Rational Endorsement

Will Fleisher

This is a pre-print of an article published in *Philosophical Studies*. The final authenticated version is available online at:
https://doi.org/10.1007/s11098-017-0976-4

Abstract

It is valuable for inquiry to have researchers who are committed advocates of their own theories. However, in light of pervasive disagreement (and other concerns), such a commitment is not well explained by the idea that researchers believe their theories. Instead, this commitment, the rational attitude to take toward one's favored theory during the course of inquiry, is what I call endorsement. Endorsement is a doxastic attitude, but one which is governed by a different type of epistemic rationality. This inclusive epistemic rationality is sensitive to reasons beyond those to think the particular proposition in question is true. Instead, it includes extrinsic epistemic reasons, which concern the health of inquiry more generally. Such extrinsic reasons include the distribution of cognitive labor that a researcher will contribute to by endorsing a particular theory. Recognizing endorsement and inclusive epistemic rationality thus allows us to smooth a tension between individual rationality and collective rationality. It does so by showing how it can be epistemically rational to endorse a theory on the basis of the way this endorsement will benefit collective inquiry. I provide a decision theoretic treatment for inclusive epistemic rationality and endorsement which illustrates how this can be accomplished.

1 Introduction

Consider the following case:

> Ellie is an entomologist, studying the brilliantly colored Madagascan Sunset Moth. Somewhat unusually, the coloration on this moth's wings are not the result of pigmentation, but of complicated light interference and polarization patterns caused by the micro-structure of its wings. The common view among Ellie's colleagues is that the function of this coloration is aposematic, i.e., to warn potential predators of the moth's toxicity.
>
> However, Ellie is aware of a new hypothesis, that the color and polarized light patterns are actually a signaling method between moths. She is not confident in the truth of this hypothesis, as no behavioral studies have yet been done. In fact, she is more confident in the aposematic theory. However, since few entomologists are working on the signaling theory, she decides to pursue it.
>
> As she garners evidence for the signaling theory, she advocates for it in published work and in discussion, defending it from various objections. She sometimes asserts its truth, or defends it as the best view. It becomes her favored theory to defend, and she remains committed to it in her work over time, even though most of her colleagues disagree with her. However, outside of professional activities, Ellie will admit that she still does not believe it. She is not that confident it is right.[1]

Ellie's situation is a familiar one: researchers often find themselves committed to theories in the face of disagreement with their peers. Moreover, it seems clear that Ellie is not behaving irrationally. Indeed, her actions are consistent with being an excellent researcher. A healthy community of inquiry, in any field, needs researchers who will strike out to pursue theories which are not currently the most probable, or best-confirmed. Inquiry needs researchers like Ellie.

Ellie does not believe that the signaling theory is true.[2] Nor, I would suggest, should she believe it. Nonetheless, she is committed to the view, defends it, advocates for it, directs her future research on the basis of it,

[1] This example is inspired by the actual case of the Madagascan Sunset Moth, *Chrysiridia rhipheus*. For details on the signaling hypothesis, see Yoshioka and Kinoshita (2007).

[2] Nor does she believe that it is approximately enough true, or that it is the best theory, or even that it is empirically adequate in van Fraassen's sense (1980).

and even asserts that it is the right theory. These actions are generally associated with beliefs, and with epistemic standards which Ellie clearly does not meet. She does not know the signaling theory is true, so she neither obeys the knowledge norm of assertion, nor does she base her actions only on premises she knows.[3] So how can we account for the apparent rationality of her behavior?

In order to explain and justify Ellie's actions, we must recognize a distinct attitude that researchers have toward their theories. I call this attitude *endorsement*. As I will argue, endorsement is the rational attitude to take toward one's favored theory during the course of inquiry.

Endorsement is governed by what I call *inclusive* epistemic rationality. This is distinct from the veritistic (or accuracy-first) epistemic rationality that plausibly governs full belief and credence (Goldman 1986; Joyce 1998; Pettigrew 2016). It is also distinct from a true pragmatist view, which denies any distinction between pragmatic and epistemic rationality (Rinard 2015). This notion of inclusive rationality is necessary to account for the actions of excellent researchers like Ellie. Sometimes researchers are sensitive to reasons which speak in favor of pursuing a theory, but which are not reasons to think the theory is true. Yet these are not simply pragmatic reasons: they are considerations about what is good for inquiry. I will call such considerations *extrinsic* epistemic reasons (following Steel 2010).

In this paper, I will argue that recognizing endorsement is necessary in order to provide a proper account of inquiry, and the norms that govern it. The paper has two goals. The first goal is to characterize when it is epistemically rational to endorse a theory. The second goal is to show that the right account of rational endorsement will allow us to smooth some apparent tensions between individual and collective rationality. The successful completion of these two projects shows that a theory of inquiry which includes endorsement provides a better account of inquiry, and prescribes better norms for inquiry, than one which does not recognize the attitude.

In what follows, I will briefly give an account of the nature of endorsement. Then, I will describe the valuable features of inquiry to which endorsement contributes (§3). In §4, I will give an account of inclusive

[3]One can even swap out "knows" here for "is justified in believing," and the knowledge norm for some other norm. Plausibly, Ellie will also fail to meet such weakened requirements. For the knowledge norm of assertion, see Williamson (1996, 2000). For an overview of the norms of assertion literature, see Pagin (2016); Weiner (2007).

epistemic rationality. Then, in §5, I will discuss some extrinsic epistemic reasons derived from the social epistemology literature regarding the distribution of cognitive labor. In §6.1, I argue that endorsement allows us to account for the way individual researchers are sensitive to these extrinsic reasons, and thereby helps smooth a tension between individual and collective rationality. Finally, I will provide a formal decision-theoretic framework for describing rational endorsement which reflects the foregoing considerations (§6.2).

2 Endorsement

Endorsement is a propositional, doxastic attitude; i.e., it is an attitude one takes toward a proposition. It is the appropriate attitude to take toward one's favored theory during the course of inquiry, within the domain of a cutting-edge research field.

This notion of endorsement is inspired by work on the distinction between acceptance and belief, especially by the work of L. Jonathan Cohen (1989a, 1989b, 1995), Isaac Levi (1974, 1980, 2004), Patrick Maher (1993), and Bas van Fraassen (1980). However, there are a dizzying array of different notions of acceptance[4] and my concept of endorsement is distinct from these previous ideas of acceptance in a variety of ways. Most importantly, endorsement is distinctively epistemic and provisional.[5] It is sensitive to both intrinsic and extrinsic epistemic considerations (see section 3 below for more on this distinction). It is an attitude one takes during the "context

[4]There are at least five kinds of "acceptance" notions that appear in philosophy: Cohen's notion from epistemology (Alston 1996; Cohen 1989a), the notion from the philosophy of language (R. C. Stalnaker 1987), the notion from the philosophy of science (Kaplan 1981a, 1981b; Levi 1974; Maher 1993; Van Fraassen 1980), the concept of acceptance from the metacognition literature (Frankish 2004; Proust 2013), and the genus conception of acceptance (Shah & Velleman 2005). For more on the various notions of acceptance, see McKaughan (2007).

[5]I am not the first to notice the need for a provisional, acceptance-like attitude. Goldberg notices this need in the context of pervasive disagreement in philosophy (2013a; 2013b). Other examples of varying degrees of similarity can be found in Firth (1981); Lacey (2015); R. Laudan (1987); McKaughan (2007); Whitt (1985, 1990). I take this convergence to be good evidence in favor of the existence of the attitude I call endorsement. However, my account is significantly divergent from prior accounts, and I apply the theory to both philosophy and science.

of pursuit" (L. Laudan 1978), and is importantly involved with planning future research.[6]

The following is a characterization of endorsement. It describes a set of features of the attitude which serve to distinguish endorsement from other kinds of mental states, including belief. I take it that this characterization is compatible with any particular view of the nature of mental states, with the possible exception of eliminativist views.

> **Endorsement:** Endorsement is a doxastic propositional attitude. S endorses p in a research domain d only if:
>
> 1. S is disposed to assert that p, or otherwise express commitment to p (in d).
> 2. S takes herself to be obligated to defend p (in d).
> 3. S treats p as a premise in her further reasoning (in d).
> 4. S shapes her research program in d (in part) based on p.
> 5. S is resiliently committed to p (in d).
> 6. S takes p to be a live option (i.e., she does not know p is false).
> 7. In endorsing p, S aims to promote healthy inquiry.

Endorsement is specific to a research domain: a subject endorses something for the purposes of a particular domain of inquiry. Characterizing these domains might be tricky in some cases, but I am conceiving of them as standard, familiar divisions of inquiry: fields and sub-fields of research (e.g., physics, or epistemology). Which theories are candidates for endorsement depends on the particular domain of inquiry.[7]

[6]Following Laudan, a number of philosophers of science have appealed to this notion of the context of pursuit. This context is the stage of inquiry when researchers pursue promising but as-yet unconfirmed theories. For an overview of this literature, see McKaughan (2007) and Whitt (1990). Other examples include R. Laudan (1987); McKaughan (2007, 2008); McMullin (1976); Nickles (1981); Šešelja, Kosolosky, and Straßer (2012); Šešelja and Straßer (2013, 2014); Whitt (1985, 1992). There is insufficient space here to show all of the applications of endorsement, and inclusive epistemic rationality, to the pursuit literature.

[7]Endorsement is a propositional attitude. However, this does not mean that the theories being endorsed need to be understood as propositions. I want to remain neutral on the nature and structure of scientific theories (Frigg & Nguyen 2016; Winther 2016). Technically speaking, the propositions that the attitude is taken toward can be propositions about the theory. So, if one has the view that, for instance, theories are actually models, then the proposition one endorses could be just "Theory A is an accurate enough model" or "Theory A is the best model," or some other variation on these lines.

Endorsement is a "fragmented" attitude, meaning that it is compartmentalized rather than being a global feature of the subject's mental state. It is fragmented in two respects: first, it explains the seemingly inconsistent actions of subjects who are committed to a position during research, but are not confident enough to believe it or act on it outside of the research domain. And second, what one endorses in one domain can be inconsistent with what one endorses in a different research domain.[8]

Endorsement is similar to belief in a number of ways. However, the above characterization provides for distinguishing the two attitudes, given appropriate interpretation of its conditions.

First, condition 5 requires resiliency, as endorsement is more resilient than belief. Endorsement can be permissibly maintained in the face of significant contrary evidence, objections, or purported counter-examples. If one discovers such evidence against a proposition one believes, one should give up that belief. But endorsement is an attitude suitable even in the face of pervasive disagreement and significant contrary evidence. It is the attitude a committed advocate has toward a theory. In this way, endorsement licenses maintaining healthy disagreement. The importance of this will become clear in §3.[9]

A second way endorsement is distinguished from belief is the obligation to defend (condition 2). Quite often, subjects will not be required to defend their beliefs (though of course in other cases they will be). But taking up an endorsement attitude will almost invariably involve some obligation to defend the view. After all, taking a position and advocating for it in a field of research is to be involved in defending the view.[10] One way endorsement is valuable for inquiry is because it motivates and licenses such behavior (as I will argue below in §3).

Endorsement also involves directing one's research on the basis of what is endorsed (condition 4). So, if a researcher endorses a theory, and there is a particularly important method for testing that theory, or a characteristic

[8]For more on the notion of fragmentation, see Egan (2008); Elga and Rayo (2015); Lewis (1982); Rayo (2013); R. C. Stalnaker (1987).

[9]Note that this resiliency is quite distinct from Leitgeb's notion of stability. Stability involves having good reason to expect no evidence that would warrant giving up the belief (2014). Belief is stable but not resilient, and endorsement is resilient but not stable.

[10]This obligation to defend is similar to the role Kaplan (1981a, 1981b) sees for his notion of acceptance. Later, however, Kaplan (1995) explicitly equates this kind of acceptance with belief, which I think is a conflation.

kind of methodology tied to the theory's tradition or heuristic, then this will affect what the researcher should be doing.[11] That is, the researcher should plan to engage in the method, or the associated methodology, because they endorse the theory. This, along with the strong obligation to defend, means that endorsement carries with it certain "pragmatic commitments" that belief does not.[12] Endorsing a theory means directing one's research toward that theory. Although beliefs clearly govern our behavior, believing that P does not generally commit one to researching about P.

The sixth condition concerns the fact that endorsement is an attitude one takes toward theories which are potential answers to the questions which guide a domain of inquiry. It is not the kind of attitude that an engineer takes toward Newtonian mechanics when using that theory for the purposes of bridge-building. For endorsement to play it's role in helping to motivate researchers, it must be that the researchers consider the theory in question to be a live option. A subject endorsing p advocates for p, and wants to see it turn out to be true. This is only possible if endorsements are limited to live options.

The seventh condition is about the aim of endorsement. The aim of endorsement is to promote healthy inquiry. "Healthy" here is just meant as a neutral term to express positive evaluation. Endorsement aims at the collective good of inquiry. Endorsing p aims at promoting the health of inquiry by allowing the subject to appropriately engage in community discourse and debate, and by motivating her to commit to a research program which promotes the community's goals. This connects the aim of endorsement with truth, but only indirectly. Endorsing promotes the community's learning more truths.[13]

The last two conditions also help illustrate the way endorsement is distinct from belief. Belief aims at truth. It seems plausible that there is a sense in which believing falsehoods is simply impermissible, and it is this requirement that generates various other epistemic requirements. However, endorsement's connection with the truth (of the proposition endorsed) is more closely akin to guessing. When guessing, the hope is to

[11] For the notion of a heuristic, or characteristic methodology, see Whitt (1992) and Šešelja and Straßer (2014).

[12] For more on this notion of pragmatic commitment, see Van Fraassen (1980) and Whitt (1990).

[13] I tend to think good inquiry is that which leads to the community learning interesting truths, but "healthy" here could refer to meeting a variety of epistemic standards.

guess the truth, but with little expectation of reliability in achieving this. It is permissible to guess (and endorse) unreliably. This is why endorsement is only appropriate to theories which are not known to be false, but these theories need not be the most likely theory to be true. In contrast, permissible belief requires some kind of reasonable expectation of achieving its aim of getting at the truth, and so believing unlikely things is inappropriate.

Another thing which distinguishes belief and endorsement is that there are different norms for rational belief than there are for rational endorsement. Where to draw the lines between what characterizes the attitude and what the norms of the attitude are is a bit difficult. There is significant overlap between the two projects. However, this does not seem terribly problematic, as belief is generally taken to have constitutive rationality requirements and a constitutive truth aim (Shah & Velleman 2005).

The biggest normative difference between endorsement and belief involves the following principle, which I take to be a bedrock intuitive assumption about belief: it is irrational to believe some proposition p if one takes $\neg p$ to be more probable than p. Put simply, you should not believe something you think is more likely false than true. This principle is not true for endorsement. One should not knowingly endorse something false, but one can endorse something unlikely to be true. Endorsement is an appropriate attitude for theories which should be pursued and advocated for, but which are (at least as yet) unconfirmed.[14]

With endorsement on the table, we now turn to the role it is meant to play in inquiry, which will in turn enable us to derive appropriate norms for rational endorsement.

There are two projects that recognizing endorsement contributes to:

Vindication Explain, rationalize, and justify our current practice.

[14]One might worry that this will permit endorsement of theories which should be ruled out, either for epistemic or moral reasons, e.g., that anthropogenic climate change is not occurring, or pseudo-scientific racist theories. However, there are two ways of resisting the idea that endorsing these theories would be appropriate. First, I think moral reasons are over-riding, so if one has a moral reason not to endorse a theory, this will mean that one should not do so, all-things-considered. Second, and more directly, endorsement is an attitude taken to live options, not options known to be false. The two examples here are both known to be false, and so are not potential candidates for endorsement. Thanks to Briana Toole for discussion on this worry.

> **Prescription** Offer normative guidance to philosophers, humanists, and scientists about the appropriate attitudes to take to their favored theories.

There are both normative and descriptive elements to the vindication project: I want to offer a description that accurately reflects and explains the actions of at least some researchers, while also justifying them. However, the theory of endorsement also provides normative guidance for how to structure inquiry, and this is reflected in the prescription project.

These two projects give us some guidance on what an account of the norms of endorsement should look like. The norms need to be attainable by subjects, without the subjects' being specifically knowledgeable about "endorsement," which is not a univocal pre-theoretical notion. To pursue the vindication project, the norms need to be attainable, so they are met by some actual researchers. For the purposes of providing prescription and justification, the norms need to reflect the features of inquiry which we take to be valuable.

3 The Value of Endorsement

In this section, I want to briefly characterize the valuable features of inquiry that endorsement promotes.

Endorsement is governed by inclusive epistemic rationality, which includes extrinsic reasons. This is justified as a distinct and significant kind of rationality because it includes sensitivity to features of healthy inquiry. These are features which an accuracy-only notion of epistemic rationality leaves out. Because of this, endorsing in accordance with inclusive epistemic rationality promotes better inquiry. There are a number of such features that endorsement promotes.

A central feature of inquiry promoted by endorsement is the resilient commitment of researchers to their theories. It is undesirable for a researcher to drop a costly research program at the first sign of strong contrary evidence. We want researchers who are motivated to continue to defend their theories in the face of difficult objections, because sometimes objections and counter-evidence turn out to be misleading. Endorsement is an attitude that is characterized by this kind of commitment, and the norms of rational endorsement reflect this fact.

Having a resilient commitment to a theory means a researcher will commit significant time and energy to exploring implications of the theory. They will be motivated to develop the best version of the theory, the best possible defense of the theory, and to explain away anomalies and purported counter-examples.

Thus, having researchers who endorse their theories will contribute to the vivacity of debate. To borrow a famous notion, this will encourage a robust "marketplace of ideas" where views will get their full, fair hearing.[15] Researchers will be motivated to be strong advocates for their favored views, and allow them to have their "day in court." This kind of debate is valuable as a feature of inquiry, and a practice of rational endorsement is well-positioned to promote it. The norms of endorsement, therefore, should reflect this.

Similarly, it is valuable for inquiry to avoid certain kinds of undesirable deference on the part of researchers. Usually, when we encounter the advice of experts in a certain field (when we are not members of that field), we should defer to the experts' judgment. When a layperson discusses a theory's import to the broader society, they should defer to experts in the field. This also seems like a good rule for policy-makers. But it would be undesirable to have this deference when actually engaged in research. We do not want all researchers (even novices) to simply defer to the best experts in their field; this would stifle creativity and progress. This non-deference is clearly a necessary requirement of healthy philosophical discourse, for example. And I think the same is true throughout inquiry. Endorsement is also well-positioned to encourage this. Rational Endorsement involves advocacy of a theory, and may be justified even when one's confidence in the theory is lowered by disagreement.

It is highly intuitive that it is valuable to have researchers who serve as committed advocates for views, and who are willing to defend less well-developed views in the face of disagreement. In addition, there is significant empirical evidence that human inquiry and reasoning is improved by group arguments with these features. Groups that engage in debate, and begin with disagreement, are better at getting at the truth, and at producing instances of good argumentation. This has been supported by

[15]The analogy of free expression of ideas to an economic marketplace seems to trace back to Mill through Supreme Court Justice Oliver Wendell Holmes, though it is perhaps an imperfect metaphor for Mill's own view (Gordon 1997).

a number of psychological experiments involving a variety of reasoning and decision-making tasks (Bonner, Baumann, & Dalal 2002; Geil 1998; Kerr, MacCoun, & Kramer 1996; Kerr & Tindale 2004; Laughlin, Bonner, & Miner 2002; Laughlin & Ellis 1986; Mercier 2016; Mercier & Sperber 2011; Resnick, Salmon, Zeitz, Wathen, & Holowchak 1993). When people are challenged in their arguments (or even when they expect to be so challenged), they produce better arguments, recognize problems with others' arguments, and avoid making fallacious inferences.[16]

In addition to the empirical psychology literature, there is also support in the form of case studies from science. In particular, a recent paper by De Cruz and De Smedt (2013) appeals to a case study from paleoanthropology to argue that disagreement is valuable for inquiry. They consider the case of the alleged discovery of *Homo floresiensis*, a proposed hominin species. Paleoanthropologists disagree about whether the skeletal remains of small hominins found in a cave on the Indonesian island of Flores are those of a new species, or whether they are actually the remains of pathological modern humans (i.e., humans with some sort of hereditary condition leading to small stature, myoencephaly, and various other slight differences from more typical human adults).

De Cruz and De Smedt argue that the disagreement among scientists over this question has led to three significant benefits. First, it results in the generation of new evidence, as proponents of each view are motivated to seek new and better evidence to convince their peers. Second, it leads to a reassessment of existing evidence and old assumptions. This is because it motivates scientists to look for ways in which old evidence might provide support for their view, and to look at old assumptions that conflict with their view, but are inadequately substantiated. Third, disagreement helps overcome confirmation bias, since researchers seek evidence and objections to views they oppose, and their disagreeing peers force them to notice and account for objections to their own view.

Endorsement enables the kind of resilient commitment and advocacy of a theory which leads to the valuable disagreement described by both the psychological literature and scientific cases studies.

An additional valuable features of inquiry will be explored in more depth below, in section 5: appropriate distribution of cognitive labor dur-

[16]For an overview of the empirical literature regarding the benefits of debate and disagreement in group reasoning, see Mercier and Sperber (2011) and Mercier (2016).

ing the course of research.

4 Inclusive Epistemic Rationality

Ellie's case (from the very beginning of the paper) is an example of a familiar and widespread occurrence. There are many inquirers who are motivated to contribute positively to inquiry. They are not motivated by personal gain or by fame or fortune (or at least not *only* by such things). Rather, they pursue and commit to a theory partially out of a concern for contributing to healthy inquiry. Ellie just wants to do her part in getting at the truth. Reasons which bear on the health and success of inquiry are genuine epistemic reasons. A view which did not distinguish these reasons from purely pragmatic ones would fail to make an important distinction.[17]

Endorsement is to be distinguished from belief, on the one hand, and from various practically-oriented acceptance notions, on the other, on the basis of the kind of epistemic rationality that governs it. What it is *epistemically* rational to endorse depends on reasons beyond those on which rational belief is based. Yet these reasons are also genuinely epistemic, as we see in Ellie's case. We can still distinguish between this kind of inclusive epistemic rationality and the "anything goes" pragmatist view of Rinard (2015).[18]

Rinard argues that there is no sense of epistemic rationality that can be distinguished from practical rationality (2015). Belief formation is governed by the same rational standards as any other act. Rationality is thus univocal: epistemic considerations are only relevant to an act's choiceworthiness to the degree that a subject happens to value them. I'm calling such a view "anything goes," because any reason or consideration relevant

[17]Of course, none of this is to deny that some researchers really are motivated by prudential reasons, especially fame and prestige. This motivation is not even always a bad thing for science (see Kitcher (1990); Strevens (2003)). I am merely suggesting that some of us are sometimes motivated by a desire to contribute to inquiry.

[18]Throughout, I use terms like "reasons," "considerations," and "values," and treat them as though they are interchangeable. I think my view is compatible with a wide variety of views about the nature of normativity, so one can simply plug in one's favored view from the meta-ethics and meta-epistemology literature. For an overview of available theories, see Alvarez (2016); Broome (2015); Finlay and Schroeder (2015); FitzPatrick (2004); Gert (2009); Parfit and Broome (1997).

to the agent's interests gets counted in determining what is rational for that agent to do.

The anything goes view fails to adequately explain cases like Ellie's, and fails to distinguish them from cases where the subject is motivated entirely by pragmatic reasons. There is a clear sense in which Ellie is justified merely by appeal to considerations about what is good for inquiry. Of course, there might be pragmatic reasons for what she does as well (e.g., it might be good for her career). But there need not be any such practical reasons: Ellie would be justified merely by appeal to what is good for inquiry. Distinguishing between a sense of rationality which is inclusively epistemic, and a sense which is "anything goes," helps us to distinguish a case like Ellie's from one where the subject is motivated only by pragmatic concerns.

Contrary to the arguments of Rinard and other pragmatists, it has long been standard to suggest that there is a sense of epistemic rationality (or justification) which is importantly distinct from prudential rationality. That is, we can evaluate a belief based on purely epistemic merits. However, what considerations, values, and reasons count as epistemic, in this sense, is plausibly more narrow than the inclusive epistemic rationality I will propose below.

Examples of this kind of epistemic consideration are how well a belief is supported by evidence, how coherent a set of beliefs is, or how reliably they were produced. Such concerns might be overridden by moral and even sometimes prudential considerations, but they are clearly separable from non-epistemic considerations.[19]

When evaluating beliefs (and degrees of belief) epistemically, it is plausible that the only relevant considerations concern accuracy. That is, the rationality of a belief is determined solely by features that reflect on *its* truth, or how likely it is to be true. Similarly, whether one's credences are rational is determined by their accuracy (i.e., by how close they are to the vindicated credence function) and features which indicate accuracy. There are two primary motivations for thinking this. The first is the recent success of accuracy-based epistemic utility theory. This program has generated arguments for probabilism, conditionalization, and various other coherence norms by appeal to accuracy alone (Easwaran 2013, 2016;

[19]For a small sample of arguments in favor of a distinctively epistemic domain of evaluation, see Goldman (1986); Shah and Velleman (2005); Sosa (2009, 2015).

Fitelson 2016; Fitelson & Easwaran 2015; Greaves & Wallace 2006; Joyce 1998; Konek & Levinstein 2016; Pettigrew 2016).

The second motivation for thinking that the epistemic rationality of belief and credence is limited to accuracy involves epistemic bribery counterexamples to epistemic consequentialism for belief (Berker 2013; Firth 1981; Greaves 2013; Jenkins 2007). These are cases of intuitively impermissible epistemic bribery designed to show that it is irrational to believe something on the basis of epistemic gains later, or elsewhere in one's belief state. That is, it is irrational to trade having a false belief now to gain more truths later, or to trade a false belief about P for true beliefs about Q and R. In Firth's original example, he describes an atheist researcher who has the opportunity to gain a great many further truths by taking on a belief in god (which he thinks is false).[20]

Intuitively, the atheist scientist is unjustified or irrational in believing in God in order to gain further epistemic value down the line. There is something inappropriate about accepting epistemic badness now for epistemic goodies later. This means that not only is the epistemic rationality of a belief sensitive only to accuracy considerations, the rationality of any particular belief is only sensitive to considerations concerning the accuracy of *that very belief* (and *mutatis mutandis* for degree of belief).

Endorsement, on the other hand, should be sensitive to considerations beyond the truth of the particular proposition in question. As an attitude meant to facilitate inquiry, it needs to be sensitive to the valuable features of inquiry described above (in section 3). In contrast to belief, it can be rational to endorse a proposition because it will lead to downstream epistemic goods, even where that comes at the potential cost of the accuracy of the particular attitude. So, it might be rational for the atheist scientist

[20]Borrowing the formulation in Berker (2013):

"Suppose I am a scientist seeking to get a grant from a religious organization. Suppose, also, that I am an atheist: I have thought long and hard about whether God exists and have eventually come to the conclusion that He does not. However, I realize that my only chance of receiving funding from the organization is to believe in the existence of God: they only give grants to believers, and I know I am such a bad liar that I won't be able to convince the organization's review board that I believe God exists unless I genuinely do. Finally, I know that, were I to receive the grant, I would use it to further my research, which would allow me to form a large number of new true beliefs and to revise a large number of previously held false beliefs about a variety of matters of great intellectual significance. Given these circumstances, should I form a belief that God exists? Would such a belief be epistemically rational, or reasonable, or justified?"

to *endorse* theism, even in the face of the evidence they take themselves to have for atheism, precisely because this will lead to the improvement of inquiry in general. So while I agree about the irrationality of *belief* in the atheist scientist's case, I think there is an attitude to theism that it is appropriate for the scientist to take in order to learn more in the future: endorsement.

Thus, we need a kind of rationality that is sensitive to considerations besides accuracy, but remains purely epistemic. I call this *inclusive epistemic rationality*. We can characterize this notion by appealing to a distinction between *intrinsic* and *extrinsic* epistemic reasons (or values), inspired by Daniel Steel (2010).[21] An epistemic value is *intrinsic* if manifesting the value constitutes an attainment of truth, is necessary for truth, or indicates truth. A reason is intrinsic if it is a reason to believe that a proposition is true, or a reason to accept it as true. Being true is an intrinsic epistemic value, as is consistency.[22]

Epistemic reasons are extrinsic when they are about features which tend to promote attaining the truth, but are neither necessary features of truth, nor reliable indicators. An example Steel suggests is testability (2010, 15). That a hypothesis or theory is amenable to testing is not a reliable indicator that it is true; many testable claims are false. Instead, valuing testability is a methodological commitment which promotes the truth. Considerations of what makes for valuable, productive inquiry are extrinsic reasons. That there is vivacity of debate, or that researchers are motivated to defend a theory, are not necessary conditions of truth. Rather, they promote truth in the long run. In order for endorsement to play the appropriate role in inquiry that I suggest it does, it needs to be sensitive to such considerations. It needs to be sensitive to extrinsic epistemic reasons, as well as intrinsic ones. Hence, it is governed by an inclusive epistemic rationality: one that includes both intrinsic and extrinsic reasons.

[21] Steel's discussion of this idea is brief, so I am uncertain whether my use of this distinction precisely tracks his (2010, 18). Jenkins (2007) appeals to a distinction that is very similar to mine. She distinguishes between "extraneous consequences" of a belief in P, and those which "which directly concern P itself" (37).

[22] The notion of intrinsic epistemic value is related to the idea that belief is "transparent," in the sense explored by Shah and Velleman (2005). Transparency here means that the reasons to believe p are simply reasons for p, or evidence for p. Whenever one considers the question of "whether to believe p," this question is equivalent to the question of "whether p." Beliefs are transparent is because they are only appropriately sensitive to intrinsic epistemic values.

Extrinsic epistemic reasons are reasons for endorsing a theory which stem from the fact that doing so will promote healthy inquiry. They are reasons which (indirectly) promote epistemic goals. However, this is not yet enough to distinguish them from pragmatic reasons. In order to more fully characterize and distinguish this kind of reason, I will provide a response to a potential objection to the idea that these reasons are genuinely epistemic.

This objection concerns the way extrinsic epistemic reasons promote epistemic goals of inquiry. One might admit that these are genuine reasons, because they promote good epistemic outcomes. But then one might still deny that they count as epistemic, because of the way this promotion works. Instead, it might be suggested that these are really pragmatic reasons; just ones aimed at epistemic ends. This worry can be illustrated by appeal to cases which involve "sandwich reasons."[23]

Suppose (plausibly enough) that hunger degrades intellectual performance. Then, eating a sandwich will count as promoting healthy inquiry, because it will lead to better intellectual performance. Suppose also that, if I endorse theory A, this will involve working in a lab near a good sandwich shop, so that I will be less hungry and thus (slightly) better in intellectual performance each day at work.

The worry is that my theory of inclusive epistemic rationality seems to predict that I have a sandwich-based (extrinsic) epistemic reason to endorse theory A, but intuitively this is not an epistemic reason. That is, it is intuitively implausible that I could have a reason to endorse a theory based on the fact that doing so will get me more sandwiches (even if the sandwiches really will improve my epistemic performance).

Happily, there are additional resources to draw upon in distinguishing extrinsic epistemic reasons from mere "sandwich reasons." Extrinsic epistemic reasons are reasons which are internal to a domain of inquiry. Whether something counts as a reason within a certain domain depends on both the goals of that domain and the internal standards of that domain. Performances and states within a domain are evaluated based on whether they promote the goal of the domain, and on whether the way this goal is promoted meets the standards of the domain.[24]

[23] I learned of cases like this one from Nomy Arpaly (2017), who attributes them to Sophie Horowitz. The two objections I am considering here are largely inspired by Arpaly's paper.

[24] Here I am drawing from Sosa (2015), especially Chapter 8.

This way of categorizing reasons can be illustrated by appeal to the domain of chess. The (proximate) aim or goal of a player in chess is winning the game. In order to be a chess-reason, something must promote this goal. For instance, if castling, given the circumstances, will increase the likelihood of my winning the game, this is a chess-reason for me to castle. If I decide to castle because I know it will impress my friend who is watching the game, this is not a chess reason because it does not promote the aim of winning.

There are, however, ways of failing to be a chess reason other than not aiming at the goal proper to the domain. Suppose I have a tattoo on my arm, and if I move my Rook to my opponent's side of the board, this will reveal the tattoo to him. I know that this will intimidate him and cause him to play cautiously and predictably. Thus, the move will promote the goal of winning the game. However, it does so in a way which violates the internal standards of the practice of chess. Intimidating an opponent is not a move which is proper within the domain of chess. If I were playing a different game, this might be totally acceptable (e.g., in the board game *Diplomacy*, which notoriously involves manipulating and deceiving one's opponents). But the standards internal to chess rule this out as a chess-appropriate move.

Domains of inquiry are like chess: they have their own standards which help determine what counts as a reason. Precisely what the internal standards are in research domains is a complicated question, and one I don't think we can answer in full generality. Different domains have different standards based on the kind of research involved. This fact is familiar: it is well-known that different fields have varying standards of evidence. Thus, what counts as an *intrinsic* epistemic reason is clearly determined in part by internal standards of the field of inquiry (though in a way that cannot be totally disconnected from truth if the field is well-functioning). I am suggesting that the same is true for extrinsic epistemic reasons.

Precisely how the internal standards of a research domain are set is a question I will remain neutral on. However, there are at least two clear options. First, the standards could be set merely via convention, or social construction: the social practices of the researchers determine what the standards are.[25] The other way the standards could be set is by appeal to

[25] I think Sosa (2015) has something like this in mind for determining epistemic standards for belief-formation performances.

the essential features of a particular kind of research. The thought here is that research fields involve investigation of different kinds of phenomena, and investigation using different methods, in a way that is non-arbitrary. So the standards of astrophysics will be different than the standards of biochemistry, for reasons having to do with the nature of the phenomena investigated. The essential nature of these investigations then determines the appropriate internal standards of the field. Either way of standard-setting (socially constructed or essential) will work to rule out sandwich reasons.

Domain internal standards will rule in, as extrinsic epistemic reasons, considerations like distribution of resources (and of cognitive labor), but rule out "sandwich reasons." This is clear from scientific practice, and explains our intuitions about sandwiches. Choosing a theory on the basis of the proximity of the relevant lab to a sandwich shop will fail to meet the internal standards of a scientific research domain.

In light of the responses to this objection, we have the following characterization:

> **Extrinsic epistemic reasons** are reasons which indirectly promote epistemic goals of inquiry, and promote them in a way that meets the internal standards of the relevant domain of inquiry.

The expanded, inclusive epistemic rationality of endorsement is part of what makes it a more appropriate attitude to take during the course of inquiry. As we will see, it will allow us to appeal to resources unavailable for belief, including a type of epistemic utility and decision theory that is not appropriate for belief.[26]

Below, I will present a few particular examples of extrinsic epistemic reasons, and show how we can build a decision theory for rational endorsement which incorporates appropriate sensitivity to extrinsic epistemic considerations. In particular, I will focus on some extrinsic reasons provided by insights in the social epistemology literature. The appeal to endorsement, as well as the appeal to inclusive epistemic rationality, is justified in part because it includes additional considerations that are clearly

[26]In order to side-step worries about attitude voluntarism , we can treat inclusive epistemic rationality as providing evaluative standards (rather than deontic norms). This is a common move to make in epistemology: see, e.g., Fitelson and Easwaran (2015). For more on the distinction between deontic and evaluative norms, see Smith (2005).

relevant to our epistemic practice, but which are left out of a traditional, belief-based view.[27]

5 Extrinsic Reasons from Social Epistemology

Endorsement is the appropriate attitude to have toward favored theories during the course of inquiry. This is due in part to its sensitivity to extrinsic epistemic reasons. This sensitivity allows us to solve several epistemological problems in ways that are not available to accounts that only recognize belief. Specifically, it eases certain tensions between individual rationality and collective rationality in inquiry.

One social epistemology issue that endorsement can help with is the appropriate distribution of cognitive labor. The contemporary literature on the topic began with Kitcher's (1990) paper. One of the primary concerns of this literature is an apparent tension between individual and collective epistemic rationality. Once we recognize endorsement, and its sensitivity to extrinsic epistemic reasons, this apparent tension can be easily resolved.

The problem is this: it seems that the rational thing to do, epistemically, is to pursue the most probable theory. However, if every researcher individually follows this advice, it will lead to clearly bad distributions of labor. To see this, suppose that we have two candidate theories, A and B. A is 80% likely to be true while B is 20% likely. We do not want every researcher working on A; after all, A might be false. It would be better to have at least some researchers working on B. But if every researcher does the (apparently) individually rational thing, then all researchers will work on A, and none on B.

In order to resolve this apparent tension, Kitcher (1990) and Strevens

[27] Although there is not room to explore the thought here, I think it is (at least very often) *irrational* to believe one's favored theory. In brief, there are three main considerations that should lower our confidence in theories in cutting-edge research domains: pervasive disagreement, the pessimistic meta-induction, and under-determination of theory by evidence. These problems are characteristic of cutting-edge domains, and so the subjective probability (confidence, or credence) we assign to the theories should be too low to justify belief. Indeed, in many such cases our confidence in the theory should be less than half, in which case full belief is clearly unwarranted. Moreover, as I have argued, the epistemic rationality governing belief is not sensitive to extrinsic reasons which do and should govern our decisions about which theories to be committed to, and to pursue.

(2003) appeal to economic modeling, and the priority rule in science (the convention of awarding all credit for a discovery to the individuals or lab that first succeeds in discovering it). First, they use an economic model to find the optimal distribution of labor, given certain assumptions. Then, they argue that the prestige-seeking behavior of scientists, coupled with the priority rule, will lead to (or at least promote) this distribution of labor.

We can call this kind of economic model of science a *marginal contribution reward* (MCR) system (following Muldoon (2013)). This method considers rules of behavior as they apply to representative members of a research community. The goal is to determine what sort of reward structure is necessary to make it individually rational for a subject to behave so that, when combined with everyone else's, the behavior contributes to the optimal distribution of labor. If the reward structure is appropriately constructed, then it will be rational for an arbitrary, representative member of the research community to behave in a way conducive to community goals.

For ease of exposition, I will focus on a particular version of the MCR strategy, the one developed by Strevens (2003).[28] This is not because I am endorsing Strevens' approach (though I admit some affinity for it). Rather, I am appealing to it as an example, to show that endorsement is compatible with solutions to the distribution of cognitive labor problem, in a way that belief cannot be. I am offering a "proof of concept." I want to show how to implement solutions from the social epistemology literature in the account of individual epistemic rationality using endorsement.

Strevens, like Kitcher, appeals to pragmatic considerations in his MCR account (Kitcher 1990; Strevens 2003). Specifically, he suggests that researchers should be (and are) rewarded with prestige in proportion to their contribution to a project. Without this appeal to pragmatic considerations, MCR accounts cannot deliver the goods of individually rational behavior leading to collectively rational results. In order to achieve collective *epistemic* rationality, they appeal to individual *practical* rationality. However, by appealing to endorsement, we can instead represent the subject's individual rationality as taking the collectively rational into account,

[28] I focus here on MCR models, however, other kinds of modeling might also be useful for discovering extrinsic epistemic reasons, e.g., Agent-Based epistemic terrain modeling (Muldoon 2013; Muldoon & Weisberg 2011; Thoma 2015; M. Weisberg & Muldoon 2009; Zollman 2009). The endorsement framework could easily implement constraints derived from such modeling approaches.

and we can do this by appealing to the same conceptual resources as already used by the MCR account.

In short, what I will show is that we can build the reward structure recommended by the MCR account into the epistemic utility function of the researcher. This is possible because endorsement is appropriately sensitive to extrinsic epistemic values.

In pursuit of this project, I will briefly describe Strevens' version of MCR. The decision-theory for endorsement incorporating MCR will be explored in section 6.

Suppose there are two scientific research projects, P_1 and P_2. There is a "success" function for each project, describing how likely the project is to produce truth given a number of people working on it. That is, it takes a number of researchers (n) as input and outputs the probability that the project will be successful.[29] Other things being equal (it seems plausible to assume), projects that are more likely to bear truth will have bigger values for any particular number of researchers n. Call the success functions for P_1 and P_2, $s_1(\cdot)$ and $s_2(\cdot)$, respectively. Suppose also that P_1 is the more likely project to pay off, and so $s_1(n) > s_2(n)$. As mentioned above, there will be an optimal distribution of researchers based on these success functions: where P_1 is more likely to get us the truth, and N is the total number of researchers (or research hours), what is to be optimized is:

$$s_1(n) + s_2(N - n)$$

This will give us the best chance of getting at the truth.

The problem, as discussed above, is that on their own, if each researcher is trying to give themselves the best opportunity of personally getting to the truth, they will all choose to work on the project with the better success function: the one more likely to get at the truth. In this case, this will mean everyone is working on P_1, and this will not (generally) be the optimal distribution. Inquiry as a whole is not best served by having all researchers working on the same project.

Strevens shows that what we need in order to reach the optimal distribution is for each researcher to maximize their own *marginal contribution*. Here, "marginal contribution" means the increase to the probability the project will pay off that is provided by the researchers joining the

[29] This can also be done in terms of research hours, rather than individuals. Number of researchers is more natural for my account, but either will work.

project. That is to say, my marginal contribution is how much more likely the project is to pay off after I work on it. It is the difference I make to the probability of success. We can represent marginal contribution of P_1 as a function $m_1(\cdot)$, defined as:

$$m_1(n) = s_1(n+1) - s_1(n)$$

Strevens shows that rewarding the researcher based on their marginal contribution, a reward scheme he calls *Marge*, will lead to a better distribution of labor. This claim is based on a few plausible assumptions. In particular, it requires that the success functions are increasing, but have decreasing marginal gains. That is, every researcher added to the project increases its probability of paying off. However, each additional researcher adds less probability than the last one, i.e., the project becomes saturated. This seems like a perfectly reasonable assumption: the more researchers already working on something, the less good it will do to add another one.[30]

So in our example, when more researchers are already working on P_1, the marginal contribution for a new worker joining it will be smaller. Meanwhile, the marginal contribution for joining P_2 will be higher. That is, as n, the number of researchers working on P_1 increases, $m_1(n)$ decreases and $m_2(n)$ increases. This can result in its being much more lucrative to work on P_2, even though the overall probability of its paying off might be significantly lower. Thus, implementing the *Marge* reward scheme will lead to better distributions of labor, since researchers will be incentivized to work on projects with less probability of paying off, but to which their own work will contribute much more.[31]

One benefit of Strevens' theory is that it vindicates aspects of scientific practice. Given human nature, it seems good that we can implement practical rewards that make individual practical rationality line up with the collective good (a notion familiar from economic theory). However, prudential motivations can lead to distorting influences, as researchers have monetary and career incentives to *seem* to contribute to inquiry without actually doing so (fraud, bribery, plagiarism, etc.).[32] This will result in

[30]These assumptions can be relaxed to obtain very similar results (Kitcher 1990).

[31]Kitcher and Strevens appeal to an MCR scheme in order to vindicate the priority rule in science. I will leave aside the differences between the *Priority* reward scheme and *Marge*, as the details do not affect the project here.

[32]Though see Bright (2016) for how "pure" alethic goals can lead to fraud, too.

some failures of Strevens' idealizing assumptions. So we might need some additional work to ensure collective rationality, beyond simply appealing to practical considerations.

More importantly, however, researchers motivated by practical concerns should not be the only ones who can rationally act so as to benefit inquiry as a whole. Being motivated by personal gain should not be a requirement of *epistemic* rationality! Instead, we should borrow the insights of the MCR tradition, in order to make individual *epistemic* rationality consonant with collective epistemic rationality. We should be able to characterize the purely epistemic norms in such a way that someone with only epistemic motivations can, also, *rationally* contribute to a healthy distribution of cognitive labor. Thanks to inclusive epistemic rationality, with its sensitivity to extrinsic reasons, we are in a position to do so.[33] This is because facts about what will lead to a better distribution of labor are extrinsic reasons which affect what it is rational to endorse. In order to reflect this, below I offer a decision-theoretic account of rational endorsement which builds a sensitivity to marginal contribution into the epistemic utility function for researchers. This provides a formal account of individual epistemic rationality which coheres with collective epistemic rationality.

6 Rational Endorsement

Rational endorsement is governed by inclusive epistemic rationality. The picture I have been developing comes to this: the rational theory to endorse is the one supported by the weight of both intrinsic and extrinsic epistemic reasons. How should we weigh these reasons when making a decision about what to endorse? One plausible answer is to turn to our standard theory of weighing considerations for action: decision theory.

In this section, I will give a formal characterization of rational endorsement, using the tools of decision theory. This account is meant to reflect the norms and features of endorsement that have already been character-

[33]It is worth noting that Kitcher briefly mentioned a solution somewhat similar to mine, but dismisses it as "redefinition" (1990, 14). I think this is a mistake. The project is not merely to stipulate that individual rationality is sensitive to concerns of collective rationality, but to show the explanatory payoff of a theory which ties them together in a coherent and rigorous manner.

ized. The basic idea is that one should endorse a theory just when doing so will maximize expected inclusive epistemic utility.

For clarity and ease of exposition, I am going to first provide an explanation of how the decision theory works (§6.1). I will then give the formal characterization in a separate subsection (§6.2), for readers who are interested in the technical details.

6.1 Rational Choice of Endorsement

The decision theory for endorsement is an epistemic one, because it involves an epistemic utility function. However, the epistemic utility here is distinct from the kind currently most popular in the literature. The contemporary notion of epistemic utility is "accuracy-first." On this view, accuracy is the sole determinant of how epistemically valuable an outcome is. Sets of beliefs, or credences, have a certain value based on how close they are to the truth. This kind of epistemic utility theory was pioneered by Joyce (1998), and has been fruitfully applied in a large, rapidly expanding literature.[34]

Accuracy-focused utility theory is perfectly appropriate for determining norms of belief and credence, because such attitudes are appropriately sensitive only to intrinsic epistemic reasons. However, since rational endorsement is governed by inclusive epistemic rationality, we need a more inclusive notion of epistemic utility. Helpfully, there is an older tradition of epistemic utility that we can appeal to, owing especially to the work of Levi (1974, 1980, 2004), Kaplan (1981a) and Maher (1993). The theory I am giving here is largely inspired by Maher's.

Using Maher-style epistemic utility, we begin with a standard subjective utility function for an agent, one that reflects the agent's values and desires. We then add constraints on the utility function which ensure that it is appropriately epistemic. These constraints require that the subject values truths over falsehoods, values the avoidance of contradictions, and values interesting truths over prosaic ones. This was Maher's strategy for ensuring that the subject would have "scientific values." What I want to add to this strategy is additional constraints which will ensure that the

[34]For more on this kind of epistemic utility, see Easwaran (2013, 2016); Fitelson (2016); Fitelson and Easwaran (2015); Greaves and Wallace (2006); Joyce (1998); Konek and Levinstein (2016); Pettigrew (2016).

subject is appropriately sensitive to extrinsic epistemic values. The new constraints that I add below are designed to implement the solutions borrowed from the social epistemology literature reviewed above, and in general to make the agent sensitive to both intrinsic and extrinsic considerations.

The version of the decision theory I offer below employs only a small number of constraints. It includes three of Maher's constraints. The first of these is a weak truth constraint: that the subject prefers to endorse a theory a when a is true. This constraint is meant to help ensure that researchers only endorse theories which are live options: theories that have some probability of turning out to be true, even if they are less likely than their negation, or than their competitors. This constraint is weak, however. It is compatible with a greater utility for endorsing b over a, even if b is false and a true. This weakness in the constraint allows the decision theory to model the fact that it can be rational to endorse something even when it is unlikely. This fits with endorsement's aim at promoting healthy inquiry.

The second constraint encodes a preference for information: a preference for more informative theories over less informative ones, as long as the more informative theory is not contradictory (since contradictions are guaranteed to count as highly informative, since they entail everything). The third constraint adds a general prohibition on contradiction.[35] These constraints, along with the standard norms of rational credence, are meant to capture the sensitivity to intrinsic epistemic reasons.

In addition to the constraints from Maher, there are two constraints meant to enforce sensitivity to extrinsic epistemic considerations. The first is taken from Strevens' *Marge* reward scheme, which I will simply call the *MCR constraint*. It ensures that the subject prefers to have a higher marginal contribution to their project. That is, they gain more utility from

[35] I think this constraint will be operative in most domains. However, there are a number of specific research domains where this might need to be relaxed. For instance, we might need to relax it in characterizing the early stages of the pursuit of quantum mechanics, where the initial theory was known to be inconsistent (Faye 2014). More obviously, research about the applicability of paraconsistent logic, and dialetheism will violate this constraint (Priest 2006; Priest & Berto 2013; Priest, Tanaka, & Weber 2015). Even if all inconsistent theories are in fact false (because inconsistent), we can still model them using inclusively rational endorsement, as long as we relax this constraint. Thanks to Eddy Chen, Graham Priest, and Branden Fitelson for discussion on these points.

working on a project to which their own work will contribute more. This constraint appeals to the notion of a marginal contribution, and to a success function, described above (§5). To satisfy this constraint, the subject's utility will be higher when their marginal contribution is higher.

This constraint thus builds Strevens' solution to the distribution of cognitive labor directly into the account of individual rationality. What it is rational to endorse depends on one's marginal contribution. This is justified because the epistemically-motivated rational inquirer values being part of healthy inquiry, and given Strevens' arguments, striving to contribute the most will lead to better distribution of labor. Endorsement must be sensitive to these considerations, since it is associated with research planning and pursuit judgments.[36]

The second new constraint is meant to reflect the resilient commitment of endorsement. This constraint, which I call *Inertia*, requires that the subject's utility for endorsing a proposition significantly increase after they endorse it. Then, in any subsequent decisions about what to endorse, the previously endorsed proposition's utility will outweigh even significant drops in the probability of that proposition. Thus, the subject's endorsement will survive even very difficult evidence against the theory. So the resilient commitment which characterizes endorsement will have the appropriate rational standing.

I suspect these constraints will prove adequate to capture all of the reasons, both intrinsic and extrinsic, relevant to endorsement.[37] This is because the MCR constraint makes the subject sensitive to a variety of other extrinsic epistemic reasons. Anything that makes the researcher more productive or effective will alter the success function, and so alter the researcher's marginal contribution. This means that extrinsic reasons will impact the success function of theories, as well as the degree of marginal contribution of that researcher in particular. So, sensitivity to marginal contribution thereby involves sensitivity to a variety of extrinsic reasons which impact the success function of the theory.

[36] As I suggest above, Strevens' model might not turn out to be the best one. If so, we can simply adopt the better model and give it the same treatment. Again, the purpose here is not to come down in favor of one solution in that domain, but to show how we can use such solutions in a framework for rational endorsement that smooths the tension between individual and collective epistemic rationality.

[37] Assuming, as above, that Strevens' MCR is the right view of how to ensure appropriate distribution of labor.

For example, extrinsic reasons such as the local availability of resources, or the particular talents of the researcher in question, will affect the researcher's own marginal contribution. Whether a researcher has access to an fMRI machine might genuinely impact how valuable for inquiry it would be for that researcher to pursue a project involving a theory that is best tested by fMRI experiments. And a researcher sensitive to their own marginal contribution will avoid endorsing a theory, and planning a research program, if they do not have access to the right equipment.

Similarly, a researcher's available network of contacts with other researchers, in her own field and relevant other fields, will affect how much marginal contribution she makes to the project of confirming her endorsement. Even just how excited a researcher is about a particular theory, or the kind of work necessary to test it, can make a relevant impact. People are more motivated to work on things they like, so their own affinities are legitimate extrinsic epistemic factors.

Thus, although these constraints were designed to allow the account of endorsement to help with distribution of cognitive labor, it also will help with ensuring sensitivity to other kinds of extrinsic epistemic reasons.

Though I prefer the epistemic utility function utilizing just these constraints, the framework developed is actually more flexible than this. It can accommodate treating other considerations as independent constraints on the utility function.[38]

The purpose of the decision theory for endorsement is not so that scientists can sit down and calculate what it is most rational to do. Rather, it serves as a set of evaluative constraints on researchers' preferences between different choices of endorsements. I think that one of its more valuable applications is to allow us to evaluate the actual decision-making procedures that researchers use in theory choice. This is much the way that the heuristics and biases literature in social psychology uses standard decision theory as its theory of rationality, and then uses this to evaluate actual human decision-making, and determine where such decision-making falls off the rails.[39]

[38]For details on how this would work, see footnote 42.

[39]For an overview this literature, see Kahneman (2013). The literature began with Tversky and Kahneman (1975). For a philosophical application of this idea, see J. Weisberg (2016).

6.2 Decision Theory for Endorsement

Here, I provide the formal details of the epistemic decision theory. The underlying formal framework for this is borrowed from Joyce (1999). For ease of exposition, this theory is evidential.[40]

In the decision theory for endorsement, each decision problem, $\mathbf{D} = (\Omega, \mathbf{A}, \mathbf{S}, \mathbf{O})$, is composed of these elements:

1. A partition of acts of endorsement, $\mathbf{A} = \{A_a, A_b, A_c, ...\}$, where A_a is the act of endorsing the proposition a.
2. A partition of states of the world, $\mathbf{S} = \{S_1, S_2, ...S_n\}$.
3. A partition of outcomes, $\mathbf{O} = \{A_i \& S_i | \forall A_i \in \mathbf{A} \textit{ and } \forall S_i \in \mathbf{S}\}$.
4. A σ-algebra Ω such that it is the smallest such algebra containing $\mathbf{A}$, $\mathbf{S}$, and $\mathbf{O}$ and their closure under negation and disjunction.
5. A credence function $P(\cdot)$ defined on Ω.
6. A utility function defined over $\mathbf{O}$ so that the utility of endorsing a in state S_1 is $u(A_a \& S_1)$.

The decision rule is then the standard requirement to choose the act that maximizes expected value:

$$EV(A_a) = \sum_{S_i \in \mathbf{S}} P(S_i | A_a) u(A_a \& S_i)$$

The act-space here consists of acts of forming endorsements, indexed by theories or propositions that count as eligible options for endorsement. This will be domain-specific: in each research domain, there will be considerations appropriate to that domain that researchers will need to be sensitive to in deciding what the available options are.

My theory introduces no new difficulties with the probability function and state-space, and we can treat these in the standard way. The state-space can be the set of possible worlds (R. Stalnaker 1999; R. C. Stalnaker 1987), or some other partition that appropriately accounts for the researchers uncertainties (Jeffrey 1990; Joyce 1999).[41]

[40] Expanding this to a causal decision theory is a relatively simple matter, but it adds some complications to the formalism which are irrelevant to our purposes here. For the procedure for the expansion to CDT, see Joyce (1999).

[41] The current version of the theory, since it is an evidential decision theory, makes this easier in some respects because it is partition invariant.

What sets this theory apart from a standard evidential decision theory is the constraints on the utility function. I will call them "M-constraints" after Maher (1993). They ensure that the theory will be appropriately sensitive to epistemic concerns, both intrinsic and extrinsic.

The set of M-Constraints:

1. **Respect for Truth**: $u(A_a \& S_i) \geq u(A_a \& S_j)$ for all $S_i \in a$ and all $S_j \notin a$ (i.e., higher utility whenever the state is one in which a is true).

2. **Respect for Information**: $u(A_{a\&b} \& S_i) \geq u(A_a \& S_i)$, for all $a, b \in \Omega$ s.t. $a \& b \nvDash \varnothing$. (i.e., prefer more specific, informative things, as long as they are not contradictory).

3. **Contradiction Suboptimal**: $u(A_a \& S_i) \geq u(A_\varnothing \& S_i)$ for any S_i and any a.

4. **MCR**: the utility of a state where the subject's marginal contribution is higher is at least as great as one where it is lower. Moreover, the difference in utility between two states should increase proportionately with differences between the marginal contribution of the agent in those states. Expressed formally:

   (a) When $S_i, S_j \in a$, and $m_a(A_a) = I$ in S_i, and $m_a(A_a) = J$ in S_j, and $I > J$, then $u(A_a \& S_i) > u(A_a \& S_j)$ and

   $$\frac{u(A_a \& S_i)}{u(A_a \& S_j)} \propto \frac{I}{J}$$

   (b) When $S_i \notin a$ and $S_i \notin b$, if $m_a(A_a) > m_b(A_b)$, then $u(A_a \& S_i) > u(A_b \& S_i)$ and

   $$\frac{u(A_a \& S_i)}{u(A_b \& S_i)} \propto \frac{m_a(A_a)}{m_b(A_b)}$$

5. **Inertia**: The utility of endorsing a theory becomes much higher once one endorses it:

   Where u_1 is utility function at time t_1: If the subject comes to endorse a between t_1 and t_2, then $u_2(A_a \& S_i) \gg u_1(A_a \& S_i)$, for all S_i.

With these constraints, we have an inclusive epistemic utility function. This provides a helpful characterization of rational endorsement.[42]

7 Conclusion

A field of inquiry in which researchers endorse their theories will be a healthier, more productive field. Recognizing endorsement, along with its associated normative framework, allows us to complete both the vindication and justification projects introduced in section 3. We can justify the actions of excellent researchers like Ellie, who are resiliently committed to their own theories, and who advocate strongly for them, even though they should not believe them. It also provides normative guidance for researchers who want to contribute to healthy, productive fields of inquiry.

The theory of rational endorsement provided here also shows how endorsement can help smooth the tension between individual and collective rationality. It provides an avenue for doing so that would not be available to a theory relying instead on belief. It achieves this by appeal to inclusive epistemic rationality, which is sensitive to both extrinsic and intrinsic epistemic considerations. This sensitivity is modeled by an epistemic decision theory which uses a set of constraints on the utility function.[43]

[42] Although I prefer a theory which uses just the few constraints listed above, the framework is actually more flexible than this. There is a simple way to expand the decision theory to take into account additional extrinsic reasons more directly. We can do this by following Levi (1974, 1980) and Pettigrew (2014), and using a composite utility function. This function is composed of the weighted average of several sub-utility functions, each of which represents sensitivity to a different extrinsic reason.

Suppose $u_{mcr}(\cdot)$ is the sub-utility function structured as in the above. Also, $u_{heu}(\cdot)$ is the sub-function structured by heuristic power, where $u_{heu}(A_a \& S_i) > u_{heu}(A_b \& S_i)$ just when the heuristic of a is better (or more powerful) than b's. Furthermore, let $u_{exp}(\cdot)$ be the sub function structured by the novelty of explanations, where $u_{exp}(A_a \& S_i) > u_{exp}(A_b \& S_i)$ just when a provides more novel explanations of phenomena than b.

To obtain the overall epistemic utility function, $U(\cdot)$, we weight these individual sub-functions, then add them together. Let $\alpha_1 + \alpha_2 + \alpha_3 = 1$, where the size of each α is determined by how important the subject takes the different considerations to be. Then,

$$U(A_a \& S_i) = \alpha_1 u_{mcr}(A_a \& S_i) + \alpha_2 u_{heu}(A_a \& S_i) + \alpha_3 u_{exp}(A_a \& S_i)$$

Finally, this composite utility can be plugged into the expected utility calculation, as in §6.2.

[43] **Acknowledgments:** Thanks to Sara Aronowitz, Bob Beddor, David Black, Matt Dun-

References

Alston, W. M. (1996, January). Belief, Acceptance, and Religious Faith. In J. Jordan & D. Howard-Snyder (Eds.), *Faith, Freedom, and Rationality: Philosophy of Religion Today* (pp. 3–27). Rowman & Littlefield.

Alvarez, M. (2016). Reasons for Action: Justification, Motivation, Explanation. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Summer 2016 ed.).

Arpaly, N. (2017). Epistemology and the Baffled Action Theorist. , *manuscript*.

Berker, S. (2013). Epistemic teleology and the separateness of propositions. *Philosophical Review*, *122*(3), 337–393.

Bonner, B. L., Baumann, M. R., & Dalal, R. S. (2002). The effects of member expertise on group decision-making and performance. *Organizational Behavior and Human Decision Processes*, *88*(2), 719–736.

Bright, L. K. (2016). On Fraud. *Philosophical Studies*, *forthcoming*, 1–20.

Broome, J. (2015, October). Reason versus ought. *Philosophical Issues*, *25*(1), 80–97. doi: 10.1111/phis.12058

Cohen, L. J. (1989a, July). Belief and Acceptance. *Mind*, *98*(391), 367–389.

Cohen, L. J. (1989b, January). What Use Are Beliefs That We Do Not Take to Be Warranted? *Analysis*, *49*(1), 7. doi: 10.2307/3328887

Cohen, L. J. (1995). *An Essay on Belief and Acceptance* (Reprint edition ed.). Oxford: Oxford University Press.

De Cruz, H., & De Smedt, J. (2013). The Value of Epistemic Disagreement in Scientific Practice. The Case of Homo Floresiensis. *Studies in History and Philosophy of Science Part A*, *44*(2), 169–177.

Easwaran, K. (2013). Expected Accuracy Supports Conditionalization—and Conglomerability and Reflection. *Philosophy of Science*, *80*(1), 119–142.

Easwaran, K. (2016). Dr. Truthlove Or: How I Learned to Stop Worrying and Love Bayesian Probabilities. *Noûs*, *50*(4), 816–853.

Egan, A. (2008). Seeing and believing: Perception, belief formation and the divided mind. *Philosophical Studies*, *140*(1), 47–63.

Elga, A., & Rayo, A. (2015). *Fragmentation and Information Access.*

Faye, J. (2014). Copenhagen Interpretation of Quantum Mechanics. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2014 ed.).

can, Andy Egan, Adam Elga, Branden Fitelson, Georgi Gardiner, Alvin Goldman, Daniel Rubio, Joshua Schecter, Susanna Schellenberg, Ernest Sosa, and an anonymous referee. Thanks also to audiences at The Rutgers-Penn-Princeton Social Epistemology Workshop, the Ninth Workshop in Decision, Games, and Logic, and the Vancouver Summer Philosophy Conference. Special thanks to Megan Feeney.

Finlay, S., & Schroeder, M. (2015). Reasons for Action: Internal vs. External. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2015 ed.).

Firth, R. (1981). "Epistemic Merit, Intrinsic and Instrumental&Quot. *Proceedings and Addresses of the American Philosophical Association*, *55*(1), 5–23.

Fitelson, B. (2016). *Coherence*. manuscript.

Fitelson, B., & Easwaran, K. (2015). Accuracy, Coherence, and Evidence. In T. S. Gendler & J. Hawthorne (Eds.), *Oxford Studies in Epistemology Volume 5*. Oxford: Oxford Studies in Epistemology.

FitzPatrick, W. J. (2004, January). Reasons, Value, and Particular Agents: Normative Relevance without Motivational Internalism. *Mind*, *113*(450), 285–318. doi: 10.1093/mind/113.450.285

Frankish, K. (2004). *Mind and Supermind*. Cambridge: Cambridge University Press.

Frigg, R., & Nguyen, J. (2016). Scientific Representation. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2016 ed.).

Geil, D. M. M. (1998, July). Collaborative Reasoning: Evidence for Collective Rationality. *Thinking & Reasoning*, *4*(3), 231–248. doi: 10.1080/135467898394148

Gert, J. (2009). Desires, Reasons, and Rationality. *American Philosophical Quarterly*, *46*(4), 319–332.

Goldberg, S. (2013a). Defending Philosophy in the Face of Systematic disagreement. In D. Machuca (Ed.), *Disagreement and Skepticism*. (pp. 277–94). New York: Routledge.

Goldberg, S. (2013b, May). Inclusiveness in the face of anticipated disagreement. *Synthese*, *190*(7), 1189–1207. doi: 10.1007/s11229-012-0102-2

Goldman, A. I. (1986). *Epistemology and Cognition*. Harvard University Press.

Gordon, J. (1997). John Stuart Mill and the "Marketplace of Ideas". *Social Theory and Practice*, *23*(2), 235–249.

Greaves, H. (2013, October). Epistemic Decision Theory. *Mind*, *122*(488), 915–952. doi: 10.1093/mind/fzt090

Greaves, H., & Wallace, D. (2006). Justifying Conditionalization: Conditionalization Maximizes Expected Epistemic Utility. *Mind*, *115*(459), 607–632.

Jeffrey, R. C. (1990). *The logic of decision*. Chicago: University of Chicago Press.

Jenkins, C. (2007). Entitlement and Rationality. *Synthese*, *157*(1), 25–45.

Joyce, J. M. (1998). A Nonpragmatic Vindication of Probabilism. *Philosophy of Science*, *65*(4), 575–603.

Joyce, J. M. (1999). *The Foundations of Causal Decision Theory*. Cambridge University Press.

Kahneman, D. (2013). *Thinking, Fast and Slow* (1st edition ed.). New York: Farrar, Straus and Giroux.

Kaplan, M. (1981a). A Bayesian Theory of Rational Acceptance. *The Journal of Philosophy*, *78*(6), 305–330. doi: 10.2307/2026127

Kaplan, M. (1981b). Rational Acceptance. *Philosophical Studies*, *40*(2), 129–145.

Kaplan, M. (1995, January). Believing the Improbable. *Philosophical Studies*, *77*(1), 117–146.

Kerr, N. L., MacCoun, R. J., & Kramer, G. P. (1996). Bias in judgment: Comparing individuals and groups. *Psychological review*, *103*(4), 687.

Kerr, N. L., & Tindale, R. S. (2004). Group performance and decision making. *Annu. Rev. Psychol.*, *55*, 623–655.

Kitcher, P. (1990, January). The Division of Cognitive Labor. *The Journal of Philosophy*, *87*(1), 5–22. doi: 10.2307/2026796

Konek, J., & Levinstein, B. (2016). *The Foundations of Epistemic Decision Theory.* manuscript.

Lacey, H. (2015, October). 'Holding' and 'endorsing' claims in the course of scientific activities. *Studies in History and Philosophy of Science Part A*, *53*, 89–95. doi: 10.1016/j.shpsa.2015.05.009

Laudan, L. (1978). *Progress and its problems: Towards a theory of scientific growth* (Vol. 282). Univ of California Press.

Laudan, R. (1987). The rationality of entertainment and pursuit. In *Rational Changes in Science* (pp. 203–220). Springer.

Laughlin, P. R., Bonner, B. L., & Miner, A. G. (2002). Groups perform better than the best individuals on letters-to-numbers problems. *Organizational Behavior and Human Decision Processes*, *88*(2), 605–620.

Laughlin, P. R., & Ellis, A. L. (1986). Demonstrability and social combination processes on mathematical intellective tasks. *Journal of Experimental Social Psychology*, *22*(3), 177–189.

Leitgeb, H. (2014). The stability theory of belief. *Philosophical Review*, *123*(2), 131–171.

Levi, I. (1974). *Gambling with Truth: An Essay on Induction and the Aims of Science*. Cambridge: The MIT Press.

Levi, I. (1980). *The enterprise of knowledge: An essay on knowledge, credal probability, and chance*. Cambridge, Mass: MIT Press.

Levi, I. (2004). *Mild contraction: Evaluating loss of information due to loss of belief*. Oxford; Oxford; New York: Clarendon ; Oxford University Press.

Lewis, D. (1982). Logic for Equivocators. *Noûs*, *16*(3), 431–441.

Maher, P. (1993). *Betting on theories*. Cambridge; New York, NY, USA: Cambridge University Press.

McKaughan, D. (2007). *Toward a Richer Vocabulary for Epistemic Attitudes* (Unpublished doctoral dissertation). University of Notre Dame.

McKaughan, D. (2008). From Ugly Duckling to Swan: C. S. Peirce, Abduction,

and the Pursuit of Scientific Theories. *Transactions of the Charles S. Peirce Society*, *44*(3), 446–468.

McMullin, E. (1976). The Fertility of Theory and the Unit for Appraisal in Science. In R. S. Cohen, P. K. Feyerabend, & M. Wartofsky (Eds.), *Essays in Memory of Imre Lakatos* (pp. 395–432). Reidel.

Mercier, H. (2016, September). The Argumentative Theory: Predictions and Empirical Evidence. *Trends in Cognitive Sciences*, *20*(9), 689–700. doi: 10.1016/j.tics.2016.07.001

Mercier, H., & Sperber, D. (2011, April). Argumentation: Its adaptiveness and efficacy. *Behavioral and Brain Sciences*, *34*(2), 94–111. doi: 10.1017/S0140525X10003031

Muldoon, R. (2013). Diversity and the Division of Cognitive Labor. *Philosophy Compass*, *8*(2), 117–125.

Muldoon, R., & Weisberg, M. (2011). Robustness and Idealization in Models of Cognitive Labor. *Synthese*, *183*(2), 161–174.

Nickles, T. (1981). What is a Problem That We May Solve It? *Synthese*, *47*(1), 85–118.

Pagin, P. (2016). Assertion. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2016 ed.).

Parfit, D., & Broome, J. (1997). Reasons and Motivation. *Proceedings of the Aristotelian Society, Supplementary Volumes*, *71*, 99–146.

Pettigrew, R. (2014, August). *M-Phi: L. A. Paul on transformative experience and decision theory I.*

Pettigrew, R. (2016). *Accuracy and the Laws of Credence*. Oxford University Press Uk.

Priest, G. (2006). *In Contradiction: A Study of the Transconsistent*. Oxford University Press.

Priest, G., & Berto, F. (2013). Dialetheism. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Summer 2013 ed.).

Priest, G., Tanaka, K., & Weber, Z. (2015). Paraconsistent Logic. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2015 ed.).

Proust, J. (2013). *The Philosophy of Metacognition: Mental Agency and Self-Awareness*. OUP Oxford.

Rayo, A. (2013). *The Construction of Logical Space*. Oxford University Press.

Resnick, L. B., Salmon, M., Zeitz, C. M., Wathen, S. H., & Holowchak, M. (1993). Reasoning in conversation. *Cognition and instruction*, *11*(3-4), 347–364.

Rinard, S. (2015). No Exception for Belief. *Philosophy and Phenomenological Research*, *91*(2).

Šešelja, D., Kosolosky, L., & Straßer, C. (2012). The Rationality of Scientific Reasoning in the Context of Pursuit: Drawing Appropriate Distinctions.

Philosophica, *86*, 51–82.

Šešelja, D., & Straßer, C. (2013). Kuhn and the Question of Pursuit Worthiness. *Topoi*, *32*(1), 9–19.

Šešelja, D., & Straßer, C. (2014). Epistemic Justification in the Context of Pursuit: A Coherentist Approach. *Synthese*, *191*(13), 3111–3141.

Shah, N., & Velleman, J. D. (2005). Doxastic Deliberation. *The Philosophical Review*, *114*(4), 497–534.

Smith, M. (2005). Meta-Ethics. In F. Jackson & M. Smith (Eds.), *The Oxford Handbook of Contemporary Philosophy* (pp. 3–30). Oxford University Press.

Sosa, E. (2009). *Reflective knowledge: Apt belief and reflective knowledge* (Vol. 2). Oxford University Press.

Sosa, E. (2015). *Judgment and Agency*. Oxford University Press.

Stalnaker, R. (1999). *Context and content: Essays on intentionality in speech and thought*. Oxford ; New York: Oxford University Press.

Stalnaker, R. C. (1987). *Inquiry*. Cambridge, Mass.: MIT Press.

Steel, D. (2010). Epistemic Values and the Argument From Inductive Risk. *Philosophy of Science*, *77*(1), 14–34.

Strevens, M. (2003, February). The Role of the Priority Rule in Science. *The Journal of Philosophy*, *100*(2), 55–79.

Thoma, J. (2015). The Epistemic Division of Labor Revisited. *Philosophy of Science*, *82*(3), 454–472.

Tversky, A., & Kahneman, D. (1975). Judgment under uncertainty: Heuristics and biases. In *Utility, probability, and human decision making* (pp. 141–162). Springer.

Van Fraassen, B. C. (1980). *The scientific image*. Oxford; New York: Clarendon Press ; Oxford University Press.

Weiner, M. (2007, March). Norms of Assertion. *Philosophy Compass*, *2*(2), 187–195. doi: 10.1111/j.1747-9991.2007.00065.x

Weisberg, J. (2016). Belief in Psyontology. *The Philosopher's Imprint*, *Forthcoming*.

Weisberg, M., & Muldoon, R. (2009). Epistemic Landscapes and the Division of Cognitive Labor. *Philosophy of Science*, *76*(2), 225–252.

Whitt, L. A. (1985). *The Promise and Pursuit of Scientific Theories* (Unpublished doctoral dissertation). Dissertation.

Whitt, L. A. (1990). Theory pursuit: Between discovery and acceptance. *PSA: Proceedings of the biennial meeting of the philosophy of science association*, *1*, 467–483.

Whitt, L. A. (1992). Indices of Theory Promise. *Philosophy of Science*, *59*(4), 612–634.

Williamson, T. (1996, October). Knowing and Asserting. *The Philosophical Review*, *105*(4), 489–523. doi: 10.2307/2998423

Williamson, T. (2000). *Knowledge and its Limits*. Oxford University Press.

Winther, R. G. (2016). The Structure of Scientific Theories. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2016 ed.).

Yoshioka, S., & Kinoshita, S. (2007). Polarization-sensitive color mixing in the wing of the Madagascan sunset moth. *Optics Express*, *15*(5), 2691. doi: 10.1364/OE.15.002691

Zollman, K. J. S. (2009). Optimal Publishing Strategies. *Episteme*, *6*(2), 185–199.