EXPLORING INDUCTIVE RISK

KEVIN C. ELLIOTT AND TED RICHARDS

Reviewed by Daniel J. Hicks

Exploring Inductive Risk
Kevin C. Elliott and Ted Richards (*eds*)
Oxford: Oxford University Press, 2017, £64
ISBN 9780190467715

The role of ethical, political, social, and other non-epistemic values in science has recently emerged as a mainstream topic within the philosophy of science. Articles on science and values appear regularly in prominent philosophy of science journals, and the last few meetings of the Philosophy of Science Association (PSA) have included multiple sessions on values in science.

The argument from inductive risk (AIR) has played a major role in the mainstreaming of science and values. In its simplest version, AIR states that a decision to accept or reject a hypothesis should take into account the potential non-epistemic consequences of error (for example, avoidable deaths due to acting on a false hypothesis); and taking into account the non-epistemic consequences of error requires appeal to non-epistemic values; and thus the decision to accept or reject a hypothesis should appeal to non-

epistemic values. This conclusion is incompatible with the value-free ideal, which asserts that non-epistemic values have no legitimate role to play in accepting or rejecting a hypothesis.

AIR was discussed as early as 1896 by William James (Magnus [2013]), but it is most frequently associated with Richard Rudner ([1953]) in the Cold War era and Heather Douglas ([2000]) today. In *Exploring Inductive Risk* (EIR), editors Kevin Elliott and Ted Richards provide an introduction to the topic of inductive risk (IR) and aim to expand its scope. The collection includes eleven major chapters, with introductory and concluding chapters by Elliott and Richards, and a brief forward by Douglas. The contributors apply AIR, and IR considerations more generally, to topics such as biomedical research, climate policy, social science, and even the discovery of the Higgs boson. Several of the chapters would work well in undergraduate courses; others are more technical philosophical discussions that connect science and values to areas such as decision theory.

In their introduction and conclusion, Elliott and Richards frame the collection in terms of 'three major themes' (p. 11). First, some key concepts in the IR literature—including the concept of inductive risk itself—are understood in different ways by different contributors, and so there is work to be done clarifying these concepts. Second, several of the contributions respond to an objection or alternative to AIR that is frequently associated with Richard Jeffrey. (This association will be discussed near the end of this review.) The third theme concerns responses to the question, 'how should we handle the balancing of values when inductive risk is present' (p. 270)? That is, in a paradigmatic case of IR, scientists are faced with a tradeoff between values such as human health and economic activity. AIR, by itself, does not provide much guidance about the relative importance of these values.

For the purposes of this review, I found it interesting to use some Kuhnian concepts to organize the major chapters. We can think of IR as a paradigm, in the sense of a set of concepts and a template or model for applying these concepts to particular cases. The first set of chapters fit easily into this template. Most of these chapters are highly accessible and include self-contained introductions to IR, making them particularly useful for introductory courses. We might call them 'canonical IR'.

The second set of chapters expand the IR paradigm to novel areas without breaking or challenging it. The third set are more critical, and in Kuhnian terms might be read as identifying anomalies that do not fit within IR's conceptual framework. Papers from these two sets are better suited to graduate seminars.

The fourth set of chapters delve deep into technical philosophical analyses, and will generally be inaccessible except to other specialists. They work within the IR paradigm, but in a much more technical way than canonical IR. We might call them 'normal IR'.

The first set, the 'canonical IR' chapters, includes contributions from Andreasen and Doty, Resnik, Stanev, Stegenga, and, to some extent at least, Plutynski. Except for Andreasen and Doty's chapter, they all consider IR in biomedical research and policy-making. Resnik applies IR to dual-use research, that is, research that might be used either to prevent novel pandemics or to create powerful bioweapons. He focuses specifically on inductive risks in estimating the probability and magnitude of harmful effects from publishing or suppressing key details about this research. Stanev examines composite outcome events, such as the disjunction of 'death or non-fatal heart attack or re-hospitalization' (p. 174), as used in clinical trials. It is far from obvious that an intervention that significantly decreases the chance of a non-fatal

heart attack but slightly increases the chance of death should be counted as successful. Stegenga also considers clinical trials, used as evidence of safety and efficacy in policy decisions by the US Food and Drug Administration, and argues that 'the epistemic standard with which the FDA assesses experimental pharmaceuticals is low' (p. 17). Plutynski applies IR to cancer screening, specifically considering the variety of harms associated with over-diagnosis. Her chapter considers a much broader array of harms than the other chapters in this set, including reduced quality of life due to unnecessary treatment, psychological and financial costs to patients, economic inefficiency, and the downstream epistemic harm of distorting mortality rates due to the increased discovery of tumors that are unlikely to cause serious illness. To this extent, Plutynski's contribution goes beyond canonical IR.

Andreasen and Doty offer the only chapter on IR in social science. Their contribution is also notable for their close focus on researchers' actual statistical practice. Many canonical discussions of IR talk in generic terms about type I and type II statistical errors. By contrast, Andreasen and Doty discuss IR in choosing a particular statistical test, whether to use null hypothesis statistical testing (NHST) at all, and the operationalization of variables. In part because of this concreteness, their chapter would be especially useful in the context of statistics, social science, and data science courses.

The contributions by Havstad and Brown and by Powers extend the IR paradigm without challenging it. Havstad and Brown provide one of two chapters on science advising in climate policy (the other is Frank's, discussed below). They critique two models of science policy advising: Pielke's 'honest broker' and Edenhofer and Kowarsch's 'pragmatic-enlightened' model. Both propose that value-laden decisions can be 'deferred' from scientists to policy-makers. In response, Brown and Havstad argue that 'choice amplification' means that the space of 'potential [...] policy pathways' (p. 122) is impossible to navigate without making value-laden decisions. Thus, deferral is impossible.

Powers applies IR to the heteronormative framing of endocrine disruption as 'demasculinzation', arguing that IR applies to the choice of terminology, and not just accepting or rejecting hypotheses. Somewhat like Plutynski, this chapter is notable for the kinds of harms considered. The tradeoff, as Powers sees it, is between language that reinforces heternormativity and more evocative and politically effective rhetoric (pp. 251–2).

The third set, including the chapters by Bluhm and by Biddle and Kukla, engage with the IR paradigm more critically. Bluhm applies IR to the design of clinical trials, and identifies tradeoffs between establishing causal knowledge with certainty and producing the kind of understanding needed for clinical relevance. Based on this analysis, she criticizes certain background assumptions about the production of evidence. Biddle and Kukla make the most explicitly critical contribution in the volume. They argue that the term 'inductive risk' should be reserved for one specific case, and that most writers (including Douglas) are talking about a broader category of 'epistemic risk' or 'phronetic risk'. Beyond this semantic point, Biddle and Kukla make a number of substantive criticisms of the IR literature. Among other points, the literature represents inductive/phronetic risk management as 'operating at the level of the psychologies of individual researchers' (pp. 190–1), and thereby neglects the way the culture of research communities or the organization of institutions creates and responds to inductive/phronetic risk.

Last but not least, the fourth set comprises two more technical discussions by Staley and Frank. While IR is usually applied to fields with clear social significance—biomedical research, environmental sciences—Staley considers IR in particle physics, and specifically the decision whether to announce the discovery of the Higgs boson. Staley points out that the epistemic and non-epistemic consequences of an incorrect announcement are entangled (my pun), as an incorrect announcement could lead to reduced research funding. Staley also discusses the role of IR in the formal decision theory developed by Wald, Churchman, and Levi.

Frank's contribution focuses on Richard Jeffrey's response to AIR. Jeffrey proposed that scientists should not accept or reject hypotheses, but instead offer decision-makers Bayesian probability distributions. Non-Bayesian versions of this response—recently offered by Sandra Mitchell ([2004]) and Gregor Betz ([2013])—propose that scientists can offer policy-makers, as Betz puts it, 'hedged hypotheses that make the uncertainties explicit'. Frank identifies three necessary conditions for applying this response, and argues that two of them fail in the case of climate science and policy.

Viewing IR as a Kuhnian paradigm also suggests that we can use EIR as a window into the social organization of the science and values intellectual community. I was pleased to find that seven of the seventeen contributors (41%) are women, and six of the fourteen chapters (43%) have at least one woman as an author. This is significantly better than philosophy of science as a whole; 17% of PSA members were women in 2014 (Philosophy of Science Association [2018]).

However, I was more disappointed when I examined whose work was discussed in this volume. Specifically, using text mining methods, I identified philosophers who were named in the index, then counted the number of mentions each received in the text. Heather Douglas had the most mentions, which reflects her work in making IR a mainstream topic. However, of the indexed philosophers who aren't contributors to the volume, the top nine are men, and the top five are Rudner, Hempel, Jeffrey, and C. W. Churchman. Helen Longino is the only woman in the top ten. The contrast between Sandra Mitchell (nine mentions), Gregor Betz (forty), and Richard Jeffrey (fifty-eight) is particularly striking. Mitchell, Betz, and Jeffrey all gave related responses to the argument from inductive risk, but in this volume the response is strongly associated with the two men only.

There are other notable omissions of women philosophers. While Kristin Shrader-Frechette's work on risk and environmental policy anticipated some of Douglas's work (Shrader-Frechette [1991]), she is cited once and does not appear in the index. Deborah Mayo's ([1988]) 'error statistical philosophy' is highly relevant to IR; she is cited once, with her frequent collaborator, econometrician Aris Spanos; and Spanos, but not Mayo, appears in the index.

In short, EIR suggests that we—the science and values community, in which I include myself—are more demographically diverse than other subfields of philosophy. But it also suggests that underneath this diversity lies a gendered distribution of epistemic authority (Krishnamurthy and Wilson [2015]). Both women and men contribute to the IR literature; but, other than Douglas, the contributions of only a few men are taken to warrant extended discussion.

This critical point is not directed at any one chapter, nor at Elliott and Richards as the editors. Each individual chapter makes a valuable contribution to the IR literature, and the collection as a whole has

significant value as both an introduction to IR and for expanding its scope. But at the level of the science and values community, we should be more deliberate about whose work we cite and whose ideas we develop.

Daniel J. Hicks
Data Science Initiative
University of California, Davis
hicks.daniel.j@gmail.com

References

Betz, G. [2013]: 'In Defence of the Value Free Ideal', *European Journal for Philosophy of Science,* **3**, pp. 207–20.

Krishnamurthy, M. and Wilson, J. [unpublished]: 'What's Wrong with Current Citation Practices in Philosophy?'.

Magnus, P. D. [2013]: 'What Scientists Know Is Not a Function of What Scientists Know', *Philosophy of Science*, **80**, pp. 840–49.

Mayo, D. G. [1988]: 'Toward a More Objective Understanding of the Evidence of Carcinogenic Risk', *Proceedings of the Biennial Meeting of the Philosophy of Science Association*, **1988**, pp. 489–503.

Philosophy of Science Association [2018]: 'Member Demographics'.

Mitchell, S. D. [2004]: 'The Prescribed and Proscribed Values in Science Policy', in P. Machamer and G. Wolters (*eds*), *Science, Values, and Objectivity*, Pittsburgh, PA: University of Pittsburgh Press, pp. 245–55.

Rudner, R. [1953]: 'The Scientist qua Scientist Makes Value Judgments', *Philosophy of Science*, **20**, pp. 1–6.

Shrader-Frechette, K. S. [1991]: *Risk and Rationality: Philosophical Foundations for Populist Reforms*, Berkeley, CA: University of California Press.