epistemic motives are initially at play in the CO, their influence can be *corrected* by insight gained in the CJ. The same example also illustrates why the so-called *positivity bias* in research, according to which negative research results are often not published, can lead to significant research distortions (Ioannidis 2005, Schooler 2011). However, this is not an inherent characteristic but a defect of research under career pressure; many medical journals have now implemented measures to ensure the publication of negative research results.

3 Means-end inferences and the significance of hypothetical value statements in science

In its simplest form, the means-end inference—abbreviated as ME inference—has the following form, which can be formulated for both value and norm statements (separated by "/"):

> *ME (means-end) inference, simple form:*
>
> *Descriptive means-end premise:* Under the given conditions, M is a necessary—or alternatively, an optimal—means for achieving the end E.
>
> ---
>
> *Therefore—hypothetical value/norm statement:* If E is accepted as a fundamental value/norm, then (the realization of) M is a derived value/norm.

In the ME inference, it is assumed that M is either a *necessary* means—i.e., every end-realizing action entails M—or an *optimal* means, i.e., M maximizes the overall utility with regard to side effects. However, if M is merely a *sufficient* means, then the means-end inference is invalid. This is because for many sufficient means the costs of their side-effects outweigh the benefit of achieving the end. For instance, for letting fresh air into a room breaking through the wall is a sufficient means, too; an optimal means is opening a window and a necessary means is any form of opening to the outside.

The logical form of the inference from ends to necessary means, abbreviated as the NME inference, is straightforward. The descriptive premise "B is a necessary means for end A" can also be rephrased without the terms "means" and "end" through the law-like statement "B is a necessary condition for A," formally "$\Box(A \to B)$" (necessarily, if A then B), where in ethics $\Box$ represents *practical* necessity (meaning, roughly speaking, nomological consequences of boundary conditions, e.g., concerning our planet, that cannot be practically changed). The NME inference then has the following logical form, where O stands for the value operator ($\mathrm{O}A$ for "A is valuable") or the obligation operator ($\mathrm{O}A$ for "A is obligatory"):

Logical form of the NME inference:

$$\frac{\text{Descriptive premise: } \Box(A \rightarrow B)}{\text{Hypothetical value/norm conclusion: } \mathrm{O}A \rightarrow \mathrm{O}B}$$

For better understanding, the following explanations are added:

1.) In $\Box(A \rightarrow B)$, B can be a necessary condition for A as well as a necessary consequence of A; the inference holds for both cases. $\mathrm{O}A$ represents the fundamental value (or norm), and $\mathrm{O}B$ represents the derived value (or norm).

2.) The NME inference can also be transformed into the following inference using the deduction theorem: Premise: $\Box(A \rightarrow B)$, 2. Premise: $\mathrm{O}A$ / Conclusion: $\mathrm{O}B$. In this form, however, the inference becomes problematic in the sciences, as the second premise is no longer scientifically justifiable. Therefore, if the NME inference is presented in the form above, the second premise must be explicitly *marked* as a value assumption and clearly separated from the first descriptive premise.

3.) The logical form of the inference from the end to *optimal* means is more complex and will be determined further below.

4.) As mentioned, both the ME inference and the connection between values and norms—abbreviated as VN connection—are accepted as *analytically* valid in most ethical theories. By an analytically valid statement, we either mean a logical truth or an extralogical meaning convention (or a statement that logically follows from it). The analytical nature of the VN connection and the NME inference is established in different ways in deontological versus teleological ethical theories (Frankena 1963, 13–16). *Deontological* theories are based on the concept of *intrinsic value*: an action is intrinsically valuable if it obeys certain supreme moral principles, *regardless* of its consequences. Every intrinsically valuable action is obligatory—Kant (1785, BA 52) calls this a "categorical imperative"—and the ME inference is deontologically valid; the obligation of B following from the obligation of A by the ME inference is called a "hypothetical imperative" by Kant (1785, BA 40). In *teleological* theories, by contrast, the value of an action or state of affairs is determined based on the value of its *consequences*. The *overall value* of an action is defined as the "sum" of the utility values of all of its consequences, with benefits being positive and costs being negative utility values. The VN connection principle of teleological theories characterizes an action as obligatory if it maximizes overall utility (or if it is logically entailed by all utility-maximizing actions if there are multiple equally utility-maximizing options; cf. Schurz 1997, Ch. 11.2). The fact that all options for action must be physically or practically possible validates the ME inference for teleological imperatives. Applied to values, the ME inference only holds for *overall values*, not for intrinsic values. The latter would result in the notorious "sanctification of the means by the ends," which is often unacceptable. For instance, it is intrinsically valuable to keep a promise, but if this is only possible at the cost of a human life, then the harm of the action's consequence outweighs the intrinsic value of the action; this results in a negative overall value of the action A, and the NME principle remains valid, i.e., $\Box(A \to B)$ holds true, but both $\neg \mathrm{W}B$ and $\neg \mathrm{W}A$ also hold.

Based on these specifications, we come to the consequences for understanding value neutrality. As explained, the fundamental purpose of the ME inference is taken over by the expert from the knowledge users, politicians, or practitioners. The expert returns to them a recommendation of means, which must be hypothetically relativized to the assumed fundamental purpose. Only this relativization allows the knowledge users to check whether the fundamental values assumed by the expert are also their own values. The combination of the demand for value neutrality and the ME inference thus accounts for the user's claim to maturity and strengthens democracy. On the other hand, if experts omit the hypothetical value relativization and formulate their recommendations categorically, this can have the politically problematic consequence that knowledge users may be led to actions that do not align with their own interests—which can seriously undermine trust in expert judgments (cf. Elliot et al. 2017; Intemann 2024, 10).

However, the above ME schema is *oversimplified*. The reason a potential knowledge user rejects the expert's assumed ultimate purpose is usually that the costs of the recommended means outweigh its benefits from the user's perspective. The listing and evaluation of the side effects of the recommended means is *omitted* in the above schema. It can have harmful consequences if users are not informed about possible negative side effects. In the case of pharmaceutical drugs, informing about side effects has long become a matter of course, but in expert recommendations concerning the Covid-19 pandemic the duty to inform about side effects was often neglected in expert recommendations (see below). Consequently, we propose refining the above necessary-means-end inference as follows:

Side-effect-transparent schema of the necessary-means-ends (NME) inference:

Descriptive premise: M is, under the given conditions, a necessary means for achieving the end E, and this means has the side effects S.

Therefore—hypothetical value/norm statement: If E is accepted as a fundamental value/norm—and therefore the benefit of realizing E outweighs the costs of S—, then (the realization of) M is a derived value/norm.

The side-effect-transparent version of the NME informs about all the consequences of the recommended means and not just about the intended consequence. Observe the practical dialectics of the side-effect-information between the em dashes: Logically speaking, this information follows already from the premise that E is accepted in the "overall sense". However, most people do not distinguish between intrinsic and overall values. If the information about side effects were omitted, they could be uncritically inclined to agree with value E, which they might not do if they were informed about the side effects. Therefore, adding this information is so important, even if it is logically redundant. Thus, logically speaking, the side-effect-transparent NME inference still has the form $\Box(E \rightarrow M)/\mathrm{O}E \rightarrow \mathrm{O}M$, but now with the additional premise-information $\Box(M \rightarrow S)$, from which $\mathrm{O}E \rightarrow \mathrm{O}S$ follows, i.e., $\mathrm{O}E$ can only hold if the side effects S are acceptable.

Reconstructing the ME inference for *optimal means* is more complicated, as the statement that M_i is "optimal" among all possible means $M_1, \ldots, M_n$ is, strictly speaking, not a descriptive but a normative-evaluative statement and therefore belongs in the conditional part of the hypothetical conclusion. Each means M_i now has its own side-effects S_i. For a *hypothetically* assumed cost-benefit assessment of all consequences, the overall utility of these means must be determined, justifying that M_i is the means with maximum overall utility among and this overall utility is positive:

> *Side-effect-transparent schema of the optimal-means-ends (OME) inference:*
>
> *Descriptive premise:* $M_1, \ldots, M_n$ are the (practically) possible means for achieving the end E, and the means M_i have the side effects S_i $(1 \leq i \leq m)$.
>
> ---
>
> *Therefore—hypothetical value/norm statement:* If (a) E is accepted as a fundamental value/norm and (b) a cost-benefit assessment is assumed under which M_k has maximal overall utility among $M_1, \ldots, M_n$—and therefore the benefit of E outweighs the costs of the side-effects S_k—then (the realization of) M_k is a derived value/norm.

The content of the premise can be logically expressed by the implication $\Box(E \rightarrow (M_1 \vee \ldots \vee M_n))$, i.e., if E is to be attained, one of the possible means has to be realized. The content of condition (b) can be qualitatively represented by the implication $\mathrm{O}(M_1 \vee \ldots \vee M_n)) \rightarrow \mathrm{O}M_k$, i.e., if one of the possible means for E should be realized, then the best one (M_k) should be realized. The condition between the em dashes is again logically redundant and follows from the joint effect of conditions (a) and (b). In conclusion, the above inference can be formalized as follows:

OME inference:

$$\Box(E \to (M_1 \vee \ldots \vee M_n))/(\mathrm{O}E \wedge (\mathrm{O}(M_1 \vee \ldots \vee M_n) \to \mathrm{O}M_k)) \to \mathrm{O}M_k$$

In this reconstruction, the OME inference follows from the NME inference, as the premise implies $\mathrm{O}E \to \mathrm{O}(M_1 \vee \ldots \vee M_n))$, which implies the conclusion.

Different value assumptions can lead to different cost-benefit assessments and thus to different optimal choices of means. This is particularly relevant for democratic societies in a state of strong *value polarization* (cf. Abramowitz und Saunders 2008, Le Bihan 2024, 3). In this situation, it becomes advisable for value-neutral policy advice to conduct hypothetical cost-benefit assessments for *several different* value preferences that reflect the political positions of relevant population groups and are presented to the audience as "alternative options".

The problem of different value weightings became especially pressing during the *COVID-19 crisis*. The value weighting imposed on us by the coronavirus was one between equally fundamental values that came into conflict: health on the one hand, and freedom and well-being on the other. Value decisions of this kind, besides considerations of reason, always depend on actual human interests and therefore contain an inevitably subjective component. Some people were willing to forgo freedoms such as communication, sports, and culture for months for a certain statistical increase in health security, while others found this completely disproportionate. Such subjective differences in attitude must be *acknowledged*, which is why collective value decisions must be tied back to democratic majorities and cannot be dictated by experts, no matter how important expert knowledge may be for understanding the consequences of actions to be evaluated. Therefore, one should expect an expert recommendation in the COVID crisis to name not only the expected benefits but also all costs and to conduct a hypothetical value assessment.

The statement of a commission of the Leopoldina (Germany's National Academy of Sciences) of December 8, 2020, did not meet this value neutrality standard: it stated that a "hard lockdown" was "absolutely necessary from a scientific point of view".[1] This was an is-ought fallacy, as no normative conclusion can be derived from facts (see Section 5). Moreover, the evaluation of the costs of a lockdown compared to the expected benefits was missing, which led to a critical discussion (Wiesing et al. 2021). A year later, the Leopoldina statement of November 27, 2021, had made remarkable progress. It presented two possible options: Option 1: rigid lockdown for everyone, versus Option 2: contact restrictions only for the unvaccinated.[2] The

[1] Leopoldina. 7. Ad-hoc-Stellungnahme. Coronavirus-Pandemie: Die Feiertage und den Jahreswechsel für einen harten Lockdown nutzen. 8 December 2020.

[2] Leopoldina. 10. Ad-hoc-Stellungnahme. Coronavirus-Pandemie: Klare und konsequente Maßnahmen – sofort! 27 November 2021.

demand for hypothetical value relativization was thus met, giving those citizens who preferred Option 2 the possibility to reject the Leopoldina's preferred Option 1 *without* resorting to the corner of science denial and "alternative facts."

In summary, we have shown in this section that value neutrality and value commitment, from the perspective of means-end inferences, are not opposites but two sides of the same coin. Why, then, is there persistent talk of the blanket "value-free ideal" instead of value neutrality even in recent debate?[3] Presumably, this has several reasons. As Douglas and Branch (2024) explain, in Western societies in the mid-1950s, a conception of science was predominant according to which the pure sciences should only serve the search for truth and were freed from social responsibility. This conception of science also included the said blanket "value-free ideal." In the 1950s the value-free ideal was further supported by the then widespread view of the *non-cognitivism* of values, according to which value statements were non-rational emotive expressions. However, the alternative view soon gained ground in Analytic Philosophy that a rational treatment of value questions is both possible and socially necessary, and since the 1970s, the disciplines of deontic logic, rational decision theory, and analytic ethics have rapidly established.

Compared to the 1950s, today's scientific self-image has changed significantly. Leading scientific associations now assume that scientists should take partial social responsibility for the foreseeable consequences of their knowledge and therefore engage with its instrumental usability, not only in "applied" but also in "pure" sciences (Douglas/Branch 2024, 12). Although this view perfectly fits our understanding of hypothetical means-end inferences, Douglas and Branch bypass this possible solution and instead argue that the value-free ideal should be rejected and science should be made dependent on categorical values (ibid., 13). From our perspective, this is a clear non sequitur. This brings us to a further reason why the narrow-minded notion of value neutrality as the absence of values lasts so long among opponents of value neutrality: because it is used as an easily criticizable straw man. For example, Dupré (2007) has argued that even if it were possible to separate the factual content from the value content of knowledge, such a separation would be counterproductive because then science could not have practical consequences. The contrary is true, however, as scientific value recommendations are not only allowed but explicitly welcomed by the demand for value neutrality, as long as they appear in the form of hypothetical value statements.

[3]Wilholt (2009), Betz (2017), Henschen (2021), Douglas/Branch (2024), Parker (2024), and more.

4 The argument of inductive risk and its resolution: uncertainty transparency and hypothetical cost-benefit assessment

In the third value judgment debate, a novel objection to the thesis of value neutrality was raised which did not play a role in earlier controversies: the argument of inductive risk, or AIR for short. This argument, which goes back to Rudner (1953) and was developed further by Douglas (2000), assumes that in most cases, scientific hypotheses are only supported by empirical evidence with a certain *probability*. When scientists accept a hypothesis H as true, they take a certain 'inductive risk' that H is false: the so-called *error risk*. For example, if the probability of H is 95%, then its error risk is 5%. But what probability of H is still high enough to reasonably accept H as true? For a practically relevant hypothesis the decision to accept it as true means relying on it, that is, being willing to *act* on its basis. Acting on the basis of H brings a benefit if H is true, but a costs if H is false. However, if the costs are much higher than the benefit, then even a hypothesis probability of 95% may not be sufficient to justify acting on the basis of H. Therefore the acceptance of a hypothesis as 'true' inevitably involves extra-epistemic evaluations.

To avoid misunderstandings, we do not claim that all scientific judgments carry a non-negligible risk of error (Betz 201, 98). Judgments such as "The Earth has one moon" (etc.) can be considered practically certain and safely asserted as knowledge categorically. Other hypotheses, like the Big Bang theory, are not practically relevant, so their acceptance is independent of non-epistemic values for this reason. However, many scientific judgments are uncertain *and* practically relevant—and it is to these kinds of judgments that the AIR refers.

A number of authors see the AIR as 'compelling evidence' for the influence of extra-epistemic values on the context of justification.[4] Here, the opposite is to be shown: AIR is much weaker than thought. For AIR can be refuted by the demand for *uncertainty transparency*—i.e., the explicit indication of the involved error risks—and this demand not only serves value neutrality but also prevents questionable consequences of expertocracy.

To refute AIR, we use the prediction of *earthquake safety* as an example. Let H be the hypothesis that (in a given area in the next few days) *no* earthquake of a Richter-scale-magnitude greater than 6 will occur. If H is accepted, then the affected people stay in their homes, which, if H is true, brings a comparative benefit of zero (no additional costs), but if H is false, it brings very high costs C_h, which can mean injuries or death, at least for those whose houses are at risk of collapsing at this earthquake magnitude.

[4] See Wilholt (2009), Steele (2012), Douglas/Branch (2024); Holman/Wilholt (2022) even speak of a "general consensus".

On the other hand, if the affected people accept non-H, considering an earthquake likely, then the evacuation of their homes is carried out, which brings comparatively lower costs C_l, both if non-H is true and if non-H is false. If p denotes the probability of an earthquake, i.e., the counter-probability of H, then the expected utility of H, abbreviated as $E(H)$, as well as that of non-H, are as follows:

$$E(H) = (1-p) \cdot 0 - p \cdot C_h = -p \cdot C_h.$$

$$E(\text{non-}H) = -(1-p) \cdot C_l - p \cdot C_l = -C_l.$$

Accepting H is better than rejecting it, if and only if $E(H) > E(\text{non-}H)$, i.e., iff $C_l/C_h > p$.

The acceptance of the hypothesis H (no earthquake) is therefore only reasonable as long as the ratio of the costs of its erroneous acceptance to the costs of its erroneous rejection is greater than its error probability p—in our example, greater than 5/100.

It is undoubtedly true that the question of accepting non-H versus H in a qualitative sense—and thus the question of whether to evacuate or not—depends on extra-epistemic cost-benefit considerations. But does this really mean that scientists should make such decisions on behalf of the affected people, as proponents of AIR would suggest? My answer to this question is: *No*, this is actually often the worst thing scientists can do in such a situation. Instead, scientists should explicitly state the involved error risks and make them transparent to non-experts through *hypothetical* cost-benefit considerations. We call this the condition of *uncertainty transparency.*

However, if scientists present qualitative statements *categorically* and conceal the error risk, unpleasant consequences may easily ensue for both the knowledge users and the knowledge producers. The earthquake example illustrates this clearly. In October 2012, six earthquake experts and a government official in Italy were sentenced in the first instance to several years in prison because they did not predict an earthquake that occurred on April 6, 2009, in L'Aquila, in which more than 300 people died, but instead gave the all-clear.[5] The case attracted worldwide attention. *If* the judges' view were correct that the earthquake experts gave the all-clear categorically at that time, despite a slightly increased risk due to the registered tremors, then the earthquake researchers would indeed bear partial responsibility: because they acted in the sense of Rudner and Douglas's recommendation, took the cost-benefit assessment out of the hands of the affected individuals by announcing an "all-clear", and thereby concealed the involved risk. Even if the risk of an earthquake of magnitude 6 to 7 increases from only 0.5 to

[5]For the following, see Edwin Cartlidge: "Seven-year legal saga ends as Italian official is cleared of manslaughter in earthquake trial. Verdict follows conviction of deputy for advice given ahead of L'Aquila earthquake." *Science*, 3 October 2016.

1%, this may be reason enough for one person to endure the hardship of leaving their home due to their specific circumstances (for example a mother with her kids), while another person is willing to take the risk (for example a farmer who cares for his livestock). Which scientist would wish to take the decision out of the hands of the affected individuals in such a case, citing their "expert authority", and be held accountable for the consequences under threat of punishment? Probably no one.

The earthquake researchers defended themselves against the court's judgment by pointing out that they did not issue a categorical all-clear, but rather a more cautious statement, stating that the registered slight increase in tremors was not significant but still "within the normal range." In 2014, the judgment was revised, and the involved scientists were acquitted; only the government official, who had conveyed the "all-clear" message to public media, was convicted. In any case, this remarkable story clearly demonstrates that categorical evaluations of risk consequences are *entirely* outside the purview of scientists.

The counter-argument against AIR, which asserts that uncertain scientific statements should be framed as *probabilistic statements*, was first articulated by Jeffrey (1956, 237). More recently, referencing Schurz (2013), Betz (2017) and Henschen (2021) have advocated addressing the problem of inductive risk by making error risks explicit. Douglas and like-minded authors counter this position by arguing that the public and their political representatives desire definitive statements with clear action implications, such as "all-clear," from the "authority of science" (Douglas 2000, 563; 2009, 135; John 2015, 82; Wilholt 2009, 94). However, scientists must not yield to such false expectations. Rather, it is their duty to educate people about the limits of what is knowable—which in this case means informing the public about the error risks of their predictions. Nonetheless, we agree with the authors that merely providing a probabilistic significance statement, hardly understandable for laypeople, is insufficient for such public enlightenment. Besides indicating the error probability of scientific predictions or assertions, it is crucial to make this error probability comprehensible to non-experts. Here again, our proposal relies on the idea of *hypothetical action recommendations based hypothetical cost-benefit evaluations*. In our earthquake example, this would mean illustrating the practical significance of a 5% error probability as follows: *If* the potential negative consequences of an earthquake during one's stay at home are assessed to be more than 20 times higher than the costs of a temporary evacuation, *then* it is advisable to undertake the evacuation. This proposal goes far beyond Jeffrey's requirement that experts should simply report the error the probabilities: it takes justice to Douglas (2020) requirement that experts cannot merely report error probabilities and leave the decisions to others, but due to the hypothetical mode, this now takes place in a value-independent manner.

After this fundamental defense of value neutrality against the AIR, three *refinements* of our argumentation are presented in the context of the current debate.

1.) Rudner (1953, 4) and Douglas (2009, 53f.) developed a fundamental objection to the proposal that scientists should limit themselves to probabilistic statements. They argued that even if scientists only assert a probability, they have already accepted a hypothesis, namely a probability hypothesis, which itself is uncertain, thereby perpetuating the problem. If this were true, the probabilistic proposal would be subject to the problem of *infinite regress* and would collapse. But closer inspection reveals that this is not the case. Suppose there is second-order uncertainty about the given error probability of the following form: "With 95% probability, the error risk of H lies between 4% and 6%." This is sometimes assumed in Bayesianism, but by forming expectation values or confidence intervals one projects the second level information back to the first level. The only practical question affected by a second-order evaluation is whether to base the decision on the probabilistic *expectation value* of the risk (which in our example is 5%) or to quantify the risk using the lower and upper risk-limits of 4% and 6% (according to the confidence interval method; see Cox and Hinkley 1974, 49, 209). The second variant weakens the decision rule's discriminatory power and increases the possibility of stalemates. Otherwise, there are no practically relevant consequences, and at the next-higher (third) evaluation level, the practical consequences are zero. Therefore, the regress stops at the second level at the latest, and the regress argument is refuted.

2.) Douglas (2000, 563) argues in her contribution that the influence of non-epistemic values cannot be avoided by limiting statements to probabilities because an error risk arises not only in the formulation of qualitative statements but also at other points during the internal research process, particularly in *data description* (ibid., 569f.) and the *interpretation of results* (ibid., 573f.). This is true, but I don't see a fundamental problem here. All these error risks add up to the overall risk of the final hypothesis H, according to the laws of probability theory. To decide whether H should be accepted as true and used as a basis for action, it is only necessary to know this overall risk. One may object that some of these risks rely on implicit knowledge and are not reflectively available to the scientists. This may be true, but recall that value neutrality is a normative requirement: it requires that these risks *should* be made reflectively available as far as possible, and even if they are entirely unknown or based on shaky guesses, this has to be stated instead of being suppressed.

Moreover, the examples Douglas cites do not solely concern error risks but also inadequate methodological steps. Concerning data description, Douglas addresses the question whether borderline cases between non-carcinogenic and carcinogenic lesions in lab rats should be classified as carcinogenic or non-carcinogenic. She argues that this question is influenced by extra-epistemic risk assessments (ibid., 571f.). However, both options are epistemically inadequate, since non-assignable borderline cases should be marked and excluded from statistical calculations. Concerning the interpretation of results, Douglas discusses whether a found correlation between dioxin and increased cancer rates should be interpreted by a continuous increase model or a threshold model. This is also not a matter of extra-epistemic preferences but can be answered through refined empirical methods such as nonlinear curve regressions.

3.) In the discussion of AIR, the tolerable probability of the α-and β-error in the context of choosing the *significance level* is frequently referenced (cf. Wilholt 2009, Section 6; Henschen 2021; Douglas 2000, 563f.). We conclude this section by addressing this connection. Let H be a statistical correlation hypothesis: the influence of dioxin on cancer incidence in lab rats, from Douglas's example. Testing H involves comparing the cancer frequency in two samples: an *experimental group* of lab rats administered dioxin and a *control group* without dioxin. Following Henschen's (2021, 9–10) continuation of this example, assume that the cancer frequency in the experimental group is 25%, and 15% in the control group. A statistical significance test then asks: is the frequency increase by 10% due to chance, or is it "significant," i.e., attributable to a presumed causal relationship? The α-error denotes acceptance of H when H is false (or rejection of the opposing "null hypothesis," $\neg H$, when the latter is true). The probability of the α-error is the likelihood that a difference Δ at least as large as the *observed difference* $\Delta_o = 10\%$ would be randomly found between two samples. (In our example, we only consider positive differences, i.e., a 'one-sided t-test' is conducted.)

The probability distribution of the frequency differences Δ between two 100-element random samples of a binary characteristic is a normal distribution; this is depicted in Figure 2 by the solid left (blue) curve. Using the underlying distribution formula (Henschen 2021, 9), one calculates the probability of randomly observing a difference of at least Δ_o to be approximately 4%. The number 4% is the so-called "p-value" of Δ_o respectively the error probability of H—in Figure 2, this corresponds to the proportion of the area below the normal distribution curve at the right side of the value 10.

Note: strictly speaking, p is not the error probability of the hypothesis but rather that of the underlying test-statistical *procedure*—i.e., the probability of observing a frequency difference of at least Δ_o in pairs of 100-element random samples amounts to 4% (Cox/Hinkley 1974, 49, 209). Only under the assumption that the data were *representative* for H is it legitimate to interpret p as the error probability of H given Δ_o.

In test statistics, the significance level refers to a probability threshold s that the p-value of the observed difference must fall below for H to be acceptable as "significant"; this threshold corresponds to a "minimum significant difference." Typically, s is set at 5%; in our example with sample sizes of 100, the minimum significant difference is calculated to be 9%, which is indicated by the vertical line in Figure 2. This definition is made *pragmatically* in test statistics and is by no means mandatory. Rather, a categorical definition of certain "acceptance thresholds" has to be criticized, because as explained, the scope of acceptable error probabilities depends on external cost-benefit evaluations. In fact, a categorical setting of acceptance thresholds does not correspond to standard statistical practice; instead, the mentioned p-value of the observed difference is always explicitly indicated as the correlation hypothesis' error risk.

The significant sample difference decreases proportionally to the increase of the square root of the sample size $n(\Delta \sim 1/\sqrt{n})$. This is illustrated in Figure 2 by the dashed left (blue) curve, representing the steep-peaked normal distribution for a sample size of 2000. Any minimal sample difference becomes significant if the samples are sufficiently large. Therefore, the mere assertion that a "highly significant" relation was found between two variables is a weak claim without specifying the sample size, and merely indicates some possibly *very slight* correlation (here between dioxin and cancer rate). Far more important is the indication of the so-called *effect size*, i.e., the strength of the correlation, which in our example can be measured by the shift in frequency and amounts to 10% (cf. Andrade 2020).

This brings us to the second error type, the *β-error*, which, complementary to the α-error, denotes the rejection of a true H (thus the acceptance of a false $\neg H$). The probability of a β-error depends not only on the chosen significance level s for the α-error and the sample size but also on the claimed effect size of H. Suppose we are interested in dioxin-induced frequency increases of at least 15%. The probability distribution of the differences between two samples from different populations with a frequency difference of 15% is depicted in Figure 2 by the solid right (red) curve. Assume s is set at 5%, i.e., the minimal significant difference is 9%. Then the probability of a β-error is the probability that a sample drawn from a dioxin-exposed population with a cancer frequency of at least $(15 + 15 =)$ 30% will randomly deviate so far to the left that it still falls within the acceptance interval of $\neg H$, thus leading to the rejection of H. In our example, the probability of the β-error is calculated as 15%; in Figure 2 this is the proportion of the area

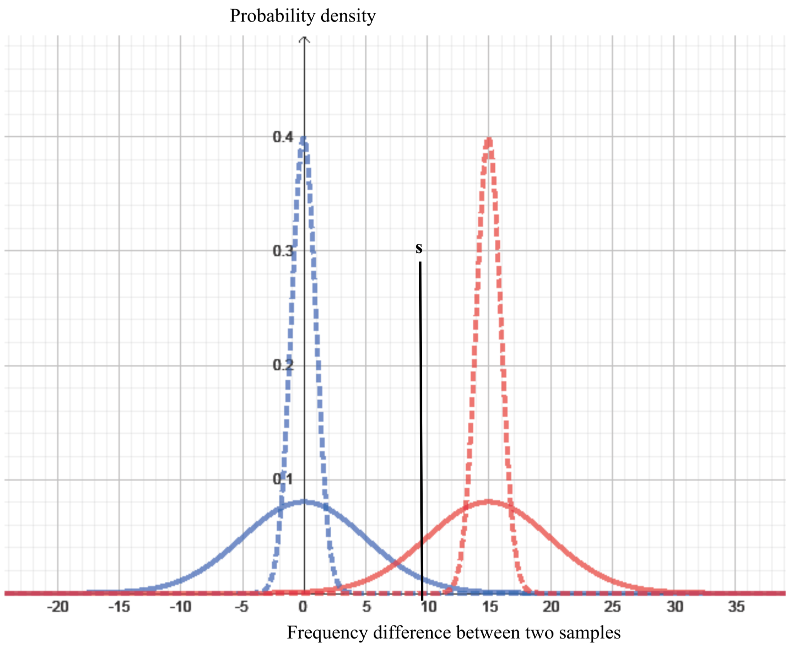

FIGURE 2. Normal distribution of frequency difference between two samples. Left (blue): Assuming the null hypothesis. Right (red): Assuming the alternative hypothesis of a frequency increase of 15%. Solid line: Sample size 100. Dashed line: Sample size 2000. Vertical line: significant difference for significance coefficient of $s = 5\%$ and sample size of 100. α-error: Proportion of area below the left-hand curve to the right of s. β-error: Proportion of area below the right-hand curve to the left of s. (Created with Geogebra.)

below the right solid normal distribution at the left side of the significance threshold s.

For constant sample size and effect size, the probability of a β-error is the greater the smaller the tolerable probability of an α-error. From this, Douglas (2000, 563f) and others have concluded that reducing the α-error risk "inevitably" comes at the cost of increasing the β-error risk (Betz 2017, 95; Parker 2024, 21f.; Wilholt 2009, 94f.). For Douglas, this is the gateway for the influence of extra-epistemic values: a low α-error risk reduces the number of falsely diagnosed dioxin-cancer correlations, thus leading to "underregulation" in the interest of the industry, while reducing the β-error risk decreases the number of overlooked true dioxin-cancer correlations, thus leading to overregulation in the interest of the health-endangered population.

The fact that the α-error risk is kept low in statistics supposedly serves the interests of the industry. However, this argument is misleading for two reasons.

Firstly, by increasing the sample size, it is always possible to circumvent the α-β tradeoff. This is illustrated in Figure 2 by the two dashed curves (left and right) for a sample size of 2000. Since the two steep-peaked distributions barely overlap, this allows the α- and β-error risks to be *simultaneously* reduced far below 0.01%.

Secondly, even when dealing with small sample sizes, the demand for uncertainty transparency requires reporting not only those sample differences that were significant at $s = 5\%$; rather, *all* found sample differences should be reported, including the non-significant ones. For each found difference Δ_o, (a) the error risk of the correlation hypothesis (α-error risk) should be specified, and (b) the error risk of the null hypothesis (β-error risk) for an assumed minimum effect size, *if* this null hypothesis were claimed at the found difference Δ_o (i.e., if s were identified with the p-value of Δ_o). Step (a) is standard practice in statistical investigations. Step (b) is often omitted but would be helpful in providing transparent information for statements with higher α-error risk and lower β-error risk, in line with Douglas. Illustrated in terms of our example: suppose we find a frequency increase of only $\Delta_o = 5\%$. Then the error probability of H given $\Delta_o = 5\%$ is calculated to be 19%, and the error probability of $\neg H$ for effect sizes of at least 15% is 4%. Summarizing, clarification of all error risks is possible even in the more complex case of statistical comparison tests, without the need for categorical valuations to be involved.

5 The metaethical justification of value neutrality

Besides its foundation in the philosophy of science, the thesis of value neutrality also rests on certain *metaethical* premises, the thorough justification of which would exceed the scope of this work; however, they should at least be mentioned in this section. The metaethical justification of the thesis of value neutrality is based on two theses:

Thesis 1 – The is-ought or fact-value dichotomy: From purely descriptive statements, value or normative statements cannot be logically or analytically derived without implicitly presupposing ethical premises.

Thesis 1 goes back to David Hume (1739/40, 177f.) concerning logic and to G. E. Moore (1903, 15f.) concerning analytical meaning postulates. However, this alone is not sufficient to justify the thesis of value neutrality. For one could hold the position that there are also so-called *value science*s that can justify categorical ethical statements with similar objectivity as factual statements are justified by empirical sciences. Those who take this position can agree with the fact-value dichotomy and still argue that fundamental extra-epistemic values have their place in science, as they obtain their objective justification in the value sciences. Several intellectual movements have held such a position; for example, (neo)marxism (Habermas 1968) or value platonism (Scheler 1913/16). Therefore, to justify value neutrality the following second thesis is additionally required:

Thesis 2 – the logical-empirical nature of the scientific method: Scientific knowledge is based on experience and logical-mathematical knowledge. Only through this type of knowledge can objective truth be achieved reliably. Value statements, on the other hand, are dependent on subjective interests or intuitions. Therefore, in the realm of values, no objectivity comparable to that of empirical sciences can be achieved.

Ad 1 – Justifying the dichotomy: Attempts to prove the is-ought dichotomy have led to a number of problems. This has led several philosophers (Searle 1969, MacIntyre 1981, Putnam 2002) to question the dichotomy thesis, which has only recently been placed on a solid logical foundation through newer investigations (Pigden 1989, Schurz 1997, Russell 2022). Three main problems arise in this context:

1.1 The conceptual fact-value distinction: The thesis of value neutrality first requires the separation of descriptive from ethical statements. For formal languages, this distinction is ensured by the different forms of expressions, but for natural language, a problem arises. Many everyday language concepts simultaneously possess factual and value content, often closely intertwined, making the conceptual separation difficult. Some critics of value neutrality, such as Putnam (2002, Ch. 2) and Dupré (2007), have claimed that for so-called "thick concepts" like "rape," a fact-value distinction is impossible. However, as Hare (1981, 74ff.) pointed out, this separation is more a matter of will than of capability and is always possible with the aid of appropriate linguistic means (see Schurz 2013, Sec. 6.2). We illustrate this with Dupré's example of the concept of rape of a woman. The descriptive part of the concept would, at first approximation, be given when a man forces a woman against her will to engage in sexual intercourse or similar acts. The conjunction of the above description and the value judgment "And this is a serious offense" can be seen as a satisfactory approximation of the descriptive-normative overall meaning of "rape," which complies with the separation requirement. The separation of the descriptive and prescriptive content is also necessary for the empirical investigation of a question that Dupré also considers to be of utmost importance, namely which measures can most effectively reduce the frequency of attempted rapes (Schurz 2013, Section 6.2).

1.2 The logical dichotomy: To justify this dichotomy one must show that purely descriptive premises do not logically entail ethical conclusions. The main problem for proving this thesis was that there are also mixed conclusions containing both descriptive and normative components, leading to the notorious "paradox of Prior" (Prior 1960). This problem was solved by applying a relevance criterion, according to which purely descriptive premises never logically entail conclusions containing *relevant* (i.e., not arbitrarily replaceable) ethical components. For example, in the logically valid inference from "Grass is green" to "Grass is green or Trump should lose the election," the disjunctive conclusion component "Trump should lose the election" is irrelevant and can be replaced by anything else without affecting validity (salva validitate). In Schurz (1997), the non-derivability of relevant ethical conclusions from descriptive premises was called the *general Hume thesis* and was proven for all multimodal logics whose axiom schemata do not contain so-called *bridge principles*, or BPs (ibid., Theorems 1, 2, 3, 7, and 8). BPs are sentence schemata that establish a connection between descriptive and ethical statements by containing the same schematic letter both inside and outside the scope of ethical operators.

1.3 The analytical dichotomy: Thirdly, it must be shown that BPs cannot be considered analytically valid meaning postulates and thus cannot be axioms for extended modal logics (Schurz 1997, Ch. 11; 2010, Sec. 3). Moore (1903, 15ff.) justified this thesis with his *open question* argument: for any descriptive condition D it can always be meaningfully asked, "but is D also good?" This shows that the good cannot be analytically reduced to a descriptive condition. Rather, the acceptance of a (substantial) BP is always linked to a certain (usually controversial) ethical position. For example, the BP of utilitarianism states that good is what maximizes the overall utility for all people (Mill 1863). This BP is criticized by alternative ethics for there being other moral intuitions not serving the maximization of utility.

However, the above applies only to so-called *substantial* BPs—these are BPs that can justify categorical value or norm statements. In contrast, so-called *functional* BPs, like the *means-end principle* discussed in Sec. 3, are analytical in nature.[6] However, with functional BPs it is provably impossible to derive *non-trivial* categorical ethical statements from descriptive premises;[7] only *hypothetical* ethical statements are derivable. This result is called the *practical Hume thesis* in Schurz (1997) and is proven for all multimodal systems axiomatized by standard axioms together with functional BPs (ibid., theorem 6, corollary 10). The practical Hume thesis is *fundamental*

[6]Logically, a functional BP is characterized by the property that it becomes a logical truth when all ethical operators are removed from it.

[7]A categorical normative conclusion from descriptive premises D is called non-trivial if D does not already entail that the prescribed (or prohibited) state of affairs is always true (or false).

to our value neutrality thesis, as it guarantees the logical possibility of the coexistence of value neutrality and means-end inferences.[8]

Ad 2 – Justifying the logical-empirical nature of the scientific method: Can there be an "objective value science" that can justify categorical ethical statements in the same objective way as empirical science justifies factual statements? Can there be moral experts who can inform people about the correct values as reliably as astronomers inform us about the stars? For this to be the case, there would have to be both substantive and intersubjectively universally valid ethical principles. However, this is doubtful. Based on a classification of ethical theories, Schurz (1997, ch. 11; 2010, Sec. 7) argues that the highest justificatory principles of these theories are either almost *empty* concerning their applications, or they are substantive but highly *controversial*. We can only summarize the most important results of this investigation here:

Utilitarianism: This position is "reductionist," meaning that the good is reduced to descriptive conditions. The fundamental principle of utilitarianism characterizes an action as good if it contributes to maximizing the overall utility for all people. However, this overall utility depends on people's empirical interests, which are culturally dependent and can change. Furthermore, the overall utility depends on the method of utility aggregation—particularly how conflicting utility values are weighed against each other, such as health versus freedom and social welfare in the example of the COVID-19 pandemic. As noted in Section 3, such trade-offs inevitably involve subjective factors.

Theories of Justice: These theories are "autonomist", meaning that they do not base the concept of the good on empirical conditions but on intuitive standards of justice. Here, too, there are different and conflicting principles of justice, particularly the principles of merit-based justice and distributive justice (see, for example, Nozick 1974 versus Rawls 1971). According to merit-based justice, distribution of wealth is considered fair if individuals have acquired their claims to property in a legitimate and equal-opportunity manner; thus, disproportionate wealth acquired through higher performance is regarded as fair. In contrast, distributive justice considers wealth, even if it was earned based on merit, to be unfair and subject to redistribution. The balancing of these two principles of justice is tied to subjective interests or intuitions and varies significantly between the conservative-liberal camp and the leftist-socialist camp.

Intuitionist-empiricist ethics: These ethical theories assume an inherent "moral sense" in humans, which relies primarily on intuitions (Hume 1939/40, 177f.; Firth 1952). The fundamental principle of these ethics characterizes a

[8]The result holds, more generally, for all conclusions of the form $\mathrm{Op}(D \to N)$ where N is a categorial ethical formula, D a possibly empty descriptive formula and Op a possibly empty sequence of quantifiers and modal operators.

state of affairs P as morally valuable if every person under normal conditions would spontaneously perceive or judge P as morally valuable. While this principle is analytically plausible, its problem lies in the fact that its scope of application is almost empty. In contrast to genuine perceptions, there are hardly any moral value judgments that are intersubjectively stable across cultures. This is convincingly confirmed for the present by the cross-cultural studies of the World Value Survey by Ronald Inglehart.[9]

In summary, fundamental categorical value judgments cannot be justified with the same claim to scientific objectivity as scientific knowledge. And if scientists occasionally make a fundamental categorical value judgment, they should, in any case, mark it as such and not present it as quasi-scientific knowledge.

6 Consequences for the debate on values in science

In this essay, we have established a form of the value neutrality thesis that differs from traditional formulations of the ideal of value freedom. We aimed to show that a harmonious coexistence of scientific value engagement and value neutrality is possible by formulating value statements hypothetically and justifying them through empirically supported means-end inferences. The method of hypothetical formulation also enabled a solution to the problem of inductive risk by disclosing error probabilities and explaining their practical implications through hypothetical cost-benefit evaluations. In this concluding section, two further issues of the recent value judgment debate are addressed through the hypothetical method.

The first of these issues is the so-called *new demarcation problem*, introduced into the debate by Wilholt (Wilholt 2009, Holman/Wilholt 2022). This is a challenge for all opponents of value neutrality. As numerous meta-studies show, many studies funded by the pharmaceutical industry suffer from a *bias* that benefits the revenues of their sponsors (and often includes the omission of negative results)[10]. The new demarcation problem now consists in the fact that proponents of scientific value dependence criticize such studies due to their capitalist value bias, while *simultaneously* considering a value bias in scientific knowledge as inherently unavoidable and even beneficial. Value proponents should, therefore, have no objections to a capitalism-friendly value bias and should regard it as equally legitimate as their own typically left-leaning value bias (see also Henschen 2021, 20). How can value proponents avoid this consequence? For example, Wilholt agrees with Douglas's thesis that, because of the AIR, extra-epistemic values inevitably influence the justification of scientific statements, but at the same time, he claims that scientists themselves establish *conventional standards*

[9] Inglehart (1997) and the website of the World Values Survey.

[10] See Wilholt (2009, 93f.), vom Saal and Hughes (2005), and Brown (2008, 191).

to distinguish legitimate from illegitimate value influences (2009, 96–99). However, Wilholt leaves us unclear about the justification of these standards, so it is not obvious why religious creationists or proponents of capitalism could not also establish standards that serve their own perspectives.

Some authors have suggested that the new demarcation problem could be solved by scientists relying on certain preferred values—namely, *democratic* values that form the foundation of our society.[11] Hilligardt (2023), however, argues that this would exclude "partisan science," which represents the values of socially underrepresented groups, such as *feminist* values. Instead, Hilligardt advocates a *pluralism* of political purposes within the sciences. A more grave objection to the democratic value dependence thesis is the *polarization objection* raised by Le Bihan (2024). Modern democratic societies are *polarized*: the politically right-leaning and left-leaning segments of the population hold opposing value preferences on many issues. The suggestion to rely on democratic values does little to address this issue, as it leaves open which side of the political spectrum these "democratic values" should be on. Upon closer examination, this suggestion even has frightening implications. For it would mean that if a political shift occurs in a democratic society, scientific knowledge and textbooks would need to be *rewritten*, as they depend on the "prevailing democratic values," which may have changed significantly at that moment. For example, if there were a shift in the U.S. from Democrats to Republicans, wouldn't this position suggest to close or reassign certain professorships and potentially entire departments? I doubt that opponents of value neutrality have such outrageous consequences in mind with their arguments. History has seen plenty of unwelcome examples of this kind. Fortunately, our society's understanding of science has largely freed itself from such political constraints, and it should remain that way.

Moreover, empirical studies show that the average political attitudes of university academics are significantly more left-leaning compared to those of non-academic populations (Duarte et al. 2015). Therefore, scientists are even more urged to frame recommendations based on their own values hypothetically rather than categorically, to avoid the risk of losing the trust of the non-academic public. This leads us to a second point of discussion in the recent controversy: the issue of *trust in science*. There is consensus that in our information society, public trust in science is fundamentally important. Proponents of scientific value influences have raised the question of whether transparency regarding these value influences would be beneficial or detrimental to public trust in science (Intemann 2024, 3). While some philosophers of science provide an affirmative answer (e.g., Elliott 2017, NAS 2018, Intemann 2024), others respond more skeptically. Some value-friendly philosophers of science (e.g., John 2018) have even argued that it is better to

[11] E.g., Schroeder (2021), Alexandrova/Fabian (2022), and Elliott (2017).

leave non-experts, who "naively" believe in the value-free nature of scientific knowledge, in their false belief, since revealing its actual value-dependence would undermine their trust in science. From our perspective, this argument is *doubly misguided*: first, the understanding of science held by these "naive non-experts" seems closer to the truth than that of their "intellectual critics"; and second, it is cynical to propose that the public should be kept in a false belief in order to better control it. In earlier times, religious leaders who were never really concerned with God but only with power argued in similar ways.

Intemann (2024, 8–10) also argues that it is ultimately a fraud when scientists hide their value assumptions from the public and present their findings as value-free knowledge. We agree with Intemann, but further emphasize that hypothetical formulations of recommendations are likely to best promote trust in science, whereas categorical recommendations are more likely to undermine this trust, especially when they oppose public interests. This has been demonstrated, among other instances, during the Covid-19 pandemic. Studies by Angeli et al. (2021) show that many non-experts lost trust in Covid experts when they realized that their recommendations were based on one-sided value preferences, such as preventing hospital overcrowding at the expense of personal freedoms. In Germany, trust in science decreased from 2022 to 2023 by 25% in the lower and middle education levels.[12] Research by Elliott et al. (2017, 13) indicates that the disclosure of values by experts either increases or decreases public trust, depending on whether the public shares those values. This is a dilemma that can only be resolved by the method of hypothetical recommendations, which protects experts from value assignments and ideological assumptions without depriving them of their usefulness to people.

In conclusion, I attempted to demonstrate that tying scientific knowledge to extra-epistemic values, as advocated by opponents of value neutrality, is not the right approach to integrating value engagement into science; rather, it ultimately leads to a dead end. The only way I see to combine the widely desired objectivity of science with value engagement and social responsibility is through the proposed method of hypothetical value statements and their justification through scientific means-ends inferences. Instead of adhering to extra-epistemic values, scientists should take *pride* in the objectivity and impartiality of their findings. At their core, these findings are neither capitalist nor communist, patriarchal nor feminist, but merely more or less true or probable. Only value-neutral truth orientation allows for stable scientific progress through the various stages of human cultural development.

[12]Wissenschaft im Dialog GmbH, Wissenschaftsbarometer 2023, Berlin 2023.

Acknowledgement. For valuable comments I am indebted to Wolfgang Spohn, Martin Carrier, Torsten Wilholt, Paul Hoyningen-Huene, Charles Pigden, Tobias Henschen, Christian Feldbacher-Escamilla and Alexander Christian.

References

Abramowitz, A. und Saunders, K. (2008): "Is Polarization a Myth?", *The Journal of Politics* 70, 542–555.

Albert, H., and Topitsch, E. (eds, 1971): *Werturteilsstreit*, Wissenschaftliche Buchgesellschaft.

Alexandrova, A., and Fabian, M. (2022): "Democratising Measurement: Or Why Thick Concepts Call for Coproduction", *European Journal of Philosophy of Science* 12: 7.

Anderson, E. (2004): "Uses of Value Judgments in Science: A General Argument, with Lessons from a Case Study of Feminist Research on Divorce", *Hypatia* 19, 1–24.

Andrade, C. (2020): "Mean Difference, Standardized Mean Difference (SMD), and Their Use in Meta-Analysis", *The Journal of Clinical Psychiatry* 81(5): 20f13681.

Angeli, F., Camporesi, S. and Dal Fabbro, G. (2021): "The COVID-19 Wicked Problem in Public Health Ethics: Conflicting Evidence, or Incommensurable Values?", *Humanities and Social Sciences Communications* 8, 1–8.

Betz, G. (2017): "Why the Argument from Inductive Risk Doesn't Justify Incorporating Non-epistemic Values in Scientific Reasoning", in: K. Elliott und D. Steel (eds.), *Current Controversies in Values and Science*, Routledge, 94–110.

Brown, J. (2008): "The Community of Science", in: M. Carrier, D. Howard, and J. Kourany (eds., 2008), *The Challenge of the Social and the Pressure of Practice*, University of Pittsburgh Press, 189–216.

Clough, S. (2008): "Solomon's Empirical/Non-Empirical Distinction and the Proper Place of Values in Science", *Perspectives on Science* 16, 265–279.

Cox, D. and Hinkley D. (1974): *Theoretical Statistics*, Chapman & Hall, 49, 209.

Doppelt, G. (2007): "The Value Ladenness of Scientific Knowledge", in: Kincaid et al. (2007), 188–218.

Douglas, *H*. (2000): “Inductive Risk and Values in Science”, *Philosophy of Science* 67, 559–579.

Douglas, *H*. and Branch, T.Y. (2024): “The Social Contract for Science and the Value-free Ideal”, *Synthese* 203, 40.

Douglas, *H*. (2009): *Science, Policy, and the Value-free Ideal*, University of Pittsburgh Press.

Duarte, J. L. et al. (2015): “Political Diversity Will Improve Social Psychological Science”, *Behavioral and Brain Sciences*, 38, e130.

Dupré, J. (2007): “Fact and Value”, in: Kincaid et al. (2007), 28–41.

Elliott, K. (2017): *A Tapestry of Values. An Introduction to Values in Science*, Oxford University Press.

Elliott, K. et al. (2017): “Values in Environmental Research: Citizens’ Views of Scientists Who Acknowledge Values”, *Plos One* 12, e0186049 (doi: 10.1371/journal.pone.0186049).

Firth, R. (1952): “Ethical Absolutism and the Ideal Observer”, *Philosophy and Phenomenological Research* XII/3, 317–345.

Frankena, W. (1963): *Ethics*, Prentice Hall.

Habermas, J. (1968): “Erkenntnis und Interesse”, Engl. transl. in J. Habermas, *Knowledge and Human Interests*, Polity Press 1987, Appendix.

Hare, R. (1981): *Moral Thinking*, Clarendon Press.

Hempel, Carl G. (1965): “Science and Human Values”, in: Carl G. Hempel: *Aspects of Scientifc Explanation*, Free Press, New York, pp. 81–96.

Henschen, T. (2021): “How Strong is the Argument From Inductive Risk?”, *European Journal for Philosophy of Science* 11: 92

Hilligardt, *H*. (2023): “Partisan Science and the Democratic Legitimacy Ideal”, *Synthese* 202: 135.

Holman, B., and Wilholt, T. (2022): “The New Demarcation Problem”, *Studies in History and Philosophy of Science* 91, 211–220.

Hume, D. (1739/40): *A Treatise of Human Nature. Vol. II/Book III: Of Morals*, Clarendon 1896.

Inglehart, R. (1997): *Modernization and Postmodernization: Cultural, Economic, and Political Change in 43 Societies*, Princeton University Press, Princeton.

Intemann, K. (2024): "Value Transparency and Promoting Warranted Trust in Science Communication", *Synthese* 203: 42.

Ioannidis, J. (2005): "Why Most Published Research Findings are False", *PLoS Medicine* 2(8), e124, 696–701.

Jeffrey, R. (1956). "Valuation and Acceptance of Scientific Hypotheses", *Philosophy of Science* 22, 337–346.

John, S. (2015): "Inductive Risk and the Contexts of Communication", *Synthese* 192, 79–96.

John, S. (2018): "Epistemic Trust and the Ethics of Science Communication: Against Transparency, Openness, Sincerity and Honesty", *Social Epistemology* 32, 75–87.

Kant, I. (1785): *Groundwork for the Metaphysics of Morals* (tr. T. K. Abbott, ed. L. Denis), Broadview Press 2005.

Kincaid. H., Dupré, J., and Wylie, A. (eds, 2007): *Value-free Science?*, Oxford University Press.

Lacey, *H*. (2013): "Rehabilitating Neutrality", *Philosophical Studies* 163, 77–83.

Le Bihan, S. (2024): "How Not to Secure Public Trust in Science: Representative Values vs. Polarization and Marginalization", appears in *Philosophy of Science.*

MacIntyre, A. (1981): *After Virtue. A Study in Moral Theory*, Dickworth.

Mill, J. S. (1863): *Utilitarianism*, Parker, Son, and Bourn.

Moore, G. E. (1903): *Principia Ethica*, Cambridge Univ. Press.

NAS (National Academies of Sciences) (2018): "Open Science by Design: Realizing a Vision for 21st Century Research", *The National Academies Press.*

Nozick, R. (1974): *Anarchy, State and Utopia*, Basil Blackwell, Oxford.

Parker, W. (2024): "The Epistemic Projection Approach to Values in Science", *Philosophy of Science* 91, 18–36.

Pigden, C. R. (1989): "Logic and the Autonomy of Ethic", *Australasian Journal of Philosophy* 67, 127–151.

Prior, A. (1960): "The Autonomy of Ethics", *Australasian Journal of Philosophy* 38, 199–206.

Putnam, *H*. (2002): *The Collapse of the Fact/Value Dichotomy*, Harvard University Press.

Rawls, J. (1971): *A Theory of Justice*, Cambridge Univ. Press, Cambridge.

Reichenbach, *H*. (1956): *The Direction of Time*, Univ. of California Press, Berkeley.

Rudner, R. (1953): "The Scientist qua Scientist Makes Value Statements", *Philosophy of Science* 20, 1–6.

Russell, G. (2022): "How to Prove Hume's Law", *Journal of Philosophical Logic* 51, 603–632.

Scheler. M. (1913/16): *Der Formalismus in der Ethik und die materiale Wertethik* (hg. C. Bermes), Felix Meiner Verlag 2014.

Schmidt, P. F. (1971): "Ethische Normen in der wissenschaftlichen Methode", in: Albert/Topitsch (eds., 1971), 353–364.

Schooler, J. (2011): "Unpublished Results Hide the Decline Effect", *Nature* 470, 437.

Schroeder, S. (2021): "Democratic Values: A Better Foundation for Public Trust in Science", *The British Journal for the Philosophy of Science* 72, 545–562.

Schurz, G. (1997): *The Is-Ought Problem. An Investigation in Philosophical Logic*, Kluwer 1997.

Schurz, G. (2010): "Non-Trivial Versions of Hume's Is-Ought Thesis and Their Presuppositions", in: Pigden, C. R. (ed., 2010): *Hume on 'Is', and 'Ought'*, Palgrave Macmillan, 198–216.

Schurz, G. (2013): "Wertneutralität und hypothetische Werturteile in den Wissenschaften", in: Schurz, G. and Carrier, M. (Eds., 2013), *Werte in den Wissenschaften*, Suhrkamp, 305–334.

Schurz, G. (2014): *Philosophy of Science: A Unified Approach*, Routledge, New York.

Schurz, G., and Gebharter, A. (2016): "Causality as a Theoretical Concept", *Synthese* 193 (4), 1071–1103.

Searle, J. (1964): "How to Derive 'Ought' from 'Is'", *Philosophical Review* 73, 43–58.

Spirtes, P., Glymour, C., and Scheines, R. (2000): *Causation, Prediction, and Search*, MIT Press.

Steele, K. (2012): "The Scientist Qua Policy Advisor Makes Value Judgements", *Philosophy of Science* 79, 893–904.

Strauss, L. (1971): "Die Unterscheidung von Tatsachen und Werten", in: Albert (Topitsch (eds., 1971), 73–91.

Vom Saal, F. S., and Hughes, C. (2005): "An Extensive New Literature Concerning Lowdose Effects of Bisphenol A Shows the Need for a New Risk Assessment", *Environmental Health Perspectives* 113, 926–933.

Weber, M. (1917): "The Meaning of 'Ethical Neutraliy' in Sociology and Economics" (Engl. trans.), in: *Max Weber in the Methodology of the Social Sciences* (ed. E. A. Shils and H. A. Finch), Free Press 1949, 1–49.

Wiesing, U. et al. (2021): "Wissenschaftliche (Politik-)Beratung in Zeiten von Corona: Die Stellungnahmen der Leopoldina zur Covid-19-Pandemie", in: *Ethik und Gesellschaft* 1 (doi: 10.18156/eug-1-2021-858).

Wilholt, T. (2009): "Bias and Values in Scientific Research", *Studies in History and Philosophy of Science* 40, 92–101.