1-2022

The Scope of Inductive Risk

P.D. Magnus
University at Albany, State University of New York, pmagnus@albany.edu

Recommended Citation
Magnus, P.D. 'The scope of inductive risk.' Metaphilosophy. 53(1): 17–24. January 2022.

The scope of inductive risk*

P.D. Magnus

December 17, 2021

This is the author's final draft.
The paper is accepted and forthcoming in *Metaphilosophy*.

Abstract

The Argument from Inductive Risk (AIR) is taken to show that values are inevitably involved in making judgements or forming beliefs. After reviewing this conclusion, I pose cases which are prima facie counterexamples: the unreflective application of conventions, use of black-boxed instruments, reliance on opaque algorithms, and unskilled observation reports. These cases are counterexamples to the AIR posed in ethical terms as a matter of personal values. Nevertheless, it need not be understood in those terms. The values which load a theory choice may be those of institutions or past actors. This means that the challenge of responsibly handling inductive risk is not merely an ethical issue, but is also social, political, and historical.

Keywords: inductive risk, values in science, social epistemology, JRD thesis

1 Introduction

The Argument from Inductive Risk (AIR) is a standard move to show the connection between values and scientific belief formation. As I explain in section 2: Although the argument is sometimes formulated in narrow terms, it points to a tension between pursuing true beliefs and avoiding false ones that is present any time someone forms a belief. Numerous examples show how this plays out in scientific contexts— from theory choice, to model building, to reporting observation— and the generality of the argument seems to show that this entanglement of science and values holds for the formation of every belief.

As I argue in section 3, however, there are situations in which it is implausible to say that the formation of the belief reflects a value judgment— cases in which a belief is formed unreflectively, either by immediate perception or habitually

*Earlier versions of this paper were given as talks at the University of Texas, Dallas and at Rensselaer Polytechnic Institute. Thanks to helpful feedback from those audiences, especially Clair Morrissey, Per Wikman-Svahn, and John Milanese. Thanks also to Jon Mandle and Ariel Zylberman.

following a conventional procedure. Even though forming the belief will have consequences which might be good or bad, in such cases it seems wrong to say that forming the belief reflects the scientist's values.

This poses a puzzle: General considerations of risk suggest that every belief adoption reflects scientist's values, but there are some specific cases where that does not seem to hold.

In section 4, I offer a response to this puzzle which highlights the possibility that the trade-off made in belief formation might reflect not the values of the individual forming the belief but instead the values of their community or the values of past actors. This means that the challenge of responsibly handling inductive risk is not merely an ethical issue but is also social, political, and historical.

2 Inductive risk

The Argument from Inductive Risk (AIR) is perhaps the most used tool for establishing a legitimate place for values in scientific theorizing.

The issue is readily seen with statistical inference. Take a schematic example: You have some data, and you want to decide whether the data supports some hypothesis. Given rates of cancer in an experimental group exposed to such-and-so chemical and rates in a control group, does the chemical cause cancer? One rate is higher than the other. You calculate the level of statistical significance, for example as a p-value. Risk enters when considering what p-value is required in order for the result to be taken as supporting a link between the chemical and cancer. A looser threshold will license many conclusions that turn out to be false— so-called false positives. A stricter threshold will block many conclusions that turn out to be true— false negatives. Setting a p-value threshold weighs the cost of false positives against the cost of false negatives.

It is harder to formulate the general issue, because both words in the phrase 'inductive risk' are potentially misleading.

Philosophers use the word 'induction' equivocally: sometimes narrowly to mean projective inference from a sample to a population and sometimes broadly to mean non-deductive, ampliative inference. The same ambiguity applies to 'inductive risk', which might apply only to induction in a narrow sense. Rudner (Rudner 1953) explicitly holds that scientific inference is statistical. More recently, Andreasen and Doty write that "Inductive risk can be broadly characterized as the risk of erroneously accepting (or rejecting) a hypothesis due to the probabilisitic nature of most hypothesis testing" (Andreasen and Doty 2017, p. 128).[1] There is risk of error in all scientific inference, though, so the issue is not merely about interpreting statistical data.

We could highlight this broader issue by speaking instead of 'ampliative risk', but both induction and ampliative inference are *inference*. So inductive or ampliative risk so-called does not apply to observation, even though there

1. See also ChoGlueck (ChoGlueck 2018) and the exchange between MacGillivray (MacGillivray 2019) and Hicks et al. (Hicks, Magnus, and Wright 2019).

are also risks of error involved in making observation reports which parallel the risks identified by the AIR.[2]

The word 'risk' suggests that it is just the danger of things going wrong— drawing false conclusions— that is at issue. Ted Richards writes that "one is no longer employing the AIR if one factors the consequences of drawing correct conclusions into one's decision making" (Elliott and Richards 2017b, p. 265, fn. 1). Regardless of how we characterize the AIR, however, the potential danger of forming a false belief must be balanced against the potential benefit of forming a true one. The latter value should figure in settling belief just as much as the former.

Rather than whinge about what the AIR is exactly, let's consider a general feature of our epistemic situation. Magnus (Magnus 2013, 2014, 2018) calls this the *James-Rudner-Douglas* (JRD) thesis: Anytime a scientist announces a judgement of fact, they are making a trade-off between wanting to believe true things and wanting not to believe false things. These are both epistemic motives, so they are not some intrusion from the outside. However, epistemology itself cannot tell us how to weigh one against the other. Scientific judgement involves assessing costs and benefits.[3]

There are no universal values that can be filled in here, just as there is no absolute, rational rule for balancing enthusiasm against scepticism. Our enthusiasms and cautions are not topic-neutral. There are matters which we care about, where the prospect of possibly believing the truth matters to us. There are others about which we are more or less indifferent.

The JRD thesis is supported by considerations like the AIR, even if the AIR itself is narrower or somewhat different. Note that the thesis— although it applies to every judgement— does not allow space for every value. Instead, it just makes space for specific conditional values. With respect to a specific claim P, these values are

- the value of believing P if P were true,
- the cost of believing P if P were false,
- the cost of not believing P if P were true, and
- the value of not believing P if P were false.

This restriction to specific values reflects the distinction Heather Douglas (Douglas 2009) makes between two different *roles* that values can play. Values play a *direct* role if a scientist considers unconditionally whether it would be good or bad if P were the case; e.g. if she believes P because things would be better if P were true. The JRD thesis does not make space for that general consideration to play a role in theory choice. Values play an *indirect* role if they "act to weight the uncertainty about the claim, helping to decide what should

2. See Magnus (Magnus 2018, p. 418).

3. The thesis is named for William James, Richard Rudner, and Heather Douglas— three prominent exponents of it.

count as *sufficient* evidence for the claim" (Douglas 2009, p. 96). Values should not play a direct role in deciding what to believe, but they not only may but *must* play an indirect role.

This distinction is important, but describing it in terms of different roles is somewhat misleading. The values which play an indirect or direct role are not merely the same values doing different jobs. Rather, they are different values. The values which necessarily play a role in any belief are the costs/benefits of believing P (or not) *conditional* on P being true (or false). The values which are epistemically irrelevant are the costs/benefits of P being true (or false).[4]

Adopting any belief is subject to the JRD thesis, so values are always in play.

The conclusion of the AIR is often formulated in ethical terms. That is, a scientist is responsible for how they form their beliefs. When the costs or benefits are large, they ought to consider the value trade-offs explicitly. Failure to do so is both an epistemic and ethical failure. It is natural, then, to see the values in play as the values that the scientist themself either does or ought to have.

This personal reading of the values is reinforced by considering specific examples like the discovery that peptic ulcer disease (PUD) is caused by bacteria. In the early 1980s, it was believed that PUD was caused by excess stomach acid. Barry Marshall and his collaborators found a correlation between the presence of the bacteria *Helicobacter pylori* (*H. pylori*) and PUD. However, it was unclear whether the bacteria caused the ulcers or was merely present because the stomach was already compromised. Animal evidence did not show an effect. Marshall decided that further experiment was required, so he ingested a vial of *H. pylori*. He got sick, and his illness along with biopsies proved that *H. pylori* could cause illness in an otherwise healthy person. Further research reinforced the connection, and Marshall won the 2005 Nobel Prize in Physiology or Medicine (along with his collaborator Robin Warren).

Matters before Marshall's self-experiment were equivocal. The evidence could be interpreted to favor Marshall's belief, but it allowed an alternative interpretation. He saw connections with earlier work and other unexplained findings, but other scientists were unconvinced. Marshall drew different conclusions than his colleagues at least in part because of their different stakes. Marshall's research funding was nearing its end. Absent a breakthrough, he would have had to pursue a job in private practice. Alternately, Marshall later wrote, "a successful infection with Helicobacter would point towards a career in clinical research, more exciting but likely to be financially insecure" (Marshall 2006, p. 269). The JRD thesis, read as a matter of personal ethics, allows us to see both Marshall and his more conservative colleagues as responsible. Marshall weighed the possibility of getting it right more heavily than the risk of getting it wrong, and his personal stakes led him to see more promise in the claim linking

4. Note that there is no fixed way to calculate one set of values from the other. For example: It would be good for me to believe I have cancer if it is true that I do, but bad for me to have cancer— yet it it would be good for me to believe I have a winning lottery ticket if I do, and also good for me to have it.

H. pylori to PUD than his colleagues saw.[5]

To sum up: The JRD thesis is a general consideration and seems to hold for every claim, every belief.

3 Problem cases

In this section, I pose four kinds of cases in which someone forms a belief that does not reflect their values. Further examples of each kind are easy enough to devise.

1. Unreflective application of conventions: Consider a discipline which takes a result with a p-value of less than .05 to be statistically significant. Many scientists understand what this means and so, at least implicitly, are on-board for the trade-offs that this involves. Yet a student or someone peripherally involved in research who is trained in the practice may not understand it at that level. The statistical test is just a method that they apply in the way that they were taught to apply it. Because they do not know about the trade-offs involved in the threshold, it seems wrong to see those trade-offs as reflecting their values.

2. Black-boxed instruments: It is a familiar fact from the sociology of scientific knowledge that instruments become *black-boxed* over time. For example, there was a time when radiation was an unfamiliar phenomenon. Scientists built their own geiger counters and were concerned with how and where geiger counters could be used as reliable instruments. For a physicist today, however, geiger counters are a standard commodity which is ordered from a catalog. Although there is skill involved in using one, the scientist need not know what every part in the geiger counter does. They need not be in a position to reflect on the trade-offs involved in building it this way rather than some other way—if they are not, then they are in no position to measure the trade-offs by the standard of their own values.

3. Reliance on opaque algorithms: People increasingly rely on the output of software to inform their decisions. In some cases, the algorithm is proprietary and the end user has no access to the details of how it works. Even if such details are available, the end user may not have sufficient understanding of them to reckon with possible errors in the output.

The output of the algorithm can be biased without any bias appearing in the code of the program itself. A pattern-recognition algorithm can learn and so replicate biases present in its data set. For example, a network trained on pictures in which two-thirds of the people cooking are women learns to strongly associate cooking with women (Zhao et al. 2017). As Barocas and Selbst observe, algorithms "can reproduce existing patterns of discrimination, inherit the

5. My use of this example follows the discussion in Magnus (Magnus, forthcoming).

prejudice of prior decision makers, or simply reflect the widespread biases that persist in society. ... Because the discrimination at issue is unintentional, even honest attempts to certify the absence of prejudice on the part of those involved... may wrongly confer the imprimatur of impartiality on the resulting decisions" (Barocas and Selbst 2016, p. 674).

Moreover, an actual case can fall short of "honest attempts to certify the absence of prejudice." Imagine an engineer on a deadline just making something that will ship and a marketing department which nevertheless assures the end user that the algorithm is not biased. The inner workings of the system are a trade secret, so the end user will be in no position to tell what risks they face when forming beliefs based on the software. So the risks do not reflect the end user's values. Moreover, since the risks are the result of corners cut at the last minute, the risks need not reflect the values of the engineers or the software company either.

Some cases exhibit all three of the features that I have mentioned so far. Consider a technician in a hospital lab who prepares blood pathology results using an array of devices, reagents, and practices. They aim to follow the standard practices, and we might even suppose that they will produce more-or-less the same report as any other competent technician. The technician may have no understanding of the underlying systematic trade-offs of the systems which they are implementing, and it does not necessarily make the lab a better one if they do. So the judgments need not reflect the technician's values.

4. Unskilled observation reports: Consider an ordinary perceptual situation. As a first example, suppose a philosopher looks at a lamp to see if the light is turned on and sees that it is. He does not consider the possibility that the sun might just be glinting through the window and striking the lampshade in a way that makes it look as if the light is on. It does not even occur to him, so he does not reflect on the possible costs of error. He *could* consider the conditional values involved, if he stopped and asked about them, but he does not.

Justin Biddle and Rebecca Kukla argue that values are incorporated into "implicit phronetic practices", the tacit knowledge involved in observation. They write, "vision already encodes a balancing of values" and "values are built into... perceptual episodes" (Biddle and Kukla 2017, p. 221). This seems plausible in the case of the philosopher looking at the lamp. He could reckon with the values involved in the perceptual judgement but does not. Allowing a glance to suffice reflects at least an implicit judgement that the stakes are low.

However, there are other cases of forming perceptual judgements where the risks could not be made reflectively available. Suppose a different philosopher is working with biologists in the rain forest, surveying areas, and reporting the presence or absence of a particular kind of grasshopper. She lacks experience and so may miss grasshoppers or misidentify them, even if she is as attentive as she can be. She lacks the expertise to even estimate what her error rates or the significance of her errors are likely to be.[6]

6. This example is adapted from Clair Morrissey's description of her own experience.

The second philosopher could not reckon with the values involved because she lacks the relevant expertise to do so. Her degree of certainty should be proportional to her reliability— but, although she knows she is substantially less reliable than the biologists she is working with, she is unable to assess how unreliable she is. She lacks enough experience to coordinate her degree of confidence to her reliability, so it is implausible to think of the perceptual episode as encoding or incorporating her values.

To sum up: Several classes of claim or belief seem like counterexamples in which the JRD thesis does not hold.

4 A puzzle and some resolutions

The two preceding sections suggest on the one had that the JRD thesis is entirely general and holds in every case (section 2) but also that it does not hold in certain counter-example cases (section 3). This is a puzzle.

One possible response would be to hold the JRD thesis as a moral demand rather than as a description. One could then see each alleged counterexample as a case in which someone falls short. The idea is that agents always ought to explicitly consider the risks of forming a belief, but that they do not always do so.

When the philosopher in the former example looks at a lamp and believes that it is on without considering costs and benefits, this reply holds that he has done wrong. However, such a norm seems to require too much of him. If he were to consider costs and benefits explicitly, then he would have to form beliefs about conditional values. In order to do that, he would need to consider the costs and benefits of *those* beliefs. And so on— the obligation would require an infinite regress of judgements.

The philosopher could not possibly reckon with the costs and benefits associated with all of the beliefs in that infinite regress. More than that, he would probably not be more responsible by following that regress out as far as he could. Suppose that he sees that the light is on and that nothing much hangs on whether he is right or wrong about it. It would be more like a neurosis than like epistemic responsibility for him to riddle out the costs and benefits which might follow from forming the belief, about the costs and benefits which might follow from assessing those costs and benefits in that way, and so on.

So it is not true that agents always ought to explicitly consider values. This parallels a familiar point about ethical consequentialism. Even if the right action is the one that produces the best possible outcome, actually trying to figure out what action would produce the best possible outcome might not be the deliberative procedure which produces the best possible outcome. The tension is resolved by noting that consequentialism provides a standard of rightness and that this standard is not— or at least not necessarily— a good decision procedure for agents to employ.[7]

7. This is literally a textbook move; see e.g., (Shafer-Landau 2018, p. 63). Thanks to John Milanese for pointing out this connection.

Making a similar move, we can construe the JRD thesis as a standard of rightness. Every belief or suspension of judgment involves trade-offs between the risk of being wrong and the chance of being right. Believing responsibly requires making these trade-offs in a permissible way. Sometimes thinking explicitly about the trade-offs is the best way to believe responsibly, but the JRD thesis is not a decision procedure. It is compatible with there being any number of cases like those posed in the previous section.

Peter Railton (Railton 1984) characterizes this move for consequentialism as a shift from subjective views (providing a decision procedure) to objective ones (providing a standard of rightness). To use that language, we can see the cases in the previous section as counter-examples to a subjective reading of the JRD thesis. If we do not read the thesis that way, then the conflict is resolved.

Although this resolves the puzzle, it also undercuts the usual assumption that the values reflected in belief are the personal values of scientists.

For some of the cases, the values are instead those of the community. A student or technician who follows a statistical procedure might unwittingly expose themselves to certain risks and avail themselves of certain benefits, but there are statisticians who have a precise understanding of those risks and benefits. The trade-off does not reflect the student's or technician's values, but the trade-off was made self-consciously by other members of the community. Moreover, other members of the community who self-consciously reckon with trade-offs might not be doing so with their own personal values. If the standards are determined by a collective administrative process, then the values are the ones arising out of the institutional context.

Alternately, there are cases in which the values are those of relevant constituencies and stake-holders. Douglas (Douglas 2009) discusses the use of analytic-deliberative procedures to inform scientific decision making. The issue becomes political rather than narrowly ethical.

Where conventions are long-standing, the trade-offs were made by earlier institutions or experts. These may be forgotten over time. So the persistence of a practice carries with it past values. Per Wikman-Svahn (Wikman-Svahn 2022) gives this the helpful name *value inertia*.

The construction of instruments also involves considerable ingenuity and careful considerations. So well-established, black-boxed instruments may also exhibit value inertia.

Value inertia can expose us now to risks which, if made explicit, we would find unacceptable. The earlier trade-offs involved in establishing conventions or instruments— even if they were made responsibly, relative to the standards of the time— might be different than ones we would make now. Dominant values may have changed, or the context of use may be different. Institutions should have resources to revisit practices and instruments when values or context change.

Nevertheless, some value inertia is the inevitable result of having persistent practices. Just as it would be pathological for a single person to try to make explicit the possible costs and benefits attending to every belief, it would be dysfunctional for an institution to constantly revisit the trade-offs involved in

every practice. And it would be dysfunctional for a discipline to refuse black-boxing any instrument. The black-boxing of an instrument is valuable precisely because it allows subsequent scientists to use it without constantly revisiting the techniques and considerations that were used in constructing it.

In other situations it may be that practices are adopted, instruments standardized, or programs published without any consideration of epistemic risks. Institutional or market pressures might encourage doing something hastily, without due consideration, and such a precedent may be repeated and become standard. If that happens, beliefs formed may not reflect anyone's values at all—not the values of current scientists, nor the values of institutions, nor the values of past actors. Even then, the JRD thesis still applies because the believers (unwittingly) have a chance of forming a true belief and run a risk of forming a false one. Taking on these risks willy-nilly may indicate a personal, ethical failing on the part of the believers, but the believers will also have been failed by their community and their institutions. Believing responsibly is not just a private matter, but requires considering how we should conduct our epistemic lives and construct our epistemic institutions so as to manage the inevitable trade-offs.

To sum up: The various kinds of examples considered in section 3 show, contra some versions of the AIR, that there are some beliefs which do not reflect the values of the believer. The JRD thesis still holds, however, because none of these cases escape the inevitable trade-off between the promise of knowledge and the risk of error. This tradeoff can reflect personal values, but it need not. It might instead reflect institutional or historical values. Or the believer might stumble blithely into it, so that the tradeoff reflects nobody's values at all.

References

Andreasen, Robin, and Heather Doty. 2017. "Measuring Inequality: The roles of values and inductive risk." In Elliott and Richards 2017a, 127–147.

Barocas, Solon, and Andrew D. Selbst. 2016. "Big Data's Disparate Impact." *California Law Review* 104 (3): 671–732.

Biddle, Justin B., and Rebecca Kukla. 2017. "The Geography of Epistemic Risk." In Elliott and Richards 2017a, 215–237.

ChoGlueck, Christopher. 2018. "The Error Is in the Gap: Synthesizing Accounts for Societal Values in Science." *Philosophy of Science* 85, no. 3 (October): 704–725.

Douglas, Heather E. 2009. *Science, Policy, and the Value-free Ideal.* University of Pittsburgh Press.

Elliott, Kevin C., and Ted Richards. 2017a. *Exploring Inductive Risk.* Oxford University Press.

Elliott, Kevin C., and Ted Richards. 2017b. "Exploring Inductive Risk: Future Questions." In Elliott and Richards 2017a, 261–277.

Hicks, Daniel J., P.D. Magnus, and Jessey Wright. 2019. "Inductive Risk, Science, and Values: A Reply to MacGillivray." *Risk Analysis.* doi:10.1111/risa.13434.

MacGillivray, Brian H. 2019. "Null Hypothesis Testing ≠ Scientific Inference: A Critique of the Shaky Premise at the Heart of the Science and Values Debate, and a Defense of Value-Neutral Risk Assessment." *Risk Analysis* 39, no. 7 (July): 1520–1532.

Magnus, P.D. 2014. "Science and rationality for one and all." *Ergo* 1 (5): 129–138. doi:10.3998/ergo.12405314.0001.005.

———. 2018. "Science, Values, and the Priority of Evidence." *Logos&Episteme* 9, no. 4 (December): 413–431. doi:10.5840/logos-episteme20189433.

———. 2013. "What scientists know is not a function of what scientists know." *Philosophy of Science* 80, no. 5 (December): 840–849. http://www.jstor.org/stable/10.1086/673718.

———. Forthcoming. "William James on risk, efficacy, and evidentialism." *Episteme.*

Marshall, Barry J. 2006. "*Helicobacter* Connections." In *Les Prix Nobel,* edited by Karl Grandin, 250–277. The Nobel Prizes 2005. Stockholm: Nobel Foundation.

Railton, Peter. 1984. "Alienation, Consequentialism, and the Demands of Morality." *Philosophy & Public Affairs* 12 (2): 134–171.

Rudner, Richard. 1953. "The Scientist Qua Scientist Makes Value Judgments." *Philosophy of Science* 20, no. 1 (January): 1–6.

Shafer-Landau, Russ. 2018. *Living Ethics: An Introduction with Readings.* Oxford University Press.

Wikman-Svahn, Per. 2022. "Values in extreme scenarios for civil protection research." In *Research Ethics for Environmental Health,* edited by Friedo Zölzer and Gaston Meskens. Routledge.

Zhao, Jieyu, Tianlu Wang, Mark Yatskar, Vicente Ordonez, and Kai-Wei Chang. 2017. "Men Also Like Shopping: Reducing Gender Bias Amplification using Corpus-level Constraints." In *Conference on Empirical Methods in Natural Language Processing (EMNLP).*