SCIENCE, VALUES, AND THE PRIORITY OF EVIDENCE

P.D. MAGNUS

ABSTRACT: It is now commonly held that values play a role in scientific judgment, but many arguments for that conclusion are limited. First, many arguments do not show that values are, strictly speaking, indispensable. The role of values could in principle be filled by a random or arbitrary decision. Second, many arguments concern scientific theories and concepts which have obvious practical consequences, thus suggesting or at least leaving open the possibility that abstruse sciences without such a connection could be value-free. Third, many arguments concern the role values play in inferring from evidence, thus taking evidence as given. This paper argues that these limitations do not hold in general. There are values involved in every scientific judgment. They cannot even conceivably be replaced by a coin toss, they arise as much for exotic as for practical sciences, and they are at issue as much for observation as for explicit inference.

KEYWORDS: values and science, inductive risk, ampliative risk, epistemic values, evidence, externality

Introduction

Recent philosophical literature on science and values has shown numerous ways in which science is not and could not be value-free.[1] The point is not just that scientific judgment respects the value of evidence and theoretical virtues, but also that it reflects whether certain outcomes would be good or bad. Nevertheless, arguments for this often yield only a limited conclusion.

First, some arguments appeal to values as a means of choosing theories in the face of empirical underdetermination. Yet, as critics note, it does not need to be values that fill the gap. In principle, a scientist faced with underdetermination between two options might instead just flip a coin. Call this the *randomizer reply*.

Second, arguments for the value-laden nature of science often apply to some but not all science. For example, Anna Alexandrova argues that sciences of well-

[1] Elliott provides a systematic discussion of different ways that science and values may be connected; see Kevin C. Elliott, *A Tapestry of Values: An Introduction to Values in Science* (Oxford: Oxford University Press, 2017).

© LOGOS & EPISTEME, IX, 4 (2018): 413-431

being (ones which make claims about health, for example) necessarily and legitimately reflect normative judgments.[2] Hilary Putnam similarly argues that sciences which employ thick concepts undercut the distinction between fact and value.[3] These arguments do not apply, nor are they meant to apply, to abstruse sciences like astronomy or particle physics. The concept of the neutrino, for example, does not seem to have any intrinsically normative dimension. Call this the *policy-relevance restriction*.

Third, many arguments show only that values enter into inferences from evidence to an underdetermined conclusion. The evidence itself is taken as given. The arguments, Kevin Elliott and Daniel McKaughn note, "incorporate nonepistemic values only as a secondary consideration for resolving epistemic uncertainty."[4] Matthew Brown decries these arguments for bringing in values too late. Adopting Brown's phrase, call this limitation the *lexical priority of evidence* over values.[5]

In sections 2 and 3, I consider several arguments that values necessarily enter into scientific inference. These arguments, as they are usually posed, treat evidence and values as separate inputs to the process. In section 4, I consider the role of values in scientific observation itself. Heather Douglas provides a clear example in which scientists had to judge whether prepared slides showed tumors or not.[6] The judgments were significant for environmental regulation, and so values were involved. Douglas argues that similar entanglement would not obtain in science without policy implications. In sections 5-7, I argue that the connection between scientific judgment and values which is typified in ampliative inference and Douglas' example holds for all scientific observation. Scientists always have a choice about how to state an observation, between more significant but riskier formulations and less significant but safer ones. This choice always involves weighing values in the sense of the goodness or badness of various possible

[2] Anna Alexandrova, "Can the Science of Well-Being Be Objective?" *The British Journal for the Philosophy of Science* 69, 2 (2018): 421–445.

[3] Hilary Putnam, *The Collapse of the Fact/Value Dichotomy and Other Essays* (Cambridge, Massachusetts: Harvard University Press, 2004).

[4] Kevin C. Elliott and Daniel J. McKaughn, "Nonepistemic Values and the Multiple Goals of Science," *Philosophy of Science* 81, 1 (2014): 2.

[5] Matthew J. Brown, "Values in Science Beyond Underdetermination and Inductive Risk," *Philosophy of Science* 80, 5 (2013): 829-839.

[6] Heather E. Douglas, "Inductive Risk and Values in Science," *Philosophy of Science* 67, 4 (2000): 559-579.

outcomes. This weighing could not be done without values (contra the randomizer-reply), it occurs in all science (contra the policy-relevance restriction), and it occurs in the very formulation of the evidence (contra the lexical priority of evidence).

Underdetermination and Tie-Breaking

Some arguments connect science and values in cases where the usual standards of evidence are insufficient to decide among competing hypotheses. The arguments hold that, in such cases, scientists may responsibly select the hypothesis which best accords with their value commitments. Call this the *tie-breaker argument*. It applies only to cases where hypotheses score equally well with respect to the evidence, and values enter only to break the tie.[7]

A more subtle version of the tie-breaker argument is given by Helen Longino.[8] She notes that connecting scientific theories to observable phenomena almost always requires auxiliary hypotheses, an aspect of underdetermination sometimes called the Duhem-Quine Problem. Values enter by way of value-laden auxiliary hypotheses, when matters cannot be settled by observation and value-neutral auxiliary hypotheses. Like the less-subtle tie-breaker argument, values are in play only when evidence does not determine theory choice.

Let's consider two standard replies to the tie-breaker argument.

One response is to claim that scientists could break ties by using a randomizer instead of making value judgments. This response has a long history, although it seems to have been suggested independently by different thinkers. Otto

[7] Magnus and Longino situate these arguments in a broader analysis of 'underdetermination;' see P.D. Magnus, "Underdetermination and the Claims of Science" (PhD diss., University of California, San Diego, Department of Philosophy, 2003) and Helen Longino, "Underdetermination: A Dirty Little Secret," *STS Occasional Papers* 4 (2016), Department of Science and Technology Studies, University College London.

[8] Helen Longino, *Science as Social Knowledge* (Princeton, NJ: Princeton University Press, 1990) and "Underdetermination."

Neurath suggests drawing lots.[9] Gregor Betz suggests rolling a die.[10] Inmaculada de Melo-Martín and Kristen Intemann suggest flipping a coin.[11]

The idea of this randomizer reply is that underdetermination only shows that something besides evidence alone must determine theory choice. Although we could follow our value commitments to select preferable theories or auxiliary hypotheses, we could instead use a procedure that is independent of our values. Randomly selecting one theory over its competitors (or one set of auxiliary hypotheses over alternate sets) makes the choice without regard to which would be better or which we would prefer.

Nevertheless, adopting such a policy would be a practical decision. As an analogy, consider a mundane case in which I cannot decide which of two restaurants to visit for lunch and so flip a coin. My values and practical reasons have no influence over the outcome of the coin toss, of course, and so the selection is value-free *to that extent*. However, my values and preferences are involved in my decision to use coin-flipping as a way of resolving the choice. I want to go to lunch at one of two places, and I do not want to spend too much time or energy deciding. If someone asks why I went to one restaurant rather than the other, a complete answer would refer not just to the random process but also to the reasons I had for adopting that method. Similarly, deciding to flip a coin in the face of underdetermination would be practical and value-driven. Breaking ties by flipping coins would keep values from directly deciding specific winning hypotheses, but it would not ultimately escape the intrusion of values and practical decisions into theory choice.[12] Just as values might lead us to prefer a specific outcome or some auxiliary hypotheses, values might lead us to choose a random method. Choosing to believe *the outcome determined by the coin toss* when evidence itself underdetermines theory choice is still a value-driven decision. So the randomizer reply fails.

---

[9] Otto Neurath, "The Lost Wanderers of Descartes and the Auxiliary Motive: On the Psychology of Decision," in *Philosophical Papers: 1913-1946*, eds. Robert S. Cohen and Marie Neurath (Dordrecht: D. Reidel, 1983 [1913]), 1-12.

[10] Gregor Betz, "In Defence of the Value Free Ideal," *European Journal for the Philosophy of Science* 3, 2 (2013): 210.

[11] Inmaculada de Melo-Martín and Kristen Intemann, "The Risk of Using Inductive Risk to Challenge the Value-Free Ideal," *Philosophy of Science* 83, 4 (2016): 505.

[12] This is one way to read Neurath's argument: Neurath thinks that we would prefer an epistemic culture that does not let our preferences directly decide theory choice. So, he argues, we should draw lots. Preferences are still at work, at the level of general policy.

A second response to the tie-breaker argument is to insist that scientists should never break ties with anything but further evidence. Rather, scientists should be agnostic when evidence is insufficient to decide between rival hypotheses. They should pursue each hypothesis until evidence is uncovered which breaks the tie without any appeal to values. Call this the *wait-and-see reply*.

In answer to the wait-and-see reply, note that remaining agnostic and collecting further evidence is not always possible. As a practical matter, we cannot pursue every hypothesis. Pursuing hypotheses that figure in separate research programs may require different techniques, and so different investments in training and equipment. Although "thought experiments can be risked without hesitation," Neurath notes, it is not possible "in the same way, to train for more than one career."[13] We could organize the scientific community so that different scientists were trained to pursue different research programs, but there is a limit to how many scientists can be trained and how many laboratories can be outfitted. Moreover, waiting for more evidence takes time. Brown argues that there are cases in which "we cannot wait for the end of inquiry for scientists to accept or reject a hypothesis, we cannot depend on anyone else to do it, and we must contend with uncertainty and underdetermination [... S]cientists find themselves in the business of accepting and rejecting hypotheses in such conditions."[14] Even if we have the option to wait for compelling evidence, the cost of waiting might be higher than the probable cost of using quicker but less-reliable methods.[15]

One may still insist that it is appropriate (insofar as possible) to respond to underdetermination by remaining agnostic and collecting more evidence. This would limit the scope of the tie-breaker argument to cases where agnosticism is impractical. It would still be a regulative ideal to abide with unbroken ties until further evidence could be uncovered to change the score.

Retreating to agnosticism in this way would avoid using values to break ties between rival theories, but it would require that the evidence itself be value-free. Brown observes that the tie-breaker argument "begin[s] from a situation where the evidence is fixed and take[s] values to play a role in the space that is left over."[16] The underdetermination of theory by data, in its very formulation, concerns the inference to theory once data are given. It presumes that evidence comes first,

[13] Neurath, "The Lost Wanderers of Descartes," 3.
[14] Brown, "Values in Science," 831-2.
[15] Elliott and McKaughn, "Nonepistemic Values."
[16] Brown, "Values in Science," 834.

before the question arises of whether values should be involved. Brown calls this presumption the *lexical priority of evidence over values*. He explains, "lexical priority means that evidence will always trump values."[17] The picture is one on which evidence and values might independently underwrite our preferring one theory over another, and the lexical priority of evidence means that considerations of evidence always override considerations of values. The preferences underwritten by values only hold sway when evidence is silent and we are forced to make a choice.

The wait-and-see reply works to limit the conclusion of the tie-breaker argument in two respects: First, values are allowed to play a role only for questions of practical importance but not for abstruse matters (the policy-relevance restriction). Second, values play a role only after the evidence itself is given (the lexical priority of evidence).

Ampliative Risk and Our Duties as Knowers

In drawing an inference from evidence, there is inevitably a tension between striving to believe the truth and striving to avoid error. Scientists might be quick to judge or more cautious. If they are too quick to judge, they risk believing in error; such a result is a *false positive* or *type I* error. If they remain agnostic, they risk the opportunity cost of not having an accurate belief that they could have had; such a result is a *false negative* or *type II* error. This tension is resolved only by assessing what the cost would be of each possible error—that is, by reckoning with values. Therefore, values enter into scientific inference. Call this the *ampliative risk argument*.[18]

The impetus to believe truth drives us to adopt the claim best supported by evidence, but the impetus to avoid error drives us to wait and demand more evidence. William James puts it in histrionic terms: "Believe truth! Shun error!—these, we see, are two materially different laws; and by choosing between them we

---

[17] Brown, "Values in science," 836.

[18] It is standard, following Hempel, to call this the argument from inductive risk. I've opted for the label 'ampliative risk' because 'induction' is ambiguous between a narrow use (enumerative induction) and a broad use (ampliative inference) (Carl G. Hempel, "Science and Human Values," *in Aspects of Scientific Explanation and other Essays in the Philosophy of Science* (New York: The Free Press, 1965), 92). See also Kevin C. Elliott and Ted Richards, *Exploring Inductive Risk* (Oxford: Oxford University Press, 2017).

may end up coloring differently our whole intellectual life."[19] A similar point is made by Richard Rudner, who writes that "our decision regarding the evidence... is going to be a function of the *importance*, in the typically ethical sense, of making a mistake in accepting or rejecting the hypothesis."[20] Heather Douglas puts the point in less grandiose terms: "Within the parameters of available resources and methods, some choices must be made, and that choice should weigh the costs of false positives versus false negatives. Weighing these costs legitimately involves social, ethical, and cognitive values."[21]

The tie-breaker argument was (at least partially) defused because not all ties need to be broken. We could wait and see. The argument here turns on the fact that waiting and seeing would itself be a choice, with benefits but also costs for our overall system of belief. As a result, values are always involved in assessing "the sufficiency of evidence, the weighing of uncertainty, and the consequences of error."[22] Douglas calls this an *indirect role* for values.

Elsewhere, I have called this the James-Rudner-Douglas or JRD thesis: "Anytime a scientist announces a judgment of fact, they are making a tradeoff between the risk of different kinds of error. This balancing act depends on the costs of each kind of error, so scientific judgment involves assessments of the value of different outcomes."[23] The JRD thesis underwrites the ampliative risk argument but, as I will argue below, it is broader and more fundamental.

Note that one could not get around the JRD thesis by flipping a coin in cases where the evidence is equivocal. Every ampliative inference involves reckoning with the risks of different kinds of error. Errors are always logically possible, precisely because the inference is ampliative. It is possible for all of the premises to be true, but for the conclusion still to be false. So one would be flipping coins always and for everything.

---

[19] William James, "The Will to Believe," in *Essays in Pragmatism*, ed. Alburey Castell (New York: Hafner Publishing Co., 1948), 100.

[20] Richard Rudner, "The Scientist qua Scientist Makes Value Judgments," *Philosophy of Science* 20, 1 (1953): 2.

[21] Heather E. Douglas, *Science, Policy, and the Value-free Ideal* (Pittsburgh: University of Pittsburgh Press, 2009), 104.

[22] Douglas, *Science, Policy,* 103.

[23] P.D. Magnus, "What Scientists Know Is Not a Function of What Scientists Know," *Philosophy of Science* 80, 5 (2013): 845. See also P.D. Magnus, "Science and Rationality for One and All," *Ergo* 1, 5 (2014): 129-138.

The ampliative risk argument as it is typically posed presumes that scientific inference leads to accepting or rejecting hypotheses.[24] The costs and benefits which scientists are supposed to reckon with are conditional on making the right or wrong choice. So a standard rebuttal is to deny that scientists should ever be flatly accepting or rejecting hypotheses.

Gregor Betz argues that scientists should report hedged hypotheses.[25] The idea is that, instead of reporting a categorical result 'H,' scientists should instead report something like 'Evidence strongly suggests but does not decisively confirm H.' Notice, however, that hedged claims still express a degree of confidence. Scientists must decide how much to hedge. Less hedging is riskier, but more hedging courts triviality. In order to escape all danger of being wrong, scientists might refuse to make any claims at all. As Stephen John observes, the price of that epistemic security would be "policy-impotence," having nothing to say that could be of any use to policy-makers or anyone else.[26] Scientists should, to use Betz's phrase, "simply admit their complete ignorance" in cases where they have no evidence whatsoever—but it would be pathological for them to plead complete ignorance just so as to avoid any chance of being wrong.[27]

Responding to Rudner, Richard Jeffrey argues that scientists should only report probabilities.[28] Even though this allows scientists to avoid deciding for or against H, they must still decide for or against the claim that the probability of H is p. A scientist might hedge this report by returning an interval p±e rather than a precise probability p. As the size of the interval is larger, the claim is safer but less useful in guiding action. The limit case, reporting that the probability of H is between 0 and 1, is a useless tautology. It does not even help to assume, as Bayesians sometimes do, that a scientist always has some degree of belief p in every hypothesis H. Since this could reflect their prior credence more than the weight of evidence, a scientist must still judge that it reflects enough evidence to merit reporting.

---

[24] For the argument formulated in terms of acceptance and rejection, see (e.g.) Rudner, "The Scientist qua Scientist;" Hempel, "Science and Human Values," 92-3; Brown, "Values in Science."
[25] Betz, "In Defence."
[26] Stephen John, "The Example of the IPCC Does Not Vindicate the Value Free Ideal: A Reply to Gregor Betz," *European Journal for the Philosophy of Science* 5, 1 (2015): 9.
[27] Betz, "In Defence," 9.
[28] Richard Jeffrey, "Valuation and Acceptance of Scientific Hypotheses," *Philosophy of Science* 23, 3 (1956): 237-246.

This reply is given already by Rudner, who argues that a hedged report is itself "nothing more than the acceptance by the scientist of the hypothesis that the degree of confidence is p or that the strength of evidence is such and such."[29] The question is precisely what the benefit would be of making a less-hedged, more-confident report (if it were true) and what the cost would be (if it were false). Settling on any particular conclusion, even if it is a conclusion about probabilities or the weight of evidence, is subject to ampliative risk. The JRD thesis applies, and so values play a role.

Brown objects that the ampliative risk argument (like the tie-breaker argument) presumes the lexical priority of evidence over values.[30] Note, however, that there can be no question here of evidence trumping values. The JRD thesis means that there can be no scientific conclusion without (at least implicitly) weighing the costs of various possible errors. There is no choice favored by the evidence alone, so *a fortiori* it makes no sense for that choice to presumptively win out.[31]

The argument from ampliative risk does not yield the blanket conclusion that, as Stijn Conix puts it, "it does not make sense to think of values and epistemic standards as taking priority over each other."[32] It makes sense to think of some values that way. If we rank theories according to how well they promote a conception of human autonomy and according to how simple they are, for example, then these separate rankings might pick out different theories as best. The values could give separate preference orderings. The point of the JRD thesis is that there are *some* values which cannot be separated in this way. The enthusiasm to reach for a possibly true belief or the risk-aversion which makes one remain agnostic even as evidence accumulates—these values do not rank the possible

[29] Rudner, "The Scientist qua Scientist," 4, emphasis in original.

[30] Matthew J. Brown, "Values in Science." Brown calls the tie-breaker argument "the gap argument," and the ampliative risk argument is his "error argument."

[31] ChoGlueck argues that the error argument (the argument from ampliative risk) is just a special case of the gap argument (the tie-breaker argument). This is directly rebutted by the fact that the tie-breaker argument is vulnerable to Brown's worry about the lexical priority of evidence in a way that the argument from ampliative risk is not. Christopher ChoGlueck, "The Error Is in the Gap: Synthesizing Accounts for Societal Values in Science," *Philosophy of Science* 85, 3 (2018): 704-725.

[32] Stijn Conix, "Radical Pluralism, Ontological Underdetermination, and the Role of Values in Species Classification" (PhD diss., University of Cambridge, Queen's College, Department of History and Philosophy of Science, 2017), 102.

choices apart from epistemic standards. Enthusiasm may make a scientist respond to preliminary evidence with belief when their colleagues demand more evidence, but it is precisely a disagreement about how much evidence is *enough*. Risk-aversion will delay accepting a belief as evidence accumulates, but only an utter sceptic would refuse to believe regardless of how much evidence there might be. These values are entangled with the application of epistemic standards, so it makes no sense to think of one as taking priority over the other.

A different way to construe the complaint about lexical priority is that the evidence has been presumed to be value-free. Although evidence and the costs of possible errors both enter into the epistemic calculation, they do so as independent variables. One may complain, as Elliott and McKaughn do, that the JRD thesis adds "nonepistemic values only as a secondary consideration."[33]

This objection is especially apt when the ampliative risk argument is posed in terms of type I and type II errors. These terms come from statistical hypothesis testing, where the problem is to specify a rule for accepting or rejecting hypotheses given a data set. The data set itself is not at issue. This construal of the argument is encouraged by Rudner's insistence that "every scientific inference is properly construable as a statistical inference."[34] It is also encouraged just by posing the argument in terms of *inductive risk*. 'Induction' and 'inductive inference' are often used narrowly to pick out inference from a sample to a population or from a finite track-record to a generalization. From given observations, the inductive problem is how to generalize or draw conclusions. Evidence is lexically and literally prior.

Of course, James and Douglas do not pose the argument in terms of statistical inference. Their arguments apply to ampliative inference generally, rather than just to inductive inference narrowly-construed. Nevertheless, posing the argument in terms of *inferential risk* makes it turn on the move from evidence to hypothesis. Douglas specifies that values, in an indirect role, "determine the importance of the inductive gaps left by the evidence."[35] Even though evidence does not recommend a conclusion without some values, the evidence is in a sense primary.

---

[33] Elliott and McKaughn, "Nonepistemic Values," 2.

[34] Rudner, "The Scientist qua Scientist," 3.

[35] Douglas, *Science, Policy,* 96.

Evidence That Matters for Policy

Douglas offers an example that rebuts this as a general worry. In the 1970s, slides of rat livers were prepared as part of a study of dioxin toxicity. These slides were evaluated by different teams of scientists over more than a decade, and different numbers of liver tumors were reported in the different evaluations. Douglas writes,

> Although not as formal as setting a level for statistical significance, the pathologists must be similarly concerned with false positives and false negatives. Suppose a pathologist chooses to take all borderline cases and judge them to be non-cancerous lesions. ... The consequences for such an approach will be an underestimation of malignancies and thus an underestimation of risk.[36]

These slides were revisited again and again precisely because the study was "important in regulation," so that an estimate of lower risk would "likely lead to a relaxed regulation" which could "cause increased harm to the public." Conversely, judging borderline cases as malignant would have erred on the side of protecting public health "at the economic costs of potentially unnecessary regulation."[37]

In the spirit of Jeffrey and Betz, one might note that scientists could reject the requirement that slides be sorted decisively into those that showed acute toxicity and those that did not. However: Although scientists could emphasize their uncertainty and the tentativeness of their conclusions to different degrees, there was no neutral way of relaying the objective situation to policymakers. Whatever report scientists gave, even refusing to report at all, would have consequences. So hedging or reporting confidence intervals could not escape the practical significance of reporting their results in one way rather than another.

This is kind of an easy case, though. Scientists knew that their observations would have consequences for regulation and public health. Their observations had a clear valence in practical and ethical terms. Douglas herself notes this policy-relevance restriction in the scope of her argument. She writes that

> there are some areas of science where making a wrong choice has no impact on anything outside of that area of research. One may think, for example, of research into the coherence properties of atom beams. It is very difficult to fathom how errors in such research could have non-epistemic consequences. Hence, scientists doing such research need not consider non-epistemic values.[38]

---

[36] Douglas, "Inductive risk," 571.
[37] Douglas, "Inductive risk," 571.
[38] Douglas, "Inductive risk," 577.

Lots of science does not have any direct practical consequences or foreseeable application. Douglas claims that, in such cases, decisions can be made on strictly epistemic grounds. In the next section, I argue that this concedes too much. Even in exotic sciences like particle physics, the JRD thesis applies.

Observation and Externality

Following Trevor Pinch, we can distinguish possible observation reports by their degree of *externality*.[39] Externality is the inverse of immediacy—that is, an observation posed at a lower degree of externality is in more direct terms.

A high externality report is riskier but potentially more significant. A lower externality report, in contrast, is safer but less interesting. A scientist initially reports their observation at some level of externality. If their report is challenged, they can redescribe the observation at a lower level of externality and offer an argument which takes the lower-externality report as a premise and yields the higher-externality report as a conclusion.

Pinch gives the example of Ray Davis' work to detect solar neutrinos in the 1960s. The detection was a complicated operation. A large tank of tetrachloroethylene, stored in an abandoned mine shaft, was used as a target. Some of the chlorine atoms interacted with solar neutrinos to produce an isotope of argon (argon-37). The accumulated argon-37 was extracted from the tank and was measured based on its characteristic decay. Finally, there were some outputs from instruments—Pinch refers to them as "splodges."

The outcome of the work could be reported at different levels of externality. From higher to lower externality, the report might be a claim about:

- The rate of particular reactions in the sun
- Neutrinos generated in the sun
- Neutrinos arriving at Earth
- Argon-37 atoms in the tank
- Splodges on the apparatus

This list reflects the levels parsed out by Pinch, except that he groups the second and third together just as "Solar neutrinos."[40] The distinction between

[39] Trevor Pinch, "Towards an Analysis of Scientific Observation: The Externality and Evidential Significance of Observational Reports in Physics, " *Social Studies of Science* 15 (1985): 3-36.
[40] Pinch, "Towards an Analysis," 17.

neutrinos generated in the sun and neutrinos arriving at Earth was ultimately important, however. Solar neutrino oscillation, a change in neutrinos as they travel to Earth from the sun, is now taken to explain Davis' result.[41]

We could also parse this more finely. If Davis reported the number of argon-37 atoms in the tank and someone challenged that report, for example, he could explain how it can be inferred from the number of argon-37 atoms extracted from the tank or from the radioactivity of the extracted material. This would report the observation at levels of externality between the last two in the list above.

And we might add further levels of even lower externality. Faced with scepticism about the external world, Davis might report his sense data and argue on that basis that there were splodges. At that extreme, he would no longer be reporting anything of scientific interest.

Davis set out to test theories about what was going on in the sun, so the scientifically most interesting claim would be an observation of specific reactions in the sun. This would involve considerable risk, however, because there were all sorts of ways in which his report about solar neutrinos could turn out to be wrong. This is not just in-principle scepticism, either, since he observed far fewer solar neutrinos than physicists predicted based on the reactions that they expected were happening. At the lowest degrees of externality, however, the report risks being trivial. The large and complicated project would hardly be justified if, at the end of it, he could report nothing more than splodges. So characterizing the observation required balancing considerations of risk against considerations of significance.

These considerations reflect the cost of believing a risky claim (if it were false) and the cost of forgoing a significant claim (if it were true). These are the simultaneous demands that we should avoid error and seek truth. The JRD-thesis, that this tension is inescapable, applies as much to the observation claim as to the conclusions of inference. This means that values enter not just into the inference from evidence, but into stating the evidence itself. In considering Jeffrey and Betz, above, we saw that scientists can assign lower probabilities or hedge their reports in order to trade significance for security. Moving to lower levels of externality is a distinct strategy for doing so. In this example, Davis could have made his report safer by widening the error bars on his observation report of neutrino flux or by maintaining precision while characterizing the observation as being about objects

[41] Note that reports at different levels of externality may exhibit the same degree of generality. For example, the rate of neutrinos generated in the sun is just as specific and concrete as the rate of neutrinos arriving at Earth.

at lower levels of externality. Just as widening the error bars too far would make the report trivial, so too would retreating to the lowest levels of externality.

Scientists in the decades following Davis' observations worked to figure out what he had observed, both what could be concluded from it but also what the correct description of it was. Settling on the right level of externality took decades.[42] Davis could not wait for these developments, but had to decide what to believe and what to report at the time. He faced uncertainty, and navigating it required weighing the potential costs and benefits of different possible beliefs.

To generalize, the argument is this: In stating their observations, scientists face a choice between different levels of externality. Observation reports posed at a high level of externality are more significant, so a scientist who declines to accept a report in those terms risks missing out on the chance to believe an important truth. Conversely, such reports are also riskier, so a scientist who accepts such a report risks believing something false. There is no purely epistemic rule which assigns costs to these risks. Instead, they are a matter of value judgment. Just as ampliative risk means that values always play a role in theory choice, values play a role in every observation.

This same pattern is seen in other cases of experimental science, and the concept of externality is helpful for describing what is going on in general terms.

Kent Staley discusses collaborative research by groups like the one at Fermilab which discovered key evidence for the top quark in the 1990s.[43] They must settle on reporting their findings in some form. This yields two conflicting but indispensable pressures, Staley argues: "[T]hey seek to avoid the embarrassment of making claims that subsequent work reveals to be false; they also seek to achieve prominence and esteem by making novel and significant claims that are upheld by further critical scrutiny."[44] Claiming to have observed the top quark was a high externality report, significant but also risky.

Boaz Miller notes that experimentalists must distinguish signal from noise.[45] There are different methods for reducing raw data, and selecting which method to

---

[42] For a summary of subsequent developments, see John N. Bahcall, "Solving the Mystery of the Missing Neutrinos," April 28, 2004. http://www.nobelprize.org/nobel_prizes/themes/physics/bahcall/

[43] Kent W. Staley, "Evidential Collaborations: Epistemic and Pragmatic Considerations in 'Group Belief'," *Social Epistemology* 21, 3 (2007): 321-335 and "Decisions, Decisions: Inductive Risk and the Higgs Boson," in *Exploring Inductive Risk*, eds. Elliot and Richards, 37-55.

[44] Staley, "Evidential Collaborations," 323.

[45] Boaz Miller, "Catching the Wave: The Weight-Adjusting Account of Values and Evidence,"

use is subject to social influences and considerations of risk. Claiming only to have observed the raw data is a low externality report, safe but insignificant. Claiming to have observed a definite signal is higher externality, potentially significant but also potentially wrong.

Objections on the Basis of Epistemic Values

In this section, I rebut some possible objections which suggest that the values involved in picking a level of externality are only epistemic values.

Here is a first try at such an objection: Describing epistemic values, Ernan McMullin writes, "One value, namely truth itself, has always been recognized as permeating science."[46] Avoiding error, too, is clearly an epistemic matter. So, one may object that I haven't shown how non-epistemic or ethical values are in play.

However, the epistemic duties to pursue truth and avoid error ultimately pull in opposite directions. So seeking truth and shunning error are not enough, by themselves, to determine belief. When they are in conflict, the epistemic duties themselves cannot tell us how to strike a balance, and we must consider how important it would be to believe a claim (if it were true) and what the cost would be of remaining agnostic (avoiding error, if the candidate belief were false). This *importance* and *cost* will be practical rather than narrowly epistemic. Selecting a level of externality requires weighing enthusiasm against caution—that is, it requires reckoning with values.

Staley describes the risk of embarrassment pitted against the thirst for esteem—practical considerations for scientists, rather than merely epistemic concerns.[47] To take a schematic example, a junior scientist who needs to publish in order to have a chance at tenure might favor reporting a result now rather than waiting for further proof. They might be wrong, but waiting could be tantamount to dooming their career. Or consider a somewhat different junior scientist who has published enough to secure tenure. The risk of publishing too soon and detracting from their overall CV might lead them to greater caution than the first scientist. In both cases, they would be acting on their duties to make true claims and avoid making false ones, but in neither case is it just truth and falsity that guide their

Studies in History and Philosophy of Science Part A, 47 (2014): 69-80.

[46] Ernan McMullin, "Values in Science," *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association*, *Volume Two: Symposia and Invited Papers* (1982), 6.

[47] Staley, "Evidential Collaborations," 323, cited above.

decisions. They are weighing the expected utility of making claims or refraining from doing so. Therefore, the values involved go beyond the merely epistemic.

Note again that the argument does not show that every value or utility is relevant to scientific judgment.[48] Imagine that one of the junior scientists has already published a paper. It might be good for them if the claims that they made in the paper were true, but that by itself does not give them any reason to believe the claims. The relevant costs and benefits are the conditional ones: what they would have to gain by believing it *if it were true*, what would it cost them to believe *if it were false*, and so on.[49] McMullin is explicit that he would view scientists weighing the expected utilities of various judgments as an intrusion of ethical (non-epistemic) values.[50]

To revise the objection: One might concede that seeking truth and shunning error alone are not enough to settle theory choice but appeal to a longer list of epistemic values.

It is hard to reply to this without some candidate for what these extra epistemic values could be. Standard lists of theoretical virtues include things like fit with evidence, coherence, consistency, simplicity, scope, and fertility.[51] However, as theoretical virtues, these do not readily apply to observation. I do not see how these would provide *any* guidance in a case like Davis' neutrino observation, nonetheless enough guidance to settle the appropriate level of externality. One might attempt to enlarge the list of epistemic values even further to include other scientifically significant considerations, but I do not see how this could be made to work without collapsing the distinction between epistemic and non-epistemic values.[52]

---

[48] My conclusion here is more modest than Staley's claim that "the decision about communicating the outcome of an experiment is subject to the full range of utility considerations applicable to any practical decision" (Staley, "Decisions, Decisions," 53).

[49] This difference between categorical and conditional values is Douglas' distinction between values in a direct and values in an indirect role.

[50] McMullin, "Values in Science," 8.

[51] The list reflects ones given by Kuhn, McMullin, and (in a critical vein) Longino (Thomas S. Kuhn, "Objectivity, Value Judgment, and Theory Choice," in *The Essential Tension: Selected Studies in Scientific Tradition and Change* (Chicago: University of Chicago Press, 1977), 320-339; McMullin, "Values in Science;" Longino, *Science as Social Knowledge*). Longino and Douglas argue that there is no legitimate distinction between values like these and any others we might revere, but I accept the distinction for the sake of argument.

[52] See Philip Kitcher, *Science, Truth, and Democracy* (Oxford: Oxford University Press, 2001),

To revise the objection again: One might insist that community standards settle the relevant considerations. There are norms in the scientific community for when and what to report, and the two junior scientists considered above only have so much leeway in what claims to publish. In the schematic case, there are professional standards which limit how enthusiastic the first scientist is allowed to be.

Although community standards might make individual values irrelevant in particular cases, the general standards themselves must strike some balance between different possibilities of error. We saw above that a practical policy of flipping a coin would still reflect values at the level of policy. The same holds, for example, for setting a statistical threshold that results must meet in order to be publishable.

Moreover, community standards cannot anticipate every possibility of novel research. There were no community standards about how to report neutrino observations which Davis could rely on to specify an appropriate level of externality. For Davis—and for the junior scientists in the schematic case—community standards will constrain their choices without fully determining what they should believe.

Here is a final attempt to reformulate the objection: One might hope that a philosophical analysis or theory of perception will determine the proper level of externality. Scientific observations could be posed at that level with only reference to epistemic values, and all other scientific claims would be inferences from claims at that base level.

If the default level of externality is not to be subject to revision or scrutiny in the course of inference, then it must favor security over significance to an extreme degree. After all, even a report of splodges presumes that there is actual equipment yielding readings and not facades or hallucinations. Insisting that scientists initially represent observations only in the most secure terms would yield scepticism or phenomenalist empiricism.[53]

---

ch. 6, who considers numerous possibilities but concludes that there is no sensible way to construe scientific significance in purely objective or epistemic terms.

[53] Staley makes a similar point in relation to research teams. He writes that "if one were to ask that groups should ideally issue statements of group belief only when there is complete uniformity in what each individual member is 'compelled' to believe based on the evidence, one would in fact be saying that ideally such groups would issue almost no statements of any interest, and thus that there simply would be very little interesting empirical science" (Staley, "Evidential Collaborations," 328).

Perhaps the default level of externality need not be neutral sense data and may be theory-laden. All that the objection strictly requires is that scientists can rely on it regardless of utility considerations or value commitments. The objection still founders, because philosophical conceptions of evidence are little help in selecting a level of externality. To take one example, Peter Achinstein provides a theory of evidence which takes empirical data or phenomena as given.[54] It is about how bare observation becomes evidence, rather than about how observation becomes credible in the first place. I do not see how the problem is any less vexed on other theories of evidence.

To sum up: Selecting a level of externality requires balancing the desire for significant findings against aversion to mistakes. This balance of enthusiasm against caution depends on conditional utilities. That is, it is a matter of values. Attempts to see these values as somehow innocently epistemic fail.

A More Qualified Objection

In this section, I consider an argument by Bryce Huebner, Rebecca Kukla, and Eric Winsberg that values play *less* of a role in science like the search for the Higgs Boson than in policy-relevant research like climate modelling.[55] The result would, perhaps, be a limited version of the policy-relevance restriction.

Huebner et al. discuss the search for the Higgs boson and portray developments at CERN in a way that is initially congenial to my argument. They write that "inductive risk balancing continues to occur in unpredictable and, perhaps, unrecoverable ways throughout the research process, even where the research does not aim at some obviously value-laden goal."[56] Yet they go on to argue that the entanglement with values is importantly less complex for the discovery of the Higgs boson than it is for climate science. The existence or non-existence of the Higgs boson is a binary question. So, they argue, there are risks only along one dimension. Climate modelling does not involve a single binary question and so "researchers can have any of a wide, multi-dimensional array of

[54] Peter Achinstein, *The Book of Evidence* (Oxford: Oxford University Press, 2001).

[55] Bryce Huebner, Rebecca Kukla, and Eric Winsberg, "Making an Author in Radically Collaborative Research," in *Scientific Collaboration and Collective Knowledge*, eds. Thomas Boyer-Kassem, Conor Mayo-Wilson, and Michael Weisberg (Oxford: Oxford University Press, 2018): 95-116.

[56] Huebner et al., "Making an Author," 112.

investments in various outcomes…."[57] Arguing in this way, one might say that straight-forward existence questions are—although not value-free—less value-laden then more complex questions which matter to policy.

This difference is not as significant as Huebner et al. suggest. When an observation is construed at different levels of externality, the evidential context changes—that is, the observation is taken to tell us about different things. That means that what is presented as a binary question hides all sorts of other possibilities. Davis' question might have been posed as whether the dominant model of the sun was correct or not, but his observations were ultimately accepted as measurements of the rate of electron neutrinos arriving at Earth. So a claim can only be construed as a binary report against an implicit, multi-dimensional background of individual and community commitments.

Conclusion

The JRD thesis points to a tension at the heart of our epistemic lives. We aim to believe true things, and we aim to avoid believing false things. Our duties as knowers require us to arrive at some balance between these, and such a balance reflects value judgments: Conditional on the claim being true or false, what would the benefit or cost be of believing or not believing?

Neither the randomizer reply, the policy-relevance restriction, nor the lexical priority of evidence undo this. Values do not enter as merely a secondary consideration, nor do they enter only in situations with obvious policy consequences, nor do they enter only into inference.[58]

---

[57] Huebner et al., "Making an Author," 112.

[58] An earlier version of this paper was presented at the Society for Exact Philosophy meeting in May 2018. Thanks to Chris Meacham and other interlocutors at the conference for helpful feedback.