Managing Values in Science: A Return to Decision Theory

Eric Winsberg

Kennedy Institute of Ethics Journal, Volume 34, Number 4, December 2024, pp. 389-418 (Article)

Published by Johns Hopkins University Press
DOI: *https://doi.org/10.1353/ken.2024.a965814*

Eric Winsberg

Managing Values in Science: A Return to Decision Theory[1]

ABSTRACT. There are many proposals in the literature on how to "manage values." Many of these proposals have in common the assumption that the relevant values in science can be "packaged for transfer": they can be put in an envelope for scientists to hand to stakeholders or policymakers, or for members of the public or ethical experts to hand to scientists. The central aim of this paper is to argue that packaging values for transfer is a practical impossibility. The central argument of the paper concerns the best stance to take on how *values in science* should be conceptualized. Specifically, I argue that we need to return to a decision-theoretic definition of values (as, I argue, is strongly suggested by a close reading of Rudner, Jeffrey, and Hempel.) Further, I argue for a picture of these values that is nonpsychologistic, stance relative, and always and everywhere entangled with credences. I call my account of the nature of values in science the Putnam-Hempel account. Unfortunately, the Putnam-Hempel account forces us to see that any proposal that depends on packaging values for transfer will ultimately suffer from great difficulties.

> I have discovered that when I say I "reject the fact/values dichotomy," I am often misheard as saying "there is no difference between facts and values." But that is not what I am saying. – Hilary Putnam (2012, 114)

1. INTRODUCTION

Science involves values. Values are invariably implicated whenever scientists endorse claims or build representational tools, like models. The values of individual scientists, however, might not be the appropriate ones to guide these decisions: we might worry that scientists will incorporate, in one sense or another, the *wrong* values. This might be because they don't share the same values as the public they serve, because the values themselves are objectionable on ethical grounds, or for any number of other reasons. This prospect is especially worrying in

Kennedy Institute of Ethics Journal Vol. 34, No. 4, 389–418 © 2025 by Johns Hopkins University Press

policy-relevant science. It is not hard to imagine that the scientists who built the models designed to project the likely impact of COVID-19 mitigation measures did not use the "right" values when broaching questions like "How much more important is saving lives than providing children with education and future economic opportunities?" and "How much personal liberty is worth sacrificing to delay the spread of COVID by a certain amount of time?" One might imagine that this is because their values were idiosyncratic or because they were simply wrong. This raises the question of how to cope with this problem, which is often discussed under the heading of *how to manage values in science*. In the literature, the idea of managing values is usually understood as finding a procedure for avoiding one or more of the problems that the so-called value-free ideal was supposed (by those who thought it was tenable) to prevent: wishful thinking, bias, lack of objectivity, political illegitimacy, and so on (Anderson 2004; Brown 2020; Elliott 2022; Lusk 2020). One well-put goal of "managing values" is to "enable[e] community members to make [scientifically informed] decisions that accord with their values" (Elliott 2021, 2). We want to enable this even when community members' values differ from the experts whose scientific advice they need.

There are many proposals in the literature on how to "manage values." Almost all of these proposals share the assumption that the relevant values in science can be what I will call "packaged for transfer." These proposals, in other words, all take the values in science to be something bite-sized, separable, and communicable. They assume that if I am a scientist, I can hand stakeholders or policymakers an envelope that lists all the values that went into my decision-making. Or if I am a member of the public, or a policymaker, or an ethical expert, I can hand scientists an envelope with a list of values that they can use in all their decisions regarding how to build their representations or regarding what purported facts to endorse. Sometimes, for short, I will call this "encapsulating" values—I use both expressions, "package for transfer" and "encapsulate" to refer to what I describe in this paragraph.

The central aim of this paper is to argue that packaging values for transfer is a practical impossibility. Thus, any proposal for managing values in science that involves values being passed back and forth between scientists and other actors will ultimately suffer from great difficulties and will not appear attractive to readers who appreciate the main points of this paper. Elsewhere, I explore the possibility, with Stephanie Harvard, that scientists working in especially policy-relevant areas, particularly

scientists building particularly policy-relevant models, must *engage their stakeholders directly in their methodological decision-making* if they hope to address the problems that the phrase "managing values in science" is directed at (Harvard and Winsberg 2023; 2024; Harvard 2024).

2. PACKAGING FOR TRANSFER: WHO NEEDS IT?

Let's begin by appreciating just how much of the literature on managing values in science depends on a conception of values as something that can be packaged for transfer. The first proposal that comes to mind is the "transparency plus backtracking" proposal, whereby scientists are encouraged to provide their stakeholders with enough information about the values behind their methodological decisions for stakeholders to consider their impact on the results and whether they would reach the same conclusions using their own values (Elliott and McKaughan 2014) (see section 6). Another proposal that requires that scientists package their values for transfer to policymakers is Stephen John's "Weber's elephant" account.[2] On this account, policymakers should allow a variety of different scientific groups to produce policy-relevant science, and what they get from each scientific group is a view of the world from the perspective of a different value framework. Several other proposals require that values be packaged for transfer in the opposite direction: from the public, or users, or ethical experts, to scientists. A recent, prominent example of such a proposal is Wendy Parker's (2024) "epistemic projection" approach, according to which users should provide scientists with a "brief" that includes things like their attitudes to inductive risk (see section 4). Other proposals of this kind include S. Andrew Schroeder's (2017) democratic solicitation proposal, where values come from the public, and Parker and Greg Lusk's (2019) proposal that "user values" be used to manage inductive risk in climate science. Finally, this type of proposal is reflected, to a certain degree, in the various proposals that suggest that the values in science should come from ethical experts (e.g., Anderson 2004; Kourany 2010).[3]

If we want to evaluate these proposals, particularly with regard to the feasibility of passing the values back and forth between scientist and the public (or their users, or ethical experts) we have to think carefully about exactly what the values in science are and exactly what role they play in the science. To begin this process, imagine that you want to cross the street, but you see a large truck coming. You ask me if I think you can make it across the street before the truck hits you. For whatever reason,

you trust my judgment on that question better than your own. Maybe I have a PhD and a university appointment in truck kinematics. But to give you a yes or no answer, I need to both exercise my epistemic judgment about the truck's arrival time *and* weigh the value of your getting hit by the truck against the value of your making it across the street. The problem of managing values is the problem of how I can deliver a package of information to you that allows you to benefit from my expert advice without my imposing the "wrong" values on you. This problem is made especially difficult if one supposes that it is very difficult for me, in my own mind, to separate my epistemic judgment from my values. As the Putnam (2012) quotation that begins this paper suggests, it is perfectly possible to think that epistemic judgments and values are conceptually distinct, but it is nearly impossible to separate them in practice. Indeed, this is more or less a crucial axiom of the science and values literature. As agents, we often simply decide how to act, and it often takes a very complex process of elicitation to suss out, even for ourselves, what combination of values and credences guided our actions. This includes the action of advising someone that it is safe to cross the street.

Of course, the difference between the real problem of managing values in science and the truck case is that in real life there is not just you and me: there is a whole community of scientists (sometimes following epistemic norms), there is a whole body politic of users of scientific information, and there are sometimes decision-makers between them who have a responsibility to act democratically. But the vignette gets at one significant, central problem associated with managing values in science: the difficulty associated with separating, in practice, credences from values in just the way that Putnam emphasized. As Richard Jeffrey emphasizes in his seminal work, in the absence of an objectively correct inductive logic, settling on a credence in a value-free way presents "great practical difficulties difficulties" (1956, 246).

One reason that the literature on how to manage values in science has failed to appreciate the difficulty of packaging values for transfer is that contributors to the literature offer a wide variety of accounts of what the values in science *really are*.[4] For example, if the values in science are conceptualized as "desirable things," then it becomes plausible that they can be easily passed from user to scientist. A central part of my argument, therefore, will involve convincing the reader that the values in science that need to be managed are far more complex than this. Indeed, it will involve convincing the reader that the values in science that need to be

managed are the "preferences over prospects" of decision theory.[5] Once this central point is appreciated, it becomes easier to see that the values that matter to science are too complex to be solicited from the public, other relevant stakeholders, or ethical experts;[6] or too complex for scientists to transparently disclose to their stakeholders or to allow stakeholders to identify the group of scientists who best share their values. In short, *the values in science cannot be packaged for transfer*.

To make all the points I want to make in this paper as crisply as possible, I will need to redefine some terms that frequently appear in the literature, though in most cases, readers will find that my use of terms aligns quite well with those of Richard Rudner (1953), Jeffrey (1956), Carl Hempel (1965a), and even Frank Ramsey (1931).

3. WHAT IS A VALUE THAT A SCIENTIST MAY NOT BE FREE OF IT? AND WHAT IS A SCIENTIST THAT SHE MAY NOT BE FREE OF A VALUE?

Many of the possible confusions that I want to alleviate in this paper can be addressed by emphasizing the tight connection between values in science and decision theory. Decision theory is a framework for understanding how individuals with their own personal preferences make choices under uncertainty. Decision theory involves three fundamental components, and each of these correspond to crucial aspects of the role of values in science. The first of these is outcomes or prospects: the possible consequences of a decision (e.g., if you decide to bungee jump into a ravine, the outcomes could be adrenaline-fueled fun, nausea, injury, or death). The second is probabilities: each outcome or prospect is associated with a probability (or credence) representing the degree of belief the agent has in that outcome occurring. The third is utilities, otherwise known as preferences or *values*: the utility of an outcome is a numerical value representing the decision-maker's preference for that outcome. The central claim that I will make here is that, in much of the "values and science literature," when it is claimed that science ineliminably involves values, the "values" at stake are literally the utilities, preferences, or values of decision theory. The answers to the two questions posed in the section title, therefore, are these:

1. A scientist is a decision-maker, and decisions can only be made on the basis of desired outcomes. Beliefs and credences alone do not provide the basis for decision-making.[7]
2. Values (of the kind of which scientists cannot be free), therefore, *are desired outcomes*. More precisely, they are the numerically weighted *preferences* decision-makers (including scientists) have *over prospects*.

> I'll say more about this below, but to say that values are numerically weighted preferences is not to say that that these numbers are measurable, or that agents are necessarily aware of them. They are what we, as theorizers, are required to posit if we want to understand decision-making as rational.

That the values in science are the same as the values in decision theory is, in fact, unsurprising: Rudner's original argument was a decision-theoretic one, and decision theory is the foundation of the Rudner/Jeffrey exchange. Rudner (1953) starts by saying that the threshold for endorsing a hypothesis depends on the seriousness of the potential mistake.[8] What he means is straightforward: when deciding if evidence for a hypothesis is sufficiently strong to endorse it, we are implicitly calculating the expected utility of endorsing it (or applying whatever alternative decision framework one believes in).[9] Thus, when we decide whether to endorse a hypothesis, we must be understood as invoking both our credences and our values. If both our credences and our values are invoked simultaneously, then it's clear that not only must our credences be numerically weighted, *but our values must be weighted too*. Otherwise, it is unclear what the link between credences and values could have been in the argument from inductive risk. How can I know that I am allowed three times as much uncertainty that the belt buckles are sound than I am that the drug is free of toxins (Rudner's famous example) unless I have the conviction that it is three times as bad to be poisoned as it is to have your pants fall down, or mutatis mutandis for any other decision about whether or not to endorse a hypothesis when evidence is of a certain strength?

Hempel, who coined the phrase "inductive risk," was also clear that the inductive-risk argument was a decision-theoretic argument. In discussing inductive risk, he says explicitly that the values in question are the utilities of decision theory:

> This basic point is reflected also in the contemporary mathematical theories of decision-making. One of the objectives of these theories is the formulation of decision rules which will determine an optimal choice in situations where several courses of action are available. For the formulation of decision rules, these theories require that at least two conditions be met: (1) Factual information must be provided specifying the available courses of action and indicating for each of these its different possible outcomes—plus, if feasible, the probabilities of their occurrence; (2) there must be a specification of the values—often prosaically referred to as utilities—that are attached to the different possible outcomes. Only when these factual and valuational

> specifications have been provided does it make sense to ask which of the available choices is the best, considering the values attaching to their possible results. (Hempel 1965b, 89)

The standard version of the inductive-risk argument points out that scientists are decision-makers when they endorse (or don't endorse) a hypothesis. Deciding to endorse a hypothesis, however, is not the only decision a scientist *qua* scientist makes. As Hempel points out, endorsing an observation statement as an available piece of evidence is a decision that requires values. And Jeffrey (1956) points out that, in an absence of an objectively correct inductive logic, it requires a decision to recommend a credence. Furthermore, as Harvard and I (Harvard and Winsberg 2022) emphasize, making representational decisions, like deciding on an experimental procedure or making a choice of idealization in a model, also requires values. Saying that a decision involves values is just a way of emphasizing that it is in fact a *decision*, since decisions can only be understood as rational given a set of preferences over prospects.

In Rudner, there are exactly four outcomes whenever we decide whether or not to endorse a hypothesis: endorsing and being wrong, rejecting and being wrong, and so on. But Jeffrey (1956) argues convincingly that this is wrong—or at least too coarse grained: there is not one outcome associated with endorsing the hypothesis that a belt buckle is nondefective when it is in fact defective. There is the outcome associated with endorsing the hypothesis and wearing the belt to class (your pants fall down in front of everyone), and there is the outcome associated with wearing the belt while mountain climbing (the possibility of falling to your death). So, the expected utility of endorsing the claim that the belt buckle is sound (the hazard) involves the probability of being wrong (the hazardous event), combined with all the further probabilities of downstream consequences (the harms) like having your pants fall down and falling to your death. The second claim I will make, therefore, following on the claim that the values in science are the utilities of decision theory (which are numerically weighted preferences over outcomes), is that the relevant outcomes are not as Rudner conceives them but as Jeffrey does.[10] They are more fine-grained than endorsing a false hypothesis, failing to endorse a true hypothesis, and so forth. They are all the subsequent possible benefits and harms like having your pants fall down, successfully summitting the mountain, and so on.

It is important to know that the utilities of decision theory are subjective—that is, they are personal, or agent-relative. To say an agent's utilities are subjective, however, is just to emphasize that they are the

preferences that agent happens to have, as individuals. It is not a normative claim about what preferences they *should* have. It is a further, open question whether any agent's subjective (or personal, or agent-relative) utilities are objectively permissible, mandatory, or impermissible. Decision theories are often normative in that they dictate which actions are rational given beliefs and values, but they are not (ethically) normative in the sense of having anything to say about what values it is permissible or obligatory to have. Note that decision theory is not a theory of right action. It is only a theory of rational action, conditional on certain preferences. Because it defines values personalistically, decision theory allows us to capture the possibility of agents preferring something that is objectively bad. Hitler preferred genocide to no genocide. This reflects that there is more to morality and politics than finding the course of action that gives the people what they want. We don't necessarily want to insist that individuals' subjective assessments are the only thing that matters for normativity, or that we can just read correct normative judgments directly off of what individuals desire. This is why, as Schroeder (2021, 2022) correctly emphasizes, the questions of what are the ethically best values and what are the democratic values can come apart.[11] Nothing in decision theory dictates which one is better—these are questions for political theory, or ethical theory, to sort out, and decision theory does not even tell us which of the two to consult.

4. WHY SHOULD PHILOSOPHERS OF SCIENCE USE A DECISION-THEORETIC DEFINITION OF VALUES? AND WHICH ONE SHOULD THEY USE?

There are a number of advantages to thinking about the values in science as the preferences over outcomes of decision theory. In fact, there is added benefit if we construe those preferences or utilities in what we might call a "logicist" conception of utilities. This is one of two different ways of thinking about values in economics and decision theory. In one conception, values or preferences are posited as a real feature of people's internal psychologies. They are actual mental states that decision-makers consult when they make decisions. On another, more logicist reading of preferences (that we might associate, canonically, with Ramsey [2016]), preferences are simply things that decision theorists need to posit in order to understand decision-makers as rational agents—and on this conception it is an irrelevant empirical question whether the preferences are out there in people's heads. Indeed, it might not even be the case that the values in question are real psychological entities that are epistemically available

to the agents making decisions. On the logicist view that I endorse, the relative value you assign to As and Bs simply is the set of dispositions you have to choose, when forced, between one A and two Bs, one B and two As, three Bs and one A.[12] As we will see, there are tremendous advantages to de-psychologizing values in this way.

The central advantage of these two features of my conception of values—that values are the utilities of decision theory and that they are de-psychologized, being simply what is required for us to posit in order to understand decisions as rational—is that it helps us to see both how complex and how ubiquitous values are in science. This, in turn, will help us to see that it would be impossible, in practice, to package them for transfer. If we look at Rebecca Korf and Kevin Elliott's (2024) catalog of the different conceptions of values in the literature ("criteria for choice," "causal factors," "desirable things," and "beliefs or attitudes about desirable things") we find that only the last of these enables us to capture this crucial idea that a scientific decision is guided not merely by scientists' valuing, say, absence of disease, but by *the degree to which* they value absence of disease relative to, say, learning loss. Value-laden disagreements during the COVID-19 crisis about what the best models to use to project the impact of nonpharmaceutical interventions were not disagreements about whether avoiding disease was valuable. They were disagreements about how valuable this was compared to avoiding other undesirable outcomes. Ultimately, any noncomparative definition of "values" fails to capture what information is necessary to package for transfer in any proposal to manage values in science.

Consider Schroeder's (2021) proposal that we solicit values from the public and encourage scientists to incorporate them into their decision-making, taking into account that such a public might be "environmentally minded." While the public may well start by telling scientists that they value the environment, they would need to tell scientists far more in order to guide their methodological decision-making. The public would need to make clear not only that they have a preference for outcomes that protect the environment but that they are willing to sacrifice other aspects of an outcome in order to reach one that is protective of the environment. And insofar as we want to achieve value management, the public would need to say how much of other things that they value they are willing to sacrifice in order to achieve environmental protection. Or consider Parker's (2024) recent "epistemic projection" proposal for how scientists can manage values. At the heart of the proposal is the idea that scientists can solicit

the "epistemic desiderata" of their users. For example, users could inform scientists of their inductive-risk desiderata by informing them of whether they prefer to err on the side of type I error or type II error. But we can see from the truck vignette why it would be difficult to make this work. Suppose I inform you that I prefer that you err on the side of protecting my life rather than on the side of ensuring I get to the other side. Unless I tell you *by how much*, you will not be able to use this information. Even in the simplest case, where all you have to do is choose between an Acme brand and a Vandelay Industries brand of truck-speed detector, and you know the Acme brand gives a higher value, you cannot determine for sure that I want you to use the Acme brand, unless you think I want you to err on the side of my safety *come what may*. But that's the rarest sort of case.[13] And if you must make multiple such methodological decisions, some of which are more than binary, and perhaps even make some decisions that lie on a continuum, "I want you to err on the side of my safety" is nowhere near detailed-enough information. Similar remarks could be made about how to manage representational risk. Users might say, "We want you to include race in your COVID vaccination model (cf. Harvard et al. 2021) *come what may,*" but their values might also dictate that they want race included in the model so long as the benefits of including it (say, avoiding racial disparities) don't come at too high a cost (say, to overall population health.) But then knowing what "too high" means will be crucial. Scientists will need to know how much to value avoiding racial disparities *relative to* avoiding total disease burden overall.

Just as a conception of value that is not fundamentally comparative will not serve us well in determining how to manage values in science, neither will a conception that is overly psychologistic rather than logicist.[14] There is an obvious reason for this: if, when it comes to the question of how to manage values in science, we ignore the role that values play in scientists' thinking *unless those values manifest at some specific psychological level*—say, of consciously motivating a scientist, or being what a scientist explicitly uses to externally justify their decisions (Ward 2021)—then we will ignore many values that influence scientists' decision-making. In other words, in conceptualizing "values" for the purpose of contemplating how to manage the role of values in science, we should seek a conception that captures as completely as possible the values that are implicated in every single decision that scientists make.

We can avoid psychologizing values by emphasizing that values are things that arise when we take a stance on some actor as being a rational

decision-maker.[15] Once we take such a stance, it is natural to think of values as one component of a disposition to act in such and such a way in such and such situation. I say they are components of dispositions because the actual dispositions are to decide to act in such and such a way, and we take the stance that this disposition can be decomposed into credences and values by treating the decider as a rational agent. So values are stance relative, not psychologistic, and a component of a stance we take on why deciders rationally do what they do. *This* is the conception of values that is most fit for our purpose.

By clearly defining values as components of a stance we take on *dispositions*, rather than as beliefs or other psychological states, furthermore, we will be better able to detect ambiguities in otherwise useful contributions to the literature and to avoid getting caught up in questions that are irrelevant to the goal of knowing how to manage values in science. Consider, for example, Zina Ward (2021), who asks if there are a variety of ways in which values can been seen to play a role in science, and who distinguishes between times when values "motivate" or "justify" scientific decisions. It often seems that Ward, when she talks about the "justifying" role that values play, is on to exactly the de-psychologized, logicist conception of values that we ought to be focused on. For example, Ward quotes me to illustrate the "justifying" role:

> [S]uch choices [of what facts to endorse or what representational decisions to make] can only be defended against some set of predictive preferences and some balance of inductive risks. In other words, any rational reconstruction of the history of climate science would have to make mention of predictive preferences and inductive risks at pain of making most of these choices seem arbitrary. … I do not mean to attribute to the relevant actors these psychological motives, nor any particular specifiable or recoverable set of interests. I am not in the business of making historical, sociological, or psychological claims. (Ward 2021, 55–56; quoting Winsberg 2012, 131)

In summary, *beliefs and credences alone do not fix actions. Only beliefs plus values give you action*.[16] Thus, I am emphasizing a de-psychologized conception of values that is only "justificatory" in the thinnest sense (the sense in which decision theory is normative—it distinguishes rational from irrational choices). At other times, Ward explains what she means by the "justifying reasons" of values in a rather different way. In a sketch of Ward's (2021, 55) that is meant to illustrate the "justifying role," a politician says to his constituents that the reason he voted for a health measure is that he cares about public health, even though he was really

motivated to vote that way because it benefits him financially. This conception of the "justifying role" of values is not what we are after at all. (Furthermore, much of what Ward discusses could be described as being concerned with the psychological role that values play—the very thing we hope to avoid with a logicist picture.) So, to be clear, the conception of the role of values that I am endorsing here is quite similar to Ward's "justifying role," but only on one way of disambiguating what she means by justificatory. It is justificatory in the sense that if a decision theorist wants to portray a decision-maker as rational in all her decisions, the theorist will have to appeal to this logicist role that values play. But it is not "justificatory" at all in the sense of having anything to do with it being a thing the decider tells interested parties in order to justify the decider's choice.

5. VALUES AND REVEALED PREFERENCES

Not only can we benefit from seeing a close connection between decision theory and the role of values in science, but we can benefit by comparing the role that values play in science to the role they play in revealed-preference theory (RPT) in economics. The idea of RPT is that people's choices are essentially constitutive of their utilities, in the same way that subjective Bayesians often take people's betting behaviors to be constitutive of their credences. I want to make a parallel claim that when scientists decide to endorse facts or make representational decisions, these are constitutive of a combination of their credences and their utilities. On my view, like in RPT, every time a scientist makes a choice to endorse a fact or makes a representational decision, she is "revealing" her preferences in the same way that a consumer does when they choose product A over product B in the market. But unlike RPT, I put the word "reveal" in scare quotes because, in the real world, decisions fail to completely determine preferences. Two reasons are especially important to us: if I watch you make a decision, I can only infer your preferences if I know both your *credences* and *the decision theory you employ*. (Are you a strict utility maximizer? Or do you have risk aversion? Or maybe you are a minimaxer?) Ramsey made the reciprocal point at the dawn of decision theory:

> The old-established way of measuring a person's belief is to propose a bet, and see what are the lowest odds which he will accept. This method I regard as fundamentally sound; but it suffers from being insufficiently general, and from being necessarily inexact. It is inexact partly because of the diminishing marginal utility of money, partly because the person may have

> a special eagerness or reluctance to bet, because he either enjoys or dislikes excitement or for any other reason, e.g. to make a book. The difficulty is like that of separating two different co-operating forces. (Ramsey 2016, 30)

Ramsey is making the point that we cannot infer someone's credences unless we know their exact utilities (even for money) and their appetite for risk. The reciprocal point is that we cannot infer someone's utilities unless we know their exact credences and their appetite for risk. In the real world we hardly ever know all of these. Thus, unlike RPT, I do not to endorse any particular operationalization of values, or any particular experimental method of measuring them. Of course this is well understood even in RPT, where economists will admit that many revealed-preference experiments are confounded. You might look as if you prefer Rice Crispies to Corn Flakes, but my experiment to determine your preference might be confounded by the fact (unknown to me) that you think you saw a spider in the Corn Flakes. More generally, all your actions are guided by both your preferences and your beliefs, and I can only observe your actions, so I can only indirectly infer your preferences. Your beliefs are always a potential confounding factor.

Indeed, scientific decisions are *confounded* revealed-preference experiments *par excellence*. When a scientist endorses a fact, we cannot tell, from the surface, to what degree this choice is influenced by their epistemic appraisal of the evidence and to what extent it is influenced by their values. Indeed, there is no philosophical reason we should suppose that scientists are even aware of what role their values and their credences are playing in their decisions. (Again, this is what is emphasized in the Putnam quotation.) This is what I refer to in this paper as the *entanglement thesis*: one cannot untangle credences and values when analyzing a decision. This cannot be done omnisciently from the inside (i.e., by the decision-maker in question) or the outside (i.e., by the observer of the decision-maker in question).

We are now in a good position to construct a theory of "values in science." The theory assumes that scientists are rational decision-makers, which means their decisions are the joint outcome of values and credences (and a decision theory). This, I argue, is the only theory that is adequate for the purpose of analyzing the significance of scientific decision-making and advancing normative proposals that concern it—that is, roughly what I take to be the project of philosophy's values-in-science literature. Thus, I propose the following picture of a value that we ought to have in mind when we talk about the values in science:

> *Values in science*: nonpsychologistic, stance-relative, intrinsically confounded-with-credences-and-decision-theory things that combine with credences in a decision-theoretic framework to rationally explain scientific decisions.

For lack of a better term, I'm going to call this the *Jeffrey-Hempel-Putnam theory of values* in science (JHP).[17]

To be sure, many philosophers of science have given very different accounts of values—accounts that are in strong tension with mine. Take, for example, this passage from Matt Brown:

> The need for value judgment arises from conflict and uncertainty among our values. If there were one, unambiguous hierarchy of values, or nonoverlapping magisteria of value, each in its own separate unambiguous realm, then the only need for judgment would come from a lack of knowledge of the right values or uncertainty about how to apply them. Instead, in our lived experience we have a plurality of values, and no generally agreed-upon, principled, decontextualized way of ranking them or integrating them prior to the way they play out in our lives. When they conflict, which they inevitably will, we must make judgments. (Brown 2020, 150)

Two specific aspects of this account are in tension with mine. First, it is not apt to speak of value judgments. As Ernan McMullin (1982, 4) points out, "When the value of something is determined by one's attitude to it, the declaration of this value is a matter of value-*clarification* rather than of judgement, strictly speaking." When it comes to values like preferring to save one life rather than preserving one year of schooling, we do not make "value judgments"; we make "value clarifications." We then make value-laden judgments like "It is safe for you to cross the street given how close that truck seems to be." Second, the idea that values are one of the two components of decision puts paid to the idea that there could be multiple overlapping magisteria of values. You may think you have conflicting magisteria of values, but the JHP demurs. It asserts that the right decision problem will always elicit what your univocal values really are. The need for value management comes from the entanglement thesis, and from the fact that not everyone shares the same values, nowhere else.

We have emphasized that all decisions by scientists are value-laden, yet some have called this into question. The JHP can help us to understand what is going on, and why it can appear that some scientific decisions are not value-laden even though they are. Ward, for example, has raised questions about "when and where justification is required in science"

(Ward 2021, 60). She is primarily concerned with the fact that scientists often seem to be blindly following methodological conventions (like using $p<0.05$). When this happens, she notes, they do not need to justify their decisions. This is true but irrelevant. Why can't a scientist who follows methodological convention avoid putting her values into play? A cheap answer is that they value following the conventions for their own sake. This might even be true, but there is a much more serious response that, once again, JHP can help explain. Note that revealed-preference theorists are well aware of the fact that, when stakes are low, such as when I choose a brand of laundry detergent, preferences can take a back seat to things that are very much like conventions. When stakes are low, consumer choices might not accurately reflect their true preferences for various reasons: consumers may simply stick to purchasing a product they have used in the past out of habit or routine, rather than comparing all available options and choosing the one that best aligns with their preferences. They may simply reach for the product that is on the nearest shelf, or for the brand their parents used.

Scientific decision-making is no different. When stakes are low, scientists will revert to conventions.[18] When scientists follow evidence evaluative conventions, does this free them from appealing to their utilities? Many have argued that it does, and the prima facie case is obvious. If there is a methodological rule that tells me, "When the evidence is such and such, you must accept the hypothesis," then what role is there for my utilities? But methodological conventions are only followed when stakes are low! When the stakes get high enough, methodological conventions go by the wayside.[19] And deciding whether the stakes are low or not *is a value-laden determination*. Even when scientists follow conventional standards (like $p<0.05$), it is a decision—because the decision to violate conventional standards is *always available*, and in fact *would be rationally required* on some sets of values. So the Geddy Lee principle continues to apply: "If you chose not to decide, you still have made a choice." The choice not to abandon conventions, which is always available, is always value-laden on the JHP account of values.

Of course, one might reply that these are just examples of times when stakes are involved, and they can't possibly show that stakes are always involved. But the point is a point of logic: if changing the standards in accord with the stakes is always possible, then not doing so has to reflect a value judgment. *Following a methodological standard is not acting without regard to your values.*

6. WHAT IS THE CONNECTION BETWEEN VALUES AND JUDGMENTS? DO SCIENTISTS MAKE "VALUE JUDGMENTS"?

A final issue to clarify is the fundamental difference between a value and a value-laden judgment. These are often confused in the literature, in part, I think, because of the confusing term "value judgment."[20] In science, we often have to make "judgments." When we do, values are usually (maybe always) implicated. But the value and the judgment are distinct.

For our purposes, what's important to note is that there is a world of difference between a scientist encapsulating what judgments she made, on the one hand, and encapsulating what values informed her judgments, on the other. Judging that the evidence is strong enough to infer that the buckle is sound is distinct from having the value that your pants falling down is ten times worse than throwing away a perfectly good buckle. And there's a huge difference between telling someone that you exercised your own judgment in evaluation of the evidence, on the one hand, and telling them the strength of your preference for your pants not falling down versus not throwing away a belt buckle, on the other.

When we are talking about packaging values for transfer, it is especially important that we take care to distinguish between values and judgments. Was a sampling method adequately balanced? Is dataset A better than dataset B? Should race be included in a vaccine model? (Harvard et al. 2021). The answers to these questions are examples of *judgments*. How much worse is it for an effective vaccine to be withheld from the market than for a dangerous vaccine to be released? Is the harm of a contaminated water supply more than ten times larger than the benefit of saving $100,000 on a municipal water system? Is the benefit of a new pesticide in increasing food supply as large as the harm of destroying a local ecosystem? The answers to these questions are examples of *values*.

It is a central dogma of the values-and-science literature that when we make methodological judgments (like the above), values (like the above) are inevitably involved. If someone is deciding whether dataset A is better than dataset B, it will be hard for them to avoid the realization that dataset A is more likely to lead to type I error, and so it will be hard for them to set aside how bad they think the consequences of such errors are. If someone is deciding whether to include race in a vaccine model, the judgment will be entangled with their views regarding whether the overall net health benefit to society should be partially sacrificed in order to avoid racial disparities. This is the *entanglement thesis*. Despite this entanglement, values and judgments are distinct. (This is Putnam's point in

the epigraph.) It is inapt to say there is "value judgment" when someone judges that race should be omitted from a vaccine rollout model. There is at best value clarification (i.e., whether they value net public health more than avoiding racial disparities, and by how much). If that's the particular attitude to public health and racial disparities that they have, no judgment is required. If they decide to put race in the model, this might involve a judgment on their part, and it involves their values, but philosophers shouldn't call it a "value judgment." The actual values comprise how they weigh the desirability of gains to net public health compared to its racial distribution. The choice to include race in the model is a complex mixture of epistemic considerations and those attitudes.

To see that this confusion sometimes arises in the literature, consider Elliott's (2021) case study on Lyme disease. In it, he notes that scientists who study Lyme disease

> appear to be placing greater value on avoiding negative side-effects from antibiotic treatment than on doing everything possible to address the experiences of those suffering from alleged long-term Lyme symptoms. … [But many patients] are willing to take risks and try almost anything in an effort to alleviate their long-term symptoms. (Elliott 2021, 5)

This is a clear example of conflicting values in the sense of preferences over outcomes: one side values the dangers of antibiotic abuse very (negatively) highly and alleviation of symptoms less highly; the other side's values are the opposite. But when Elliott talks about the scientists being transparent about their values, this is not among the various things he catalogs. Rather, Elliott focuses on whether scientists should be transparent about outstanding disagreements over the efficacy of antibiotic treatment for Lyme symptoms, about what studies have been determined to be of poor quality, and about what datasets have been set aside.

Why is it so important to carefully distinguish values from the value-laden judgments they are entangled in? The reason is that while the judgments might be easier to package for transfer, encapsulating judgments does not advance the project of managing values nearly as well as encapsulating actual values would, if encapsulating values were practical.

We can work through an example of a proposal for managing values to see why this is so. Let's use Daniel McKaughan and Kevin Elliott's (2013) "transparency plus backtracking" proposal. Daniel Steel summarizes the view as follows:

> [S]cientific information should be provided so that its recipients can *undo the effects* of non-epistemic values they disagree with, and then "track forward" to conclusions that would result *had their own values been used instead*. … And if backtracking were assiduously followed, then the others could, given the scientific information, backtrack and "track forward" to results that would be derived given their own values. (Steel 2017, 53; emphasis added)

All I aim to show here is that transparency *about values* plus backtracking is much more likely to be of use for managing values than transparency *about judgments* plus backtracking. Consider the truck case again. We can use it to see that transparency *about values* plus backtracking isn't *perfect*. If you tell me it's safe to cross the street, and you disclose to me that you value my life 1,234 times as much as me getting to the other side, but I value my life 1,269 times as much as getting to the other side, it's impossible for me to track forward to the result you *would have gotten* if you had exactly my values. But we can also use it to see how transparency *about values* plus backtracking could be very helpful. That's because if I know that your values that are wildly different from mine, I can see that your advice is probably useless to me. If I know our values are very similar, I can probably trust your advice. So knowing your values *could* be helpful to me in addressing worries about wishful thinking, bias, lack of objectivity, political illegitimacy—that is, in managing values.

Even perfect transparency about judgments, however, would lack even these moderate virtues. Suppose you tell me, "I reached the conclusion that you could safely reach the other side, but that's because I used the Acme brand of truck-speed measuring device, rather than the Vandelay Industries brand. And if I had used the latter brand, I would have reached the opposite conclusion." As a nonexpert, this disclosure is nearly useless to me. What have I learned? Even if you tell me, "I used the Acme brand because I wanted to err on the side of caution," I don't really learn very much. Maybe neither brand would be nearly cautious enough for me!

If value management involves scientists encapsulating information and giving it to their stakeholders, it has to be, at a minimum, encapsulation of values, and not of judgments.

7. WHY IS IT IMPOSSIBLE TO PACKAGE VALUES FOR TRANSFER?

Now we come to a central claim of the paper: packaging values for transfer is practically impossible. There are five nonempirical premises (all concerning the nature of values in science) for this conclusion, which I have already argued for.

1. Values are not just things that are desired but the precise relative degree to which they are desired.
2. Values are not posited as psychological entities. The people that hold them needn't be aware of the exact role they play in their decisions.
3. Values that are relevant in science are not just the positive and negative utility of the foreseeable consequences of, for example, endorsing a hypothesis and being wrong—they are the positive and negative utility of all the possible downstream consequences (from a God's eye view).
4. Values are not just relevant to endorsing or not endorsing truth-apt claims but also are involved in the making of any decision in science—either of endorsing a hypothesis or making a representational choice.
5. Values and credences are combined by decision-makers in ways about which they are not first-personally omniscient. This is the entanglement thesis.

The fifth premise is the one that comes closest to being an empirical one. But it is also a very difficult thesis to give up and still have it that values play an inextricable role in science. The latter claim is as close as one can come to a core commitment in the values-and-science literature. I take it to be true that values play an inextricable role in science first and foremost because I see the entanglement thesis as nearly self-evident. Jeffrey, Rudner, and Putnam all did too.[21] The case study I use here comes from health economics:[22] : the construction of a model to "quantify the potential impact of minimum unit pricing (MUP) for alcohol on alcohol consumption, spending and health in South Africa" (Gibbs et al. 2021, 1). In brief, drinkers in South Africa were divided by age, sex, wealth quintile, and drinker groups (heavy, occasional binge , and moderate drinkers), and elasticity of demand for alcohol was estimated for each bin of drinker. The authors used these values to estimate the change in consumption for each bin under three different possible MUP policies of R5, R10, and R15 per South African standard drink (SD) (12 grams of ethanol). The model considered the possibility of drinkers switching to homebrew alcohol, estimating that 30% of the reduction in recorded alcohol consumption could be compensated for by an increase in homebrew. Sensitivity analysis varied this assumption between 0% and 100%. The model then calculated the potential impact fractions (PIFs) to estimate the impact of a change in exposure to risk on outcomes for different diseases and injuries (HIV, road injury, intentional injury, liver cirrhosis, and breast cancer). They used a "discount rate" for future costs saved (for example, costs of people going to the hospital after a drunk driving accident in 2029 is discounted to compare it to a cost of the program in 2023).

I hope many readers will find this essentially nonempirical argument more or less convincing. But I also think that premise (2) is worth fleshing out with some empirical considerations. In other words, I think it is worth drawing attention to a case study to illustrate, empirically, just how rich and thick the set of possible downstream consequences of any scientific decision is, and, in parallel, how rich and thick the set of values that lies behind every scientific decision is.

On the basis of model results, the authors estimated that

> an MUP of R10 per SD would lead to an immediate reduction in consumption of 4.40% (–0.93 SD/week) and an increase in spend of 18.09% [with] absolute reduction [being] greatest for heavy drinkers (–1.48 SD/week), followed by occasional binge drinkers (–0.41 SD/week) and moderate drinkers (–0.40 SD/week). (Gibbs et al. 2021, 1)

The authors concluded that over twenty years, there would be "20,585 fewer deaths and 900,332 cases averted across the five health-modelled harms" (1).

As we have discussed, both representational decisions and decisions about what facts to endorse necessarily involve values. We can already see that there is an enormous number of representational decisions in a model like the MUP model. Furthermore, each representational decision invokes a wide range of values. Take, for example, the representational decision to apportion drinkers into wealth *quintiles*, rather than tertiles, deciles, or some alternative. Now consider what would happen if we were to switch from quintiles to deciles: this change would enable us to calculate a value for the amount of wealth redistribution that flows from, for example, the second wealth decile to the lowest wealth decile. *This is something we cannot see when wealth is divided into quintiles.* Note that if one thinks that impoverishing the lowest decile, even if it is to the benefit of the second lowest, is a terrible thing, then one will likely judge that the model should use deciles. On the other hand, if one thinks that the benefits of lives saved by the intervention is much greater than the harm of such a wealth redistribution, one will likely judge that the model should *not* use deciles, but quintiles, on the grounds that deciles would provide information (about wealth redistribution) that is apt to distract from the laudable goals of the intervention. This decision can also be pushed around by the value one attaches to benefits of the intervention other than lives saved.

Second, consider the discount rate in the model. A discount rate assumes a particular time preference. People have different preferences when it comes to the trade-off between present and future consumption of utility. Some individuals might have a strong preference for immediate gratification, while others might be more willing to delay consumption to achieve long-term goals or benefits. The subjective value that individuals place on future outcomes can vary depending on the time horizon. For instance, someone might have a strong preference for receiving a benefit one year from now compared to ten years from now, but they might be indifferent about receiving the same benefit in ten years versus eleven years. The way people perceive and value different time horizons is highly individual and can change over time. And the idealization of a discount rate assumes that everyone's time preference has a logarithmic function, and this is almost certainly false. For many outcomes, most of us have a time preference function, just for one example, that drops off quite sharply around the time that we expect to die, and hence is not logarithmic at all. Literally any mathematical function from time to utilities can be one's time preference function.

It has been noted by several philosophers that when a model needs a discount rate, the choice of discount rate is clearly value-laden (Frisch 2013; Schroeder 2017; Winsberg 2018). But the problem is much deeper than this. What is important to understand is that it is *not just that the modelers put a discount rate into the model explicitly*. It is that, in fact, all of the values mentioned above—like the value of impoverishing one decile at the expense of the other— could in principle be discounted in the implicit decision-theoretic reasoning of the modelers. Someone might disapprove of a policy that transfers purchasing power from the bottom decile of earners to the rest of the population in the year 2024, but (because they expect people to be lifted out of poverty over the next 30 years) they might be open to a policy that results in a similar transfer in 2054. Someone else might disagree. And such a disagreement between researchers might affect whether they decide to use deciles or quintiles. In the end, a massive number of differences in values that researchers might have, and regarding the rate at which they discount each of them, could affect the simple decision of whether to use deciles or quintiles. And the decision of whether to use deciles or quintiles is just *one* of a whole host of representational decisions that go into making a model.

Those are a small sample of the values implicated in managing representational risk in the model. There are also values associated with

inductive risk. When Gibbs et al. (2021) endorsed the facts that they did, this was likely to have downstream effects on outcomes over which people can have differing values or preferences. The most obvious one is that endorsing the fact that MUP for alcohol would lead to "20,585 fewer deaths and 900,332 cases averted across the five health-modelled harms" (Gibbs et al. 2021, 1) is likely to increase the chance that local or national governments will adopt an MUP policy. Other possible outcomes include a reduced likelihood that governments will study or adopt alternative harm-reduction policies, such as public awareness campaigns, reducing hours during which alcohol is allowed to be sold, implementing a tax on alcohol, providing treatment for alcohol abuse, increasing law enforcement on alcohol-related crimes, and regulating advertising of alcoholic beverages.

8. SOME CONCLUSIONS

We have reached the conclusion that a wide swath of proposals for how to manage values in science depends on values being packaged for transfer, and that this packaging is a practical impossibility. We have also reached the conclusion that while methodological judgments *can* be packaged for transfer, it does not contribute substantially to value management for them to be transferred from scientist to stakeholders. This leaves one tantalizing possibility: for methodological judgments to be transferred from stakeholders (e.g., the public) to scientists. If the set of values that informs the decision whether to use wealth deciles or quintiles is too rich and complex to be packaged for transfer from public to scientists, why not simply let *the outcome of the decision* be so transferred? If my friend the truck kinematician needs to choose between the Acme brand and the Vandelay Industries brand of truck-speed measuring devices, but ultimately I will be the one who suffers the consequences of the decision, why should she not involve me in the decision? And why should she not involve me in the decision, at the end of the process, about whether to endorse the fact that it is (or isn't) safe for me to cross?

We already know why I can't make these decisions myself. The decisions require a complex, entangled mixture of epistemic judgments and value clarifications. And I don't, on my own, have epistemic expertise. Collaboration arguably provides the best prospect for us mutually coming to the decisions that best combine my friend's expertise and my values.

The proposal lurking behind this tantalizing possibility is, of course, public participation in science. This is not the space to lay out a detailed proposal, and the merits of any specific proposal would have to be debated

on the basis of detailed empirical-sociological evidence regarding how the public can and does interact with scientists. The obstacles, indeed, are fairly obvious. How can scientists and the public coordinate their collaborations so that they reflect, to the largest degree possible, scientists' expertise and the public's values, rather than scientists' values and the public's ignorance? Imagine a proposal wherein a "minipublic" was assembled to collaborate with the makers of the MUP model. The members of the minipublic would need to have carefully explained to them all of the important methodological choices the model makers were facing before the minipublic could weigh in on them. But who could do this explaining other than the model makers? And how could they do this (explain all of the relevant science) without presupposing any value commitments? And perhaps most seriously, how would power be managed in such a dynamic? Even if it becomes reasonably clear that scientists prefer one modelling choice over another because of their upper-middle-class, knowledge-worker-status values, and the minipublic demurs, how would this conflict get resolved, in practice? These are all very difficult questions, and they may very well point to problems that are as bad, if not worse, than the packaging-for-transfer problems I have surveyed above.

Readers interested in managing values in science may find all of these conclusions overly pessimistic. But I would argue that it is better for philosophers to face these problems head-on, armed with a realistic picture of the role of values in scientific decision-making, than for them to produce proposals that are insufficiently attentive to real world complexities. If philosophers aim to contribute to well-informed debates about how to manage values in science, we must be as clear as possible about what it means to manage values in science, starting with what, exactly, those value are. I hope to have made a small contribution to that in this paper.

NOTES

1. Many people have been enormously helpful in producing this paper. The idea for the paper started after presenting my earlier work in Cambridge at the HPS departmental colloquium, especially because of a question from Rune Nyrop. Rune and Marion Boulicault gave valuable feedback after I circulated a draft to the values and science reading group at Cambridge. Kevin Elliot, Matt Brown, and Zina Ward gave valuable feedback in a subsequent round, as did two anonymous referees at a previous journal when the paper had a rather different form. I presented the paper at the University of Bristol and at SOCRATES at the University of Hannover, where Richard Pettigrew and

others, and Matt Brown, Jacob Stegenga and others, respectively, gave me valuable feedback. Stephanie Harvard has looked at every increment of every draft and has helped to craft, at this point, practically every sentence. The paper would have surely been dead in the water many times without her help and encouragement.

2. As of this writing, this proposal has only been defended in public presentations, not in writing. John is partly motivated by a skepticism similar to mine about packaging values for transfer from the public to scientists, but he seems to have a picture of the role of values in science wherein policymakers can identify what "value perspective" a scientific group is coming from. As we will see, my picture of the role of values in science means one can't just identify the value perspective a group has, or really even identify a set of values with a group.
3. One can see Kourany as recommending something like this: "[T]hese social values should be chosen so as to meet the needs of society … . [T]hose would be the morally justified political conditions under which scientific research would be pursued" (Kourany 2010, 68; partly quoted in Brown 2020, 29). As Brown points out, this suggest that the values should be chosen by a philosophical procedure rather than a democratic one. Anderson (2004) has a similar view, especially if one pairs her observations on the role of values in science in this piece with what she has to say about how the best values should be chosen in her book (Anderson 1995). As Elliott (2022) points out, there are problems with these proposals that go beyond the difficulties with packaging values for transfer. Moreover, once we appreciate, as we will, that the values in science are the utilities of decision theory, and see just how ubiquitous they are, it becomes clear that the deliverances of ethical experts, whatever other problems a proposal that incorporates them might face, will underdetermine the values that are needed to make all of the relevant scientific decisions—especially in model building for policy-guiding science. So even if you think the deliverance of ethical experts can be packaged for transfer, it will only be because they are limited (and therefore inadequate).
4. See Korf and Elliott (2024) for a nice survey of such accounts.
5. The terms values, preferences, and utilities are used essentially synonymously in much of decision theory, and I follow that convention here. Sometimes we find the following more specific convention: values are the things I want, preferences are the ordinal rankings of them, and utilities are the cardinals I attach to them. In health economics it is sometimes claimed that values are different than preferences, but this is mostly a point about how values are more difficult to elicit, because it requires more reflection on the part of

the subject to know what their own values are (Shiell, Hawe, and Seymour 1997). It is also sometimes claimed that utilities are what are elicited under uncertainty, and values are what are elicited in its absence. These points do not concern us here. We are not interested in methodological questions in experimental economics about how to elicit values. The "values in science" are, for reasons we discuss above, necessarily cardinal valued utilities.

6. In addition, it will follow that ethics massively underdetermines values. In the truck vignette, I might think it is worth the risk to cross the street for chocolate ice cream, but not for vanilla ice cream. You might think the opposite. Ethics will not settle our dispute.
7. This is very tightly connected to the idea in the philosophy of mind and action that actions can only be explained by both beliefs and desires. As Piñeros Glasscock and Tenenbaum (2023) write, "[T]he explanation of an action involves a 'primary reason': a belief and a desire pair that rationalizes the action by expressing the end pursued in the action (desire) and how the agent thought the action would accomplish this end (instrumental belief). So, for instance, in 'Larry went to Gus the Barber because he wanted a haircut,' Larry's action is explained by a desire (his wanting a haircut) and a belief, left implicit in this case (his believing that he could get a haircut by going to Gus the Barber)."
8. Rudner of course talked of accepting and rejecting hypotheses, not endorsing them. But Hempel, in defining inductive risk, emphasized that it is action that involves risk, and endorsing is the relevant action, not accepting. So I will use that verb here, following Harvard and Winsberg (2022).
9. It makes no difference to the present discussion if our framework is expected utility maximization (EUM) or if we use Buchak's (2017) risk-weighted expected utility maximization, or if we have ambiguity aversion. All that matters is that decisions require utilities. We can read both Rudner and Jeffrey as implicitly endorsing EUM, but this doesn't matter very much.
10. This is not the difference between Rudner and Jeffrey that is usually emphasized. Usually, it is emphasized that Rudner expected scientists to accept and reject hypotheses, while Jeffrey thought they should merely assign probabilities. It is also sometimes claimed that Jeffrey thought that by doing this, scientists could remain value free. Jeffrey neither thought this nor is it true. (See Harvard and Winsberg 2022, section 4).
11. It is sometimes said, moreover, that decision theory involving expected utility is incompatible with deontology. But for our purposes, this is a confusion directly related to the fact that decision theory is not a theory of right action. The deontologist is welcome to read everything I say here and insist

that some preference rankings are impermissible. They are welcome to say that valuing an outcome where everyone is extremely happy but one child is being tortured in the basement is impermissible. But that doesn't mean that their own decisions aren't guided by their expected utilities; it just means that some decisions should result in infinitely negative utility according to them. Decision theory isn't a moral framework, let alone a utilitarian or consequentialist one.

12. Strictly speaking this is of course not right, since you might have diminishing marginal utility in As, or it might be that one A is useless to you unless you have a second one. There is no perfect procedure for measuring your utilities (see the Ramsey [2016] quotation in section 5 for a similar point). But this is in principle what they are: disposition to choose if forced.
13. In fact, if I really want you to err on the side of caution *come what may*, I might as well not ask your advice and just decide for myself to remain safely on the original side of the street.
14. Take, for example, Korf and Elliot's (2024) third category of values conceived as "beliefs or attitudes about desirable things." Note that conceiving of values as "beliefs" is not consistent with the logicist conception of utilities—even "attitude" is not the perfect word. What we are talking about is a stance we take on people's dispositions to act.
15. I mean "stance" in roughly Dennett's (1989) sense.
16. Think again of Piñeros Glasscock and Tenenbaum's (2023) barber vignette.
17. I'm leaving Rudner out of the name only because he identifies values too closely with the "seriousness of mistakes" rather than all the downstream consequences of mistakes and correct inferences.
18. Stakes being low is only one reason that scientists might employ conventional epistemic standards. They might also tailor-make special conventional epistemic standards when stakes are high. All that matters here is that the enterprise will always be value dependent.
19. Consider, for example, first-person accounts from participants in a qualitative study of health economics modelers (Harvard, Werker, and Silva 2020), which reveal that modelers consider the stakes at hand when they decide what standards to apply.
20. This comes, I believe, from McMullin wanting to use the phrase value both for a desirable thing and as a "criteria for choice." The latter comes from the famous Kuhn (1977) essay, "Objectivity, Value Judgment, and Theory Choice," where Kuhn refers to criteria for paradigm choice as values in a few places. I think this is a source of confusion in the values-in-science literature.

The "epistemic values" of Kuhn's famous essay are not the values of values in science. (See Winsberg [2023] for details).

21. Putnam (2004) says it explicitly. Rudner (1953, 4) says it when he says that "the determination that the degree of confirmation is say, *p*, or that the strength of evidence is such and such … is clearly nothing more *than the acceptance by the scientist of the hypothesis that the degree of confidence is p or that the strength of the evidence is such and such*." And Jeffrey (1956) says it when he says that replacing hypothesis acceptance and rejection with probabilities only gets you to the value-free ideal (VFI) if you believe in an objective theory of confirmation, which he did not. I am deliberately being a little vague here about whether I think the entanglement thesis is the price of admission for the inextricability thesis, or if it's literally the price of admission for denying the tenability of the VFI. The reason is that I think that even if everyone were omniscient about their credences, it would be irrational most of the time to reveal them (see Harvard and Winsberg [2022] for details). So it would be *possible* to hold that the values are extricable but that it's never rational for scientists to extricate them, and that that's why the VFI is untenable. But I don't think anyone in the literature holds this view.
22. The case is merely illustrative: a similar exercise could be carried about using almost any policy-evaluative model or study.

WORKS CITED

Anderson, Elizabeth. 1995. *Value in Ethics and Economics*. Cambridge, MA: Harvard University Press.

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

Brown, Matthew J. 2020. *Science and Moral Imagination: A New Ideal for Values in Science*. Pittsburgh, PA: University of Pittsburgh Press.

Buchak, Lara. 2017. *Risk and Rationality*. Oxford: Oxford University Press.

Dennett, Daniel C. 1989. *The Intentional Stance.* Cambridge, MA: MIT press.

Elliott, Kevin C. 2021. "The Value-Ladenness of Transparency in Science: Lessons from Lyme Disease." *Studies in History and Philosophy of Science* 88: 1–9.

Elliott, Kevin C. 2022. *Values in Science*. Cambridge: Cambridge University Press.

Elliott, Kevin C., and Daniel J. McKaughan. 2014. "Nonepistemic Values and the Multiple Goals of Science." *Philosophy of Science* 81 (1): 1–21.

Frisch, Mathias. 2013. "Modeling Climate Policies: A Critical Look at Integrated Assessment Models." *Philosophy & Technology* 26 (2): 117–37.

Gibbs, Naomi, Colin Angus, Simon Dixon, Charles Parry, and Petra Meier. 2021. "Effects of Minimum Unit Pricing for Alcohol in South Africa across Different Drinker Groups and Wealth Quintiles: A Modelling Study." *BMJ Open* 11 (8): e052879.

Harvard, Stephanie. 2024. "Making Decision Models Fit for Purpose: The Importance of Ensuring Stakeholder Involvement." *PharmacoEconomics* 42 (3): 249–52.

Harvard, Stephanie, Gregory R. Werker, and Diego S. Silva. 2020. "Social, Ethical, and Other Value Judgments in Health Economics Modelling." *Social Science and Medicine* 253: 112975.

Harvard, Stephanie, Eric Winsberg, John Symons, and Amin Adibi. "Value Judgments in a COVID-19 Vaccination Model: A Case Study in the Need for Public Involvement in Health-Oriented Modelling." *Social Science & Medicine* 286 (2021): 114323.

Harvard, Stephanie, and Eric Winsberg. 2022. "The Epistemic Risk in Representation." *Kennedy Institute of Ethics Journal* 32 (1): 1–31.

Harvard, Stephanie, and Eric Winsberg. 2023. "Patient and Public Involvement in Health Economics Modelling Raises the Need for Normative Guidance." *PharmacoEconomics* 41 (7): 733–40

Harvard, Stephanie, and Eric Winsberg. 2024. "'Managing Values' in Health Economics Modelling: Philosophical and Practical Considerations." *Social Science & Medicine* 358: 117256.

Hempel, Carl G. 1965a. *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science*. New York: Free Press.

Hempel, Carl G. 1965b. "Science and Human Values." In *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science*, 81–96. New York: Free Press.

Jeffrey, Richard C. 1956. "Valuation and Acceptance of Scientific Hypotheses." *Philosophy of Science* 23 (3): 237–46.

Korf, Rebecca, and Kevin Elliott. 2024. "Values in Science: What Are Values, Anyway?" *European Journal for Philosophy of Science* 14 (4): article 53.

Kourany, Janet A. 2010. *Philosophy of Science after Feminism*. Oxford: Oxford University Press.

Kuhn, Thomas S. 1977. "Objectivity, Value Judgment, and Theory Choice." In *The Essential Tension: Selected Studies in Scientific Tradition and Change*, 320–39. Chicago: University of Chicago Press.

Lusk, Greg. 2020. "Political Legitimacy in the Democratic View: The Case of Climate Services." *Philosophy of Science* 87 (5): 991–1002.

McKaughan, Daniel J., and Kevin C. Elliott. 2013. "Backtracking and the Ethics of Framing: Lessons from Voles and Vasopressin." *Accountability in Research* 20 (3): 206–26.

McKaughan, Daniel J., and Kevin Elliott. 2014. "Moral Molecules and Love Drugs: Objectivity, Understanding, and Backtracking." Paper presented at the Normative Aspects of Science Communication Conference, Ames, IA. https://doi.org/10.31274/sciencecommunication-180809-101.

McMullin, Ernan. 1982. "Values in Science." *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association,* vol. 1982, no. 2: 3–28.

Parker, Wendy. 2024. "The Epistemic Projection Approach to Values in Science." *Philosophy of Science* 91 (1): 18–36.

Parker, Wendy, and Greg Lusk. 2019. "Incorporating User Values into Climate Services." *Bulletin of the American Meteorological Society* 100 (9): 1643–50.

Piñeros Glasscock, Juan S., and Sergio Tenenbaum. 2023. "Action." In *Stanford Encyclopedia of Philosophy* (Spring 2023 edition). Edited by Edward N Zalta and Uri Nodelman. https://plato.stanford.edu/archives/spr2023/entries/action/.

Putnam, Hilary. 2004. *The Collapse of the Fact/Value Dichotomy and Other Essays.* Cambridge, MA: Harvard University Press.

Putnam, Hilary. 2012. "For Ethics and Economics without the Dichotomies." In *The End of Value-Free Economics,* edited by Hilary Putnam and Vivian Walsh, 111–29. New York: Routledge.

Ramsey, Frank. P. 1931. *The Foundations of Mathematics and Other Logical Essays*. London: Kegan Paul, Trench, Trubner.

Ramsey, Frank P. 2016. "Truth and Probability." In *Readings in Formal Epistemology: Sourcebook*, edited by Horacio Arló-Costa, Vincent. F. Hendricks, and Johan van Benthem, 21–45. Cham: Springer International Publishing.

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

Schroeder, S. Andrew. 2017. "Using Democratic Values in Science: An Objection and (Partial) Response." *Philosophy of Science* 84 (5): 1044–54.

Schroeder, S. Andrew. 2021. "Democratic Values: A Better Foundation for Public Trust in Science." *British Journal for the Philosophy of Science* 72 (2): 545–62.

Schroeder, S. Andrew. 2022. "Thinking about Values in Science: Ethical Versus Political Approaches." *Canadian Journal of Philosophy* 52 (3): 246–55.

Shiell, Alan, Penelope Hawe, and Janelle Seymour. 1997. "Values and Preferences Are Not Necessarily the Same." *Health Economics* 6 (5): 515–18.

Steel, Daniel. 2017. "Qualified Epistemic Priority: Comparing Two Approaches to Values in Science." In *Current Controversies in Values in Science,* edited by Kevin C. Elliott and Daniel Steel, 49–63. New York: Routledge.

Ward, Zina. B. 2021. "On Value-Laden Science." *Studies in History and Philosophy of Science Part A* 85: 54–62.

Winsberg, Eric. 2012. "Values and Uncertainties in the Predictions of Global Climate Models." *Kennedy Institute of Ethics Journal* 22 (2): 111–37.

Winsberg, Eric. 2018. *Philosophy and Climate Science*. Cambridge, MA: Cambridge University Press.

Winsberg, Eric. 2023. "A Moratorium on Talk of Epistemic and Non-Epistemic Values in Science." *Erics's Substack,* May 23. https://winsberg.substack.com/p/a-moratorium-on-talk-of-epistemic.