

# FORGET AND FORGIVE: A PRACTICAL APPROACH TO FORGOTTEN EVIDENCE

SINAN DOGRAMACI  
*University of Texas at Austin*

We can make new progress on stalled debates in epistemology if we adopt a new practical approach, an approach concerned with the function served by epistemic evaluations. This paper illustrates how. I apply the practical approach to an important, unsolved problem: the problem of forgotten evidence. Section 1 describes the problem and why it is so challenging. Section 2 outlines and defends a general view about the function of epistemic evaluations. Section 3 then applies that view to solve the problem of forgotten evidence.

## 1. Introduction

### 1.1. *The Problem of Forgotten Evidence*

If someone forms a belief in an irrational way, say by purely wishful thinking, but then he completely forgets how he formed this belief, is the belief still irrational? For a concrete example, take Joe, who is a severe coffee addict, and who at one point formed the belief that drinking coffee is beneficial to your health. Let's suppose Joe formed this belief purely on the basis of his wish that this be true, and suppose he even engaged in this bit of wishful thinking *knowingly* at the time (he's helpless!), and he later completely forgot how he came to his view about coffee's health benefits. Is the surviving belief still irrational, or is it now rational? The case poses a problem, which we can present as a dilemma.

The **first horn of the dilemma** is just that there is certainly something uncomfortable about calling Joe's belief rational. If his belief becomes rational after he forgets how he formed it, then Joe has *laundered* the epistemic status of his belief—like the mafia launders money. Can you really do this? Can you turn an irrational belief into a rational belief just by forgetting things? To make the

---

**Contact:** Sinan Dogramaci <sinan.dogramaci@gmail.com>

problem even more vivid, think of the most appalling belief you can. Maybe someone, purely on the basis of his dislike for you, came to believe something horrible about you or your family, but the belief's original basis is now completely lost from his memory. Easily forgiven?

It feels wrong to forgive the forgetful reasoner. After all, it appears as though he never acquired, and certainly does not now retain, any positive reasons in favor of his view, and that looks bad. How can high confidence in propositions about topics like coffee's health benefits be justified without your ever having in your possession any positive reasons in favor of a particular view?

Some epistemologists who favor taking this first horn of the dilemma have ventured to suggest that you can possess positive reasons to trust a surviving belief even after forgetting its origins in virtue of a *background belief* you have that most of your beliefs have happily turned out, upon facing the test of time, to have been accurate.<sup>1</sup>

Unfortunately, the background belief suggestion is not appealing upon reflection. The problem is that it turns out that the background belief itself cannot be supported by any positive reasons, and thus anything that is based upon it cannot be said to *ultimately* enjoy the support of positive reasons. Why is it that the background belief has no positive reasons in its support? Well, if you were to possess some positive reasons that support this belief that your memory has been generally reliable, then those reasons would have to make some claim about the general accuracy of your *past* beliefs. But, what sort of reason have you to think your past beliefs were generally accurate? Any considerations you reach for here will make your reasoning *circular*.<sup>2</sup> So, even if Joe holds his belief that coffee is healthy on the basis of a background belief that his memory has generally been reliable, this will ultimately do very little to improve his ability to provide reasons in favor of his coffee belief, for the coffee belief is being based on another belief—the background belief—that itself enjoys no positive reasons in its own favor. To be clear, the background belief might well qualify as justified—perhaps it is justified by default, without the aid of any positive reasons in its favor. The point here is just that the worry about Joe not having any positive reasons to support his belief that coffee is healthy will not be satisfactorily addressed by claiming Joe should support that belief by basing it on a background belief that his memory has generally been reliable.

Turning to **the other horn of the dilemma** now, there is a real problem with saying such beliefs as Joe's are irrational. How can it be that a belief like his is irrational when Joe possesses no evidence, no reason, that favors making any *revision* to his view about coffee's health benefits? If his current belief is irrational, but he has no reasons to revise his belief, then he is in a rut: he must either hold

---

1. This suggestion is made by Christensen (1994: 74–6) and Conee and Feldman (2001: 9).

2. As Plantinga (1993: 62–3), Huemer (1999: 347) and Owens (2000: 313–4) have all noted.

on to an irrational belief, or make a revision he has no reason to make, a revision which would only land him with another irrational view. To make the problem here more vivid, just consider that even the best of us form a good number of irrational beliefs. And most of your beliefs, whether rational or irrational, outlive your memory of their original bases. You cannot be rationally required to keep track of much at all from the daily flood of evidence you take in and on the basis of which you first formed your beliefs; as Harman (1986: Chapter 4) famously argued, it would introduce a hugely wasteful amount of clutter into your mind if you stored all the evidence that would be needed to put every belief you now hold on any kind of firm foundation. You just keep the belief and forget the evidence. Joe now possesses no evidence opposing his belief that coffee is good for you, nor does he possess any evidence that his current view was unreliably formed. But, the same goes for lots of his beliefs, and lots of yours and mine, and it would be absurd to think that, just because we can't store limitless evidence, we have reasons to revise these views.

It seems you may not be able to turn these irrational views into rational ones, even if you acquire and take account of *new evidence* to form a new view. This is because you cannot just disregard the old view; new views, in particular new credences, are always based, in part, on your prior views, and the rational status of the old view threatens to infect the new view. For example, if your prior credence that a coin is biased is high, then even as you observe an even frequency of heads and tails come up, your posterior credence that it's biased will still remain a bit higher than it would have been if your prior had been lower. And if that prior credence was irrational, it seems the posterior credence will be too.<sup>3</sup> Joe's case seems on a par with the coin observer's. Suppose Joe acquires new evidence that coffee is in fact not healthy. Perhaps it is allowable that he could end up with a rational outright, or "binary", belief that coffee is not healthy (I'm not sure). However, his resultant credence in this will have been anchored down to some extent by his prior credence that coffee is healthy. Because of his irrationally high prior credence that coffee is healthy, it seems he will be left with a posterior credence that is still, to some extent, irrationally high. It seems such anchoring effects could only be fully undone if we could make Joe completely certain that his prior coffee views were formed unreliably. Rational certainty in anything is hard enough to attain. And it may be especially hard here, for, if we try to confront Joe with evidence that his view that coffee is healthy was based on wishful thinking, there *might* be anchoring that makes him rationally mistrust our evidence: he might reasonably be at least slightly suspicious that

---

3. Even in a strange case where, somehow, two different priors lead to the same posterior credence (as can happen when all old evidence is "screened off"; see Weatherston (2010) for discussion), even then it's plausible that, if the prior was irrational, then the posterior will be too. Posterior credences appear to always inherit the rational status of prior credences.

we are giving him misleading evidence, because he thinks coffee is healthy, and he knows a reliable thinker is likelier than a wishful thinker to have hit upon the truth!<sup>4</sup> I think most contemporary epistemologists will be inclined to agree that it will always be reasonable for Joe to remain at least *slightly* dubious that he engaged in wishful thinking. So, any epistemic predicament like the one Joe got himself into will leave the believer in a real rut, one that is at least very hard, and maybe even impossible, to completely get out of, and one that is very common in all of us.

The problem of forgotten evidence is thus a dilemma. Either Joe's belief is rational, in which case it has been laundered, or it is irrational, in which case he has landed in a rut. And, whatever we say, we will have to say it about a phenomenon that is very common.<sup>5</sup>

The dilemma divides epistemologists. Many say Joe's belief is irrational (or unjustified or unwarranted—I won't distinguish these), including Annis (1980: 326), Senor (1993: 468), Burge (1997: 39–40), Goldman (1999: 280–1), Goldman (2009: Section V), Huemer (1999: 348–9), Owens (2000: 157), and Greco (2005: 261). Others say the belief is now rational, including Harman (1986: Chapter 4), Pollock and Cruz (1999: 87), Conee and Feldman (2001: 9–10), McGrath (2007: 4), Lackey (2008: Sections A.3, A.6), Schoer (2008: 77–8), and Naylor (2015: 377). Interestingly, the division does not track externalist or internalist sympathies (note, for example, that Goldman and Huemer, arch-externalist and arch-internalist, are on the same team). Perhaps even more interestingly, many of these authors choose their side without relying on any argument. They simply express a strong intuition for their own verdict about Joe's rationality. Each of Annis, Burge, Goldman, Greco, Huemer, Lackey, McGrath, Naylor, and Schoer present their view as simply the intuitively clear verdict—they just have different intuitions!

I join the second group, the nice guys: I say Joe's belief is now rational. I will present a new line of argument for this conclusion, one that I hope will offer new hope for resolving a debate that looked to be hopelessly deadlocked.

### *1.2. Plan for My Argument*

My argument has an unusual character. I will be arguing for this conclusion, a conclusion that says certain beliefs are rational, not on the basis of initial premises that themselves directly say what is or is not rational, but rather on

---

4. I used an italicized '*might*' in the noted sentence because this last claim treads into hotly debated issues about undercutters, higher-order evidence, and the bracketing or screening off of first-order evidence. See, e.g., Christensen (2010), Weatherson (2010) and Lasonen-Aarnio (2014).

5. I'm treating Joe's wishful thinking as a concrete example serving as a placeholder for any case of irrational thinking where the subject has forgotten his or her evidence; our discussion applies quite generally.

the basis of certain considerations concerning the *practical function*, or *practical purpose*, that we serve by making epistemic evaluations. That is, the overall argument moves to a first-order epistemically normative evaluation for its conclusion, but it doesn't start from any first-order epistemically normative premises. It starts from some plausible considerations about why we make epistemic evaluations, ostensibly a separate subject matter. And, in fact, the resulting argument is neither deductively valid nor "inductively strong"! (It is not a deduction, an enumerative induction, an inference to the best explanation, or any sort of statistical inference.)

How will the argument be any good then? One step in the middle of the argument will make use of semantic descent. In a simple use of semantic descent, we first make a claim about some bit of language's truth-conditions and then we switch to *using* that very language to *make* a first-order claim that we showed is true. The move where such a switch happens is neither valid nor "inductively strong", but it is ordinarily unproblematic. A simple example: 'knows' is factive; therefore Alice knows no lies.

In this paper's overall argument, however, I do not begin with semantics, but rather with pragmatics. I'll begin by defending a view of the function of our epistemically evaluative language, and this view will then guide us to conclusions about the semantics of these words, in particular to a conclusion about their extensions. I'm going to argue that our use of epistemically evaluative language serves a certain function which I'll outline, and I'll argue that it is contrary to that function for us to criticize Joe's coffee belief as irrational. Rather, we best serve the function by not calling Joe's belief irrational; we best serve the function by forgiving Joe, and letting his coffee belief pass as rational. From this (pragmatic) claim that *applying* the language of irrationality to Joe's belief does not serve its function, I wish to infer the (semantic) conclusion that Joe's belief is not included in the extension of our word, 'irrational', or in the extension of our concept of irrationality expressed by our word. That is, I wish to infer that it is *incorrect*—a false application—to call Joe's belief irrational.

Why should this pragmatics-to-semantics inference be reliable? It is plausible that this inference is reliable because it is plausible that we adopted words and concepts of epistemic evaluation with extensions such that their *correct* application is the application that serves their function. We could have adopted a different concept from our actual one, one whose extension differs slightly so as to make the alternative verdicts about cases of forgotten evidence correct. We could have used words that express slightly different concepts, words that have slightly different meanings. But, since we adopted linguistic and conceptual tools that serve their function by being applied to forgive Joe, it is a much simpler and thus more plausible hypothesis that we adopted a concept, and we use a word, whose extension makes this function-serving application the correct

application. That is how my overall argument is supposed to work, and why it is supposed to be a convincing form of argument.

So, although solving the problem of forgotten evidence is certainly of sufficient intrinsic interest to write about, my hope is that this paper will also serve a second, ulterior aim. I hope to show a new way we might make progress on stalemated debates within first-order epistemology. I believe that reflecting on the function of epistemic evaluation can benefit epistemology in many ways. This paper illustrates how it sheds light on the problem of forgotten evidence.

Reflection on the function of philosophically important words and concepts has a few important precedents. Quine (1986: Chapter 1) and Craig (1990) made important proposals about the functions of attributions of truth and knowledge. (Parsons, 1990/2014, discusses Quine's wider use of the method.) The trend of reflecting on functions has caught on in recent years: see Reynolds (2002), Williams (2002: esp. Section 2.4), Johnston (2006; 2011), Fricker (2007; in press), Henderson (2009; 2011); Henderson and Horgan (in press), Divers (2010), Mercier and Sperber (2011), and Dogramaci (2012; 2015). A number of papers explicitly feature the form of argument of this paper: the premises describe the function of some philosophical word or concept, and the conclusion then uses that word or concept to make a claim about the property it expresses. One example: Boris Kment argues for a view of the function of modal discourse and thought, and then he takes this view to support the first-order metaphysical thesis that necessities and possibilities all reduce to counterfactual conditionals (Kment 2006b: 307; 2006a: Sections 1.1, 7). Williamson (2008: Chapter 5, Section 1) argues for that same metaphysical conclusion from the premise that evolution would only pressure us to develop modal cognitive capacities for thoughts about counterfactuals.

In what follows, Section 2 presents the general considerations we'll require on the function of epistemic evaluation, and Section 3 applies these considerations to the problem of forgotten evidence. As a kind of preview, here's a short and rough caricature of the argument: we criticize each others' irrational beliefs in order to promote the trustworthiness of testimony, but it can't now have any positive impact to criticize the coffee addict Joe who has lost all track of the period of time when he did something worth criticizing, so Joe gets a pass. The caricature you might easily find fault with. The real argument takes up the rest of the paper.

## **2. The Function of Normative Evaluation**

### *2.1. Having a Function Is Having a Teleological Explanation*

I claim that our epistemically evaluative practice has a function. By "our epistemically evaluative practice", I mean our actual practice of making linguistic as-

sertions that use ordinary words of epistemic evaluation, words like '[ir]rational', '[un]reasonable', '[un]justified', '[un]warranted', even the ordinary (non-technical) use of '[il]logical'. I'm targeting language that expresses a natural and intuitively significant category of evaluations: it's the *J* in the traditional *JTB* (justified true belief) theory of knowledge. I don't mean to generally include uses of 'knows' as part of our epistemically evaluative practice; however, certain uses are included, in particular those where a subject is said to not know something that's not obviously false.<sup>6</sup> And sometimes epistemic evaluations can be made with equal effect using no conventionally epistemic words, for example, saying 'Oh, come on!' in the right argumentative context. In fact, I think regular people most commonly make epistemic evaluations by calling a person *smart*, *clever*, *dumb*, or *stupid*, even though those words have many different meanings, and even though a philosopher would never use those words in a paper to mean rational or irrational (probably because they are more vague and ambiguous).

By claiming our epistemically evaluative practice has a function, I mean there exists a *teleological explanation* of why we engage in that practice. In other words, I mean this: there exists an explanation of why the practice exists that proceeds by showing how the practice efficiently promotes interests of ours, interests that it is independently plausible that we strive to promote.

My intention is to uncover the function of words and concepts we already employ, the ones we already use in ordinary epistemic evaluation. We could, of course, revise our practice and switch over to alternative concepts of epistemic praise and criticism that correctly apply to different extensions. We might want to make such a revision if we discover that we don't, upon reflection, endorse the interests promoted by our current practice's function. (There is likely a teleological explanation of why people use certain ostracizing slurs, but at least some of us don't reflectively endorse the apparently tribal interests that such usage serves.) Anyway, the function I'm going to propose for our epistemically evaluative practice serves an interest that we all would endorse promoting even upon reflection. So, we should be happy to go on employing our old language, expressing our old concepts, to assert whatever verdict about our coffee addict, Joe, turns out to serve the proposed function.

---

6. I think ordinary people recognize this as somehow problematic and criticizable, even though they don't explicitly think of this as failure to meet a justification condition on knowledge. The knowledge attributions that are excluded from my targeted category are just the ones where the attribution is clearly not due to the subject's having or lacking justification, but to the presence or absence of the truth condition, or belief condition, or anything else, such as reliability or some other anti-Gettier condition.

## 2.2. Pursuit of Truth through Reliable Belief-Forming Practices

To lead up to the main proposal I want to make and defend about the function of our epistemically evaluative practice (the offset claim labeled 'FUNCTION' at the end of this subsection), let me start off by examining and developing the following initial suggestion: *we make epistemic evaluations because it efficiently promotes our interest in having true beliefs on subject matters of practical importance.* I think this suggestion is ultimately correct, but the story is more complicated than it first seems.

The substantive claim contained in this suggestion is that we make epistemic evaluations because it *efficiently promotes* our having true beliefs on subject matters of practical importance. It's trivial that having these beliefs really is in our interest, because I use "practical importance" just to mean those subject matters that it's in our interest to have the true view on. True beliefs have obvious potential to serve our interests, at least if it's a remotely accurate description of us that we act in ways that would fulfill our desires were our beliefs true. (Many philosophers, functionalists, say this interconnection between belief, desire and action is a necessary truth about any believer/desirer/actor;<sup>7</sup> here we're only insisting on the milder claim that it's at least actually true of us.) Though we occasionally have desires that are either neutral or detrimental to our interests, by and large we desire what's in our interest, and consequently we hold beliefs on subject matters that, if those beliefs were true, would lead us to fulfill those desires, and thereby promote our own interests. By and large, then, the subject matters we care about are ones of practical importance. And then, automatically, the interest that I'm suggesting epistemic evaluations promote is something we will, by and large at least, endorse promoting, even upon reflection.

So, what can be said in support of the suggestion that our epistemically evaluative practice efficiently promotes our interest in having true beliefs on practically important subject matters? Well, from one broad perspective, it may seem obvious enough, for it may seem obvious that most linguistic communication has this function. Ordinary linguistic communication serves to transmit information from speakers to audiences, and by and large this benefits participants to the practice by efficiently spreading true beliefs, beliefs which then cause actions that promote whatever our desires are, and thereby, presumably, promote our interests (the presumption being that we desire what's in our interest). But, actually, I don't think the epistemically evaluative aspect of our overall linguistic practice serves its function in this straightforward way. Let me explain why.

True beliefs serve our interests when they are true beliefs about the means to the fulfillment of our desires and we desire what's in our interest. So, for a really easy example, a true belief about how to get food concerns a means to

---

7. For example, Lewis (1974: p.337) and Stalnaker (1984: p.15).



fulfilling a desire for an outcome, an outcome that is in our interest. But a true belief about how to be rational does not concern a means to an outcome that is in our interest. We don't have a direct interest in being rational. I know that claim—my last sentence—will initially strikes many people as wrong, but I think it sounds wrong only so long as we unreflectively associate being rational with being reliable. Being reliable certainly is in our interest, since being reliable just is having a tendency to form true beliefs. But, I do not see how being rational is at all in our interest *independently of our interest in being reliable*. It normally is in our interest to have rational beliefs, but not ultimately for any reason aside from our interest in reliability and, more fundamentally, truth. So, there is actually quite a puzzle: why do we form beliefs and why do we communicate information about how to be rational, if our interest is only in how to be reliable?

The answer is that the teleological explanation of why we make epistemic evaluations is a separate explanation from the more common and obvious teleological explanation of why we engage in most linguistic communication. Most, or at least much, ordinary communication on a subject matter is teleologically explained by our interest in true beliefs *on that subject matter*; a true belief on that subject matter helps us act in ways that fulfill our desires, and thus, normally, serve our interests. Epistemic evaluations, however, are not explained in this way. We do not make epistemic evaluations, we do not tell others what's rational and what's irrational, because having true beliefs *on that subject matter* fulfills our desires, and thereby serves our interests, in the way other beliefs normally do. The way epistemic evaluations promote our interest in true belief on matters of practical importance is more indirect.

I want to argue that we make epistemic evaluations because it influences the belief-forming practices of the community, influences them in a way that promotes the *safety* of the transmission of true beliefs via testimony, where "safe" characterizes a practice that does not easily lapse into spreading false beliefs. The idea is that epistemic evaluations play a supporting role: ordinary linguistic communication, ordinary testimony, has the primary function, that of transmitting true beliefs on subject matters of practical importance, while epistemic evaluation has the ancillary function, that of supporting the safety of the transmission of true belief by ordinary testimony. If ordinary testimony can easily lapse into spreading false belief, it is a dangerous tool for promoting our interests; it is unsafe. We benefit from keeping testimony safe. To do this, we need a way of guarding against false testifiers. Our practice of epistemically evaluating each other serves as one such guard. Let me explain how.

There are two ways to guard ourselves against false testimony. **The first way** is to flag the testimony as false or unreliable, and thereby discourage anyone's trusting it. But, if epistemic evaluations serve a valuable function, it cannot be, or rather it cannot *merely* be, that of flagging false or unreliable testimony, be-

cause this job can be done most effectively using evaluations of others as simply *unreliable*. Ordinary people often do this using plainer language than ‘reliable’; they might just say “he doesn’t know what he’s talking about” to communicate the point that anything he gets right is by luck. This is the basic idea of Edward Craig’s view of the function of knowledge attributions, which I am strongly inclined to accept.<sup>8</sup> Perhaps calling someone irrational is *a* way of communicating, in many ordinary contexts, that he is unreliable, but it is not the only, much less the best, way. So, this cannot be the primary function served by epistemic evaluations, because then they would do nothing that evaluations of others as (un)reliable, or as (non-)knowers, already do just as well or better. (I’ll return to this point shortly, in Section 2.3 below.)

It’s **the second way** of guarding against false testimony that I propose epistemic evaluations are uniquely suited to do. The second way is to influence the testifier to modify his or her belief-forming practice, so as to improve the reliability of his or her testimony. This is the proposal about the function of epistemic evaluations I want to defend and will be relying on in addressing the problem of forgotten evidence, so let’s put it front and center:

FUNCTION: We make epistemic evaluations because they help protect us from unreliable testimony. Our evaluations do this by influencing unreliable testifiers to modify their belief-forming practices, and thereby modify their testimony, so as to improve the reliability of testifiers’ reports, and thus also improve the reliability of audiences’ beliefs.

FUNCTION only claims that epistemic evaluations *help* keep testimony safe. FUNCTION does not suggest evaluations are the only or even the primary means by which beliefs and testimony are made as reliable as they actually are. What FUNCTION claims is that the primary function of evaluations is to help make belief and testimony reliable. Something else might play a fundamental role in establishing our reliability. (Maybe we have innate belief-forming mechanisms that are by default highly reliable, maybe due to evolution.)

Also note that the explanation offered by FUNCTION assumes testimony is typically sincere, that is, expresses the speaker’s actual beliefs, or else that audiences can reliably discern whether or not testimony is sincere. I won’t argue here for this background assumption of the offered explanation.<sup>9</sup>

---

8. For defenses of this view, see Craig (1990), Reynolds (2002), Fricker (2007: Chapters 5 and 6), Henderson (2009; 2011), and Henderson and Horgan (in press).

9. For discussion and various defenses, see Sperber et al. (2010), Michaelian (2013), and Sperber (2013).

### 2.3. *Arguing in Support of FUNCTION*

#### 2.3.1. The Methodology: What Are Normative Evaluations Uniquely Suited for?

In Section 2.2, I argued that epistemic evaluations don't serve their function by flagging unreliable testifiers, because this job can be done no worse, and often better, by assertions about others' reliability. That argument relied on a plausible (though empirically defeasible) methodological assumption about the function of epistemic evaluations, an assumption that we're going to continue to put to use in this paper. The assumption is that basic terms of epistemic evaluations serve a function in some unique way. They must earn their keep. Our basic words and concepts for epistemic evaluation are not, I assume, a redundant set of entries in our lexicon of basic words and concepts. Perhaps *some* words and concepts are redundant, always replaceable with another word or concept. Indeed, perhaps, given one concept of epistemic evaluation, say the concept of rationality, another one is redundant, say the concept of justification. I am invoking the methodological assumption of uniqueness only for the family of epistemically evaluative words and concepts considered as a whole.

My argument in support of FUNCTION, here in Section 2.3, relies on this methodological assumption that epistemic evaluations do not serve a function that can just as well already be served by other language, in particular by attributions of reliability. I'll first argue, in Section 2.3.2, that there are important ways that reliability attributions are useless at helping protect the reliability of testimony. Then, in Section 2.3.3, I'll argue that epistemic evaluations can do this important job, and that they can do it in the way described by FUNCTION. This is my argument that FUNCTION is true, that it gives the correct teleological explanation of our epistemically evaluative language.

#### 2.3.2. What Assertions about Reliability *Can't* Do

Assertions about what's reliable have a significantly limited ability to rationally influence another person's belief-forming practices. If you tell your uncle that Fox News is unreliable, he can only rationally decrease his confidence in Fox's reporting if he independently believes you are reliable. And, even if he does independently know you're reliable, your uncle must weigh his reasons to think you're reliable against any reasons he may have to think Fox is reliable. (The reasons on one side may be misleading reasons, but he isn't certain which side that is.)

If you want your uncle to rationally decrease his trust in Fox even more than can be achieved by your bare assertions of Fox's unreliability, you'll need to supply him with additional evidence, evidence he has independent reason to

trust. That can require extra work, but at least it's a possible avenue of influence.

Unfortunately, though, sometimes even that avenue is closed. Consider, now, cases where we have no ability to influence others by using assertions about reliability and presentations of supplementary evidence to elicit a rational revision. These cases illustrate where epistemic evaluations are our best, and often our only, hope of modifying others' unreliable beliefs and testimony.

These are cases where a subject uses an unreliable belief-forming rule, and assesses that rule's reliability using the very same rule. The procedure is, of course, circular. But this doesn't mean that we don't or even that we shouldn't use such procedures. For one thing, assessments of your own reliability cannot help eventually resulting in such circularity. And furthermore, the circularity is not the obviously unacceptable sort of circularity where an argument has its conclusion among its premises; rather, you *use* a rule to reach a verdict *about* that rule.<sup>10</sup>

For our first illustration of such a case, consider Slim, a conspiracy theorist, a sort of person whom I've met in real life too many times. Suppose I notice that Slim's basic inductive rule is incautious; it systematically has him quickly believing radical new theories on slim evidence. (Let's assume Slim doesn't have a basic rule of trusting my testimony.) If I tell Slim he's unreliable, will it influence his basic belief-forming rule? Well, I am not giving him evidence that, by *his* lights, will indicate that he is unreliable. Indeed, *nothing* can constitute evidence that will lead him to conclude that his most basic inductive rule *R* is unreliable, at least so long as he continues to *use R* to draw his conclusions about what reliably indicates what; for we can presume that *R*, like any ordinary basic inductive rule, will endorse the reliability of its own prescriptions.

The gambler's fallacy provides another illustration of inductive self-support that is immune to criticism. Suppose Gabby has been committing the gambler's fallacy, it has been cost her money at the casino, and we point this out to her, hoping to reform her practice. If Gabby is sufficiently dedicated to the gambler's fallacy and commits it yet again when examining her poor past record of success, she will conclude that her luck is due to turn around, and she should stick with reasoning by using the gambler's fallacy!<sup>11</sup>

A similar point holds for basic deductive rules like Modus Ponens. How can you rationally give any weight to the criticism that Modus Ponens is unreliable if you rely on Modus Ponens even to understand, apply and engage in the sim-

---

10. The contrast is between *premise-* or *logical-circularity*, which is clearly irrational, versus *rule-*, *pragmatic-*, or *epistemic-circularity*, which is at least not so clearly irrational; see, among many discussions, Alston (1986), Dummett (1991), Boghossian (2001), and Dogramaci (2010).

11. This is just the well-known point that so-called counter-induction is self-supporting, as observed by Black (1954: Chapter X) and Salmon (1966: 15–6, 133). Another classic discussion of inductive self-support is Lewis (1971). Also see Field (2000) for a notable discussion.

plest reasoning concerning an assertion about its reliability?<sup>12</sup> So, when it comes to influencing others' basic belief-forming rules, both inductive and deductive, there is little possibility of influencing others by making assertions that prompt a rational revision, at least if, as is plausible, a rational revision has to be a revision that your own basic belief-forming rules deem to be reliable.

Basic, self-supporting inductive and deductive rules are not the only cases where we form beliefs that cannot rationally be influenced by the reliability assertions of others. It could be that a reasoner *accepts* only rational and reliable basic belief-forming rules, but she is extremely poor at successfully *following* her own rules because she commits *performance errors*. Wishful thinking, as we all ordinarily engage in it, is arguably best categorized as a performance error: our deepest reflection or our deepest commitments do not have us accepting it, but sometimes we form beliefs using it, sometimes even quite systematically. This illustrates yet another case of resistance to correction by assertions about reliability. Imagine Gulliver believes in reincarnation because of wishful thinking. The wishful thinking can be very robust. Just telling him, "Gulliver, you know that's not true!" won't snap him out of it, because he uses his wishful thinking to come to, and stick to, the conclusion that it *is* true. Gulliver might even, on the basis of wishful thinking, start to follow a rule of trusting everything said in the Bhagavad Gita. And, now, you can't disrupt Gulliver's trust by saying, "Gulliver, [the literal reading of] the Bhagavad Gita is not reliable", because Gulliver, *trusting the Bhagavad Gita*, is of the view that it *is* reliable. This illustrates a case where, once again, assertions about reliability are powerless to rationally influence another person's beliefs, and thus powerless to correct unreliability. The cases immune to rational correction by assertions about reliability thus include not only cases of unreliability in basic rules, both inductive and deductive, but also unreliability due to sufficiently systematic performance errors, systematic enough that the reasoner uses reasoning that *commits* the performance error to assess the error's reliability. Your uncle, the dedicated Fox News consumer, may be yet another example, and again an example that is (frighteningly) common in ordinary life.

---

12. Some philosophers, Dummett (1991) and Boghossian (2001) among others, say we can and should establish the reliability of Modus Ponens itself via a rule-circular argument, an argument that uses Modus Ponens. Dogramaci (2010) raises difficulties specific to this last point, namely whether deductive—though not inductive—rules can be known to be reliable by a rule-circular argument. Still, the point would stand that we use Modus Ponens even to understand, apply and engage in nearly all ordinary reasoning, perhaps including the interpretation of, and reflection on whether to trust, others' assertions.

### 2.3.3. What Normative Evaluations *Can* Do

If anything is to influence another person's beliefs in these resistant cases, it will have to be a form of influence that prompts a *non-rational revision*. Here is where epistemic evaluations can gain purchase: criticizing another as irrational, or praising him as rational, can influence his practice, though not by prompting any rational response to your presentation of any evidence.

Criticism and praise are forms of normative evaluation, forms of evaluations that have a character that includes something not included in the character of assertions about others' reliability. Normative evaluations may include instances of blaming, resenting, shunning, deriding, excusing, or permitting. I focus on simple forms—flatly *calling* beliefs (ir)rational—only for convenience. The character of normative evaluations is, of course, extremely rich and unique in many ways. Here I only want to emphasize one way it is unique: normative evaluation has a unique role in interpersonal influence, and this role enables it to serve a unique function.

Normative evaluations offer a unique means for influencing others, a means of influence that is effective because it's a contingently true psychological fact about us that criticism and praise actually do influence us. Specifically, they influence us to modify, and reinforce, behaviors so as to do less of what's criticized and more of what's praised. There is, of course, no simple and perfectly strict relation between evaluation and behavioral modification. I don't deny the perfectly clear fact that much criticism is ineffective, and much is even counterproductive—it backfires, reinforcing the criticized behavior. This can happen especially when the audience perceives the evaluator as an "outsider" to some group the audience self-identifies with. (When you receive criticism from, say, members of an opposing political party, you might think this is reason in *favor* of the criticized practice!) My approach here claims only that it is not actually the normal effect of epistemic evaluations to be ineffective or to backfire. I claim praise and criticism tend to promote and suppress the evaluated behavior. I only claim this is an actual truth about us; I don't assume it's any kind of necessity. While this approach is hostage to empirical refutation, it seems to me a reasonable default to work with. For one thing, the prevalence of ineffective or backfiring criticism is something we might very easily overestimate due to the so-called availability heuristic: we may more easily recall the frustrating occasions when a stubborn audience ignores criticism; we may more easily forget the many mundane occasions when criticism has its usual influence.<sup>13</sup> And *praise* seems to hardly ever backfire: just about all praise, even transparent flattery, is welcomed.

Another point in defense of the effectiveness of evaluations is that often the same criticism succeeds if merely slightly modified to be made more diplomatic,

---

13. See Tversky and Kahneman (1973) for the classic discussion of the availability heuristic.

maybe, for example, so that it's directed at some third-party engaging in the same bad behavior as the audience of the evaluative assertion. (I strongly suspect most criticism is third-personal. Just think of how much nasty gossip you hear in a single day.) This relates to the aforementioned suggestion that "outsider" criticism is what we most often bristle at. That fact suggests third-personal evaluations can be especially powerful, by casting the criticized practice as a practice shunned by "our" group, a group the evaluator is inviting the audience to perceive as including them both, but excluding the evaluated third-party. Example: parent observes child exhibiting some prejudice, and then teaches child by using third-personal evaluations: 'Just look at those Sneetches discriminating against each other. Aren't they being so silly!'.<sup>14</sup>

It's certainly hard to characterize in any great detail the relation between evaluation and the behaviors it promotes or suppresses. All I am claiming is that the general contour of the relation is intuitively clear: criticism suppresses, and praise reinforces, the target behaviors. This is how normative evaluations provide us with a means for influencing one another's basic belief-forming rules, influencing them, as we often must, non-rationally. A conclusive case for any detailed view of the relation between normative evaluations and human responses would require substantial empirical psychological research. For my purposes here, I mean only to be offering a general suggestion, one that I hope will be taken as the face-value view of how evaluations influence others—just that, by and large, praise reinforces and criticism suppresses the target behaviors.

Again, our means of "persuasion" here is not rational persuasion. It does not involve the subject treating his evaluator as providing any evidence for him to respond to. This suggested description of how non-rational factors play a role in influencing each others' belief-forming practices may sound cynical. But, I don't think the non-rational role that I'm arguing normative evaluations have on group inquiry is really antithetical to our search for truth at all. On the contrary, I mean to be arguing that it is extremely valuable that normative evaluations play this role. We need normative evaluations to influence others' belief-forming practices in order to serve the function, as described in *FUNCTION*, of not simply excluding others from the circle of trust, but positively improving the reliability of bad testifiers. The non-rational influence of normative evaluations is our only hope of influencing their basic belief-forming rules. Slim, the incautious reasoner who too quickly infers a radical new theory on slim evidence, is a bad testifier, but can be made into a good testifier if he can be influenced to adopt a more cautious belief-forming rule for induction. While a handful of examples

---

14. For additional discussion and support of the idea that it's the epistemic criticisms of outsiders that we bristle at the most, see Fricker (2007: Chapter 5, esp. 115–6). As Fricker notes, our comparative unwillingness to trust outsiders may plausibly be due in large part to prejudicial stereotypes.

of “unteachable” paranoids may easily come to mind, so many others among us, especially while we are young, do learn to fix our incautious epistemic habits. Gulliver, who stubbornly repeats his wishful belief in reincarnation, cannot be made to rationally see the unreliability of a book endorsing reincarnation, but a normative evaluation can, if saying anything to him can, improve his bad epistemic habits. We bring about these improvements just by persuading our subject to revise his belief-forming practice—even while he cannot see, by his own lights, that he is incautious, or, as we’d put it, unreasonable.

Our ability to bring about these improvements does not require that Gulliver or any subject can *decide*, or *choose*, what beliefs he will adopt on any given basis. There is no commitment here to any implausible form of voluntarism about belief, only a commitment to subjects’ having a degree of control over their beliefs—whatever is required such that outside evaluations can have an influence on their belief-forming practices. That we have this much control is a modest, plausible claim. The claim that we have this much control over our beliefs has been accepted even by philosophers who strongly deny that we have *voluntary* control over beliefs.<sup>15</sup>

So, epistemic evaluations can help protect us from bad testifiers by making them into good testifiers; that is, FUNCTION is true. Epistemic evaluations must play this role, for it is often only normative evaluations that have any power to influence others. Normative evaluations do not *necessarily* influence others, but they actually tend to, and it is valuable that they do, because then they can serve this valuable function of helping maintain a safe system of spreading true belief by testimony.

It’s not a magical cure-all; evaluations cannot turn an unreliable entire community into a reliable one, but they can help maintain the reliability of a system whose members are (a) prone to veer off the straight and narrow path, the reliable path, and (b) good at detecting others’ flaws but bad at detecting their own flaws. If such a communal support system is useful for anyone, it’s useful for people like us. Though, as noted, the ultimate case for or against FUNCTION is empirical, the case looks highly favorable. Consider the very normal tendency we each have of lapsing into unreliable performance errors anytime emotions, desires, or other biases disrupt us from following our own accepted rules for inductive reasoning.<sup>16</sup> Consider how often an easy-to-use, oversimplified heuristic, though perhaps an efficient tool in the long-run, leads us astray in a given case.<sup>17</sup> Consider how our careful, reflective belief-forming rules, (system 2), regularly

---

15. See, for example, Hieronymi (2006; 2008).

16. On the psychological research into errors that are due to emotion and desire, see Kunda (1990) and Molden and Higgins (2005).

17. On errors due to heuristics and biases quite generally, see Gilovich, Griffin, and Kahneman (2002), and Kahneman (2003).



fail to catch many of the unreliably formed beliefs produced by our quick and unreflective belief-forming rules, (system 1), beliefs we can easily recognize, if we are careful and reflective, as unreliable.<sup>18</sup> We seem to be so much better at catching these “errors” in others than in ourselves.<sup>19</sup>

(An independent argument for FUNCTION is that it follows from another view I’ve defended, *epistemic communism*. Communism makes a stronger claim than FUNCTION; it says we use evaluations to modify others’ beliefs so as to promote the coordination of our basic belief-forming rules, and this coordination allows us to efficiently and safely share true beliefs by testimony. Communism is independently supported by its ability to solve some puzzles about epistemic evaluation. But since communism is a stronger view than FUNCTION, and the argument for it can be found in other papers, (see Dogramaci 2012; 2015), I’ll leave things here with the argument already given.)

#### 2.4. Clarifications about FUNCTION

Before proceeding to the next section, where we’ll return to directly tackling the problem of forgotten evidence, let me add a few clarifications about what, in endorsing FUNCTION, I have *not* said here about the teleological explanation of our epistemically evaluative practice. Skip this subsection if you want to skip subtleties.

---

18. On systems 1 and 2, the “dual process” theories of reasoning, see Sloman (1996); Stanovich (1999); Evans (2003), Gilovich, Griffin, and Kahneman (2002) and Kahneman (2003).

19. For psychological evidence that we are better at catching errors in others than in ourselves, see Pronin, Lin, and Ross (2002), Mercier and Sperber (2011) and Sperber and Mercier (2012).

Mercier and Sperber (2011) have a theory that is in many ways similar to and supportive of my project here. They wish to propose a function to explain part of our linguistic practice, and that function is to improve the reliability of cognition through social mechanisms. But our theories, though they are complimentary, focus on explaining different phenomena. Mercier and Sperber heavily appeal to data from cognitive science that shows a confirmation bias and self-serving motivations in the ways humans produce arguments in favor of their own beliefs. Mercier and Sperber cite this as crucial support for a theory that aims to explain the biased and motivated ways in which people make conscious to themselves and share with others reasons for their beliefs. The data that I, on the other hand, am most concerned to explain is the way ordinary evaluations of reliability and rationality play distinct roles in our linguistic practice.

Mercier and Sperber are *also* interested, as I am, in explaining how people evaluate the quality of testifiers’ reasoning. I endorse what they say on this, especially in Sperber et al. (2010) and Sperber (2013). But they don’t attend to the crucial distinction that I’m most concerned with, namely our evaluation of beliefs as reliable vs. our evaluation of them as rational. I’ve argued these evaluations play different roles, and I wish to explain the distinctive role of our evaluations of beliefs as rational.

For criticism of the view that epistemic evaluation has a first-personal function, one that guides individuals to improve their thinking on their own, also see Dogramaci (2015: Section 1.4).

**First clarification:** I've intended to argue in support of FUNCTION as describing the *primary* function of epistemically evaluative language. I've only made a defeasible case for this. It would be an objection to my view if it could be shown that epistemic evaluations have some other primary function, if something else better teleologically explained why we engage in this practice. But my view is consistent with the possibility that epistemic evaluations usefully contribute to other secondary functions, including the function that's primarily served by assertions about reliability. Likewise, other terms, for example knowledge attributions, might contribute to the function that FUNCTION describes for epistemically evaluative language, though I'd argue (given my sympathy, already noted, for Edward Craig's view) that knowledge attributions do not have this for their primary function. The present clarification is important because, if we observe some temptation to use epistemic evaluations in either of two incompatible ways, as perhaps in the problem of forgotten evidence, we should resolve the tension by taking the correct use of epistemic evaluations to be the one that serves their own primary function.

**Second clarification:** I said earlier that a candidate teleological explanation succeeds in explaining why a practice exists if, and only if, the practice efficiently promotes an interest that it is independently plausible that the participants would strive to promote. The interest cited in my proposed teleological explanation for our epistemically evaluative practice is our interest in having true beliefs. As already noted, this is an interest we will endorse promoting even upon reflection, and thus if—and when—we find that a certain response to the problem of forgotten evidence better serves this interest, that's the response we'll have good reason to continue engaging in. The clarification I want to add here is that, for all I'm saying or need to say in this paper, it can be (and very likely is true) that our epistemically evaluative practice serves some much deeper interest of ours, one served via our interest in true belief. Perhaps the teleological story I'm starting here is best completed by giving an evolutionary explanation of our use of epistemically evaluative language as promoting our survival interest, perhaps also adding that our epistemically evaluative practice is genetically selected and innate (similar to a Chomskyan account of our general faculty for language). This is consistent with what I've claimed, as long as true beliefs tend to promote survival. But I don't want to make any claims concerning our evolutionary history, since, (a) I don't see that I need to for my purposes in this paper, and (b) for all I know, it will turn out that the best complete teleological explanation of our epistemically evaluative practice is something else, something unexpected; for example, maybe epistemic evaluation is merely a product of some sort of non-genetic "cultural" evolution.

**Third clarification:** I have not endorsed the following view, which I'll call *instrumentalism*. Instrumentalism says a belief is epistemically rational for a sub-

ject if, and only if, the subject possesses the goal of having true beliefs and it is thereby instrumentally rational for the subject to hold the belief. Kelly (2003) gives a compelling critique of instrumentalism. Kelly points out that in some cases, while it is epistemically rational to hold some belief, it is not instrumentally rational. I happen to agree with Kelly. My proposal has been a different one from instrumentalism. I only claim that it serves an interest of ours, an interest served by having true beliefs, *to engage in our practice of epistemically evaluating each other*. I do not even claim that serving this interest is a goal we possess in any psychologically real way. (I did claim that true beliefs serve our interest when the things we desire, i.e., goals we possess, are also in our interest, but this doesn't undermine the present point. We don't need to positively desire, to have as a goal, true belief. We need only desire things in our interest, like food, and then having true beliefs about the means to those goals will be in our interest.) I do not claim that every manifestation of our epistemically evaluative practice serves an interest of ours. I only claim that the practice as a whole serves our interest. One could argue that some other practice could advance our interests better than our actual practice; but, unless the practice is equally *efficient* at serving our interest, this will not obviously offer a better teleological explanation of our actual practice. A practice that would require much more effort or more resources to engage in would not be clearly more efficient. (Consider, for example, a variant of our actual practice, one which involves withholding evaluations when certain calculations would suggest the evaluation will be unlikely to have much effect.)

**Fourth clarification:** I have not endorsed the following view, which I'll call *epistemic belief-consequentialism*. This view says that a subject's belief is rational if, and only if, it serves some "epistemic goal" for that subject to hold that belief. One paradigmatic example of what might be treated as the epistemic goal is true belief, and you might easily think that I have here endorsed a version of epistemic consequentialism with true belief as the epistemic goal. That would be a misunderstanding; I have not endorsed that view. I've endorsed only a view of epistemic evaluation. I've claimed that we engage in a way of *criticizing and praising* that efficiently promotes a certain epistemic goal, the goal of true belief. This claim is false if, and only if, our overall practice of making evaluations actually does fail to promote this interest. Epistemic belief-consequentialism, on the other hand, will be false if, and only if, it is possible for a belief to be rational while the subject's *holding that belief* fails to promote any epistemic goal. The following combination is perfectly good: we actually promote our interests, promote them quite efficiently, by engaging in an epistemically evaluative practice that has us, in some cases, criticize (praise) some individual beliefs the subject's holding of which does (doesn't) promote some (any) epistemic goal. I think this actually happens in cases, including BonJour (1980)'s unwittingly reliable clair-

voyant: the clairvoyant's having these beliefs serves the goal of true belief, but we criticize these beliefs. Another example is the unwitting victim of a Cartesian demon: we may praise false beliefs of his as rational even while his holding these beliefs does not serve, but harms, the goal of true belief (as well as any other plausible goal). The overall evaluative practice that includes these particular evaluations still efficiently promotes true belief, via its function of promoting the safety of testimony, which is so effective at spreading true beliefs throughout the community.

The case of the clairvoyant also illustrates the falsity of another view I've not endorsed, a view we can call *epistemic rule-consequentialism*. This view says that a subject's belief is rational if, and only if, the subject formed the belief by following a rule, the following of which serves some "epistemic goal". It seems to serve the paradigmatic epistemic goal, true belief, for the clairvoyant to follow her highly reliable rule. Yet her beliefs are not rational. And it might (and I believe it does) promote true belief to *criticize* the clairvoyant.<sup>20</sup>

Selim Berker (2013) has recently given extensive attention to clarifying and criticizing a broader meta-epistemological view that includes what I've just called epistemic belief-consequentialism and epistemic rule-consequentialism. The target of Berker's critique also includes the following view, which I'll call *epistemic recommended-rule-consequentialism*. This view says that a subject's belief is rational if, and only if, the subject formed the belief by following a rule, a rule from the list of rules which it serves some "epistemic goal" for us to exclusively praise. My view is that this biconditional is actually true. But, while I agree with the truth of this biconditional, I don't see it as offering any *explanation of why* this or that belief or rule is rational. My explanatory aims are only to explain why we make certain evaluations. I'm in fact not disagreeing with Berker's view, since his aim is to reject consequentialist explanatory theories of epistemically rational belief.<sup>21</sup>

### 3. Application to the Problem of Forgotten Evidence

#### 3.1. Applying FUNCTION to Joe the Coffee Addict

Let's, finally, turn back to the problem we started with. The problem of forgotten evidence was a dilemma, one I posed using the example of a coffee addict, Joe, who formed a belief, on the basis of wishful thinking, that drinking coffee is healthy, but has since entirely forgotten how he formed this belief.<sup>22</sup> We saw that

20. See Dogramaci (2012) for why criticizing the clairvoyant promotes true belief.

21. See Dogramaci (2015) for why there is *no* explanatory theory of epistemically rational belief.

22. Remember that Joe is just a concrete example serving as a placeholder for any case of irrational thinking where the subject has forgotten his or her evidence; the present discussion

there are challenges facing either way of answering the question: is his belief now rational or irrational? To solve the dilemma using the practical approach, the question we turn to now is this: which evaluation would, or would more efficiently, serve the function of epistemic evaluation, as described in FUNCTION? (As an intuitive shorthand, I'll now often speak of "serving FUNCTION" to mean serving the function that's described in FUNCTION.)

Reminder: the larger argumentative strategy relies on the premise that evaluations that serve FUNCTION will be *correct* applications, ones that apply our word or concept (of rationality) to their extension. Thus, we will ultimately be in a position to draw a conclusion about the rationality of Joe's belief, a conclusion we state by *using* our word 'rational' or our concept of rationality. And we'll want to continue to use this word, we'll want to continue to make evaluations that express this concept, because we of course endorse the goal of FUNCTION, namely the goal of promoting true belief.

FUNCTION said that we promote true belief in a particular way. It said that epistemic evaluations help protect us from unreliable testimony, not by excluding unreliable testifiers from the circle of trust, but rather by influencing unreliable testifiers to modify their beliefs and, thereby, modify their testimony. Now, Joe certainly is an unreliable testifier: he engaged in wishful thinking, an unreliable belief-forming practice, and he has now forgotten what his evidence was (if any) for his belief that coffee is healthy, so he cannot self-correct his unreliability. This may seem to suggest that Joe is a prime candidate for someone whose belief-forming practices, and thus whose testimony, would benefit from correction by the epistemic evaluations of others; Joe may seem to be someone whose belief, according to FUNCTION, we should criticize.

I will argue that is wrong. Criticism of Joe's belief would not help improve the reliability of his own, or anyone else's, beliefs and testimony. Not every instance of unreliability is treatable by epistemic evaluation. Sometimes unreliability can only be treated, or would be better treated, by a non-normative assertion, such as an assertion about reliability; sometimes it cannot be treated by any assertion. In the case of Joe, I'll argue that epistemic evaluations, which are normative, would not contribute to improving the reliability of either Joe's, or anyone else's, testimony.<sup>23</sup>

---

is general.

23. Greco (2005: Section 3) aims to argue that any plausible view of the purpose of epistemic evaluation will require that we criticize someone in Joe's position. But, Greco's focus does not involve any consideration of how evaluations influence the targeted believers, so his discussion does not engage with what I want to argue on the basis of FUNCTION.

### 3.1.1. Examining the effects of criticism

I'm going to focus on examining whether epistemic *criticism* of Joe's coffee belief would serve FUNCTION. I focus on criticism for the following reason. When criticism serves FUNCTION, it does so by acting as a *corrective*, as a means of counteracting bad practices. So, in order to serve FUNCTION, criticism must make some difference that improves the situation; epistemic criticism must make a beneficial difference to the reliability of belief-forming practices and, thereby, of testimony. If it doesn't make a beneficial difference, then criticism will not serve FUNCTION. In such a case, we best serve FUNCTION by issuing no corrective advice, no recommendation for change. In such a case, if we must make any evaluation at all, we best serve FUNCTION by making one that expresses the *permissibility* of the existing belief. I make the standard assumption that every belief falls either under the concept of epistemic irrationality or the concept of epistemic permissibility. Since our ordinary concept of epistemic rationality expresses such permission (whether or not it also expresses obligation), we thus serve FUNCTION by evaluating a belief as rational whenever we cannot promote a beneficial correction. Although epistemic praise of course plays its own important role in promoting reliable belief-forming practices and thus reliable testimony, it is not primarily corrective. We can reinforce good practices with epistemic praise, and we can use praise as an indirect corrective by praising the practices we wish our audience would revise theirs to. But it's criticism that is our primary and direct means of correction.<sup>24</sup>

So, my main concern will be to ask: would an epistemic criticism of Joe's coffee belief help counteract the unreliability of testimony? I'll look at this question in two stages. First I'll ask whether criticizing Joe's belief would help improve Joe's reliability (Section 3.2.1). Then I'll ask whether criticizing his belief would help improve the reliability of others (Section 3.2.2).

In examining this, I'm going to rely on the premise that if epistemic evaluations of Joe's belief would improve the reliability of testimony, then they would do so by suppressing future instances of wishful thinking. I take that premise to be plausible. What other unreliable belief-forming practice would we be promoting or suppressing? (I will discuss the question of targeting *forgetfulness* below; see Section 3.3.)

A pre-emptive clarification: I want to argue that criticizing Joe's belief as irrational would not make him or anyone else more reliable in the future, but, of course, if we cook up the right outlandish, artificial background context, we can make just about anything trigger just about any behavioral effect. However, the question we're interested in here, the question relevant to the actual function

---

24. Fricker (in press: Sections 2–3) also argues for a distinctively corrective function of negative evaluation.

of epistemic evaluations, is only this: in a normal case of forgotten evidence, one typical of the actual cases that make true an actually correct teleological explanation, in particular the explanation offered by FUNCTION, would normative criticism of the subject make him more reliable in the future? This is the question relevant to our own actual predicament concerning how to evaluate the beliefs of subjects like Joe. (The notion of ‘normal’ here may have some vague boundaries, but we won’t be considering any borderline or otherwise controversial cases of normality to settle the question we’re asking here.) Our question is, given any remotely normal background circumstances, would criticizing him make him, or anyone else, more reliable?

### 3.2. Criticizing Joe Won’t Serve FUNCTION

#### 3.2.1. Suppressing Future Wishful Thinking on Joe’s Part

There’s no doubt that wishful thinking is irrational, and also that it serves FUNCTION to criticize wishful thinking. Our question, however, is whether it serves FUNCTION to criticize Joe’s *later, surviving* belief, or whether, instead, we should allow the period of forgetfulness to cleanse, or launder, the originally irrational belief. If it were to serve FUNCTION to call the later belief irrational, then it must be that doing so would, somehow, influence others in a way that makes their beliefs and testimony more reliable, in particular by suppressing wishful thinking.

And, for better or worse, calling Joe’s coffee belief irrational would not help to suppress wishful thinking on Joe’s part. This is so whether or not Joe is someone who is particularly prone to future acts of wishful thinking (as his past indiscretion might indicate he is). The reason is simple: criticism would be ineffective on Joe simply because he has forgotten how he formed his belief that coffee is healthy—it’s because of the very defining feature of the case! Because he has lost all memory of how he formed his belief, Joe has no way of connecting his belief to any disposition he may (or may not) have to engage in wishful thinking.

Of course, you could *inform* Joe that his coffee belief was based on wishful thinking; perhaps doing this will be part of what you could and should do to somehow encourage him to address a disposition to engage in wishful thinking. But, once you’ve identified Joe’s coffee belief as based on wishful thinking, you’ve identified it to him as unreliable, *and now he longer needs any further normative evaluation*, that is, he no longer needs to be called irrational—it serves no further purpose, no further function. The idea here is that *without* identifying to Joe that the coffee belief was based on wishful thinking and hence is unreliable, there’s no use, no benefit, in giving it a negative normative evaluation, whereas *in conjunction with* identifying it to him as based on wishful thinking and hence as unreliable, there’s *still* no use, no benefit, in *also* giving it a negative normative evaluation. Remember our methodological assumption, which we are applying

once again here: normative evaluations must serve a function that differentiates them from assertions about reliability. If assertions about reliability do all the work that's needed to address a potential disposition Joe may have to engage in wishful thinking, there's nothing left for a normative evaluation of his coffee belief to do. If Joe has his coffee belief identified to him as unreliable (and accepts the news) and he *holds on* to it, then you can usefully criticize the belief, but the criticism here is of a completely new failure of his, failure to appropriately respond to a defeater he possesses.

The invulnerability Joe exhibits to the function of epistemic criticism is special; it contrasts with what happens when epistemic evaluations do play their proper, effective role. Consider again the many illustrative cases we reviewed in Section 2 while making the case in support of FUNCTION. When Gabby has not forgotten that her belief that the roulette wheel will next come up black is based on a belief that the recent string of reds has made black overdue, then it serves FUNCTION to criticize this belief. When Gulliver knows that his belief in reincarnation depends on his very selectively trusting certain favorable sources of testimony, again it serves FUNCTION to criticize that belief. Similarly for the other characters. It's not a waste of breath to be (diplomatically) critical of beliefs that are knowingly based on unquestioned dedication to Fox News—it's just exhausting. But, it *is* a waste of breath to criticize Joe's coffee belief, at least for the sake of improving Joe's own belief-forming habits.

### 3.2.2. Suppressing Wishful Thinking on the Part of Others

I just argued above (in Section 3.2.1) that it wouldn't help suppress future wishful thinking on Joe's part to call his belief irrational after he's forgotten his original (lack of good) evidence. And I also offered it as a plausible premise (in Section 3.1.1) that if epistemic criticism of Joe serves FUNCTION, then such criticism would help to suppress future wishful thinking. To now complete the overall argument that criticism of Joe wouldn't serve FUNCTION, we need to cover the remaining possibility.

There remains the possibility that a negative evaluation may still serve FUNCTION by suppressing wishful thinking not in Joe but in others, such as a third-party audience to the evaluation, or even others who may later be influenced by such a third-party audience. The audience to an epistemic evaluation need not be the subject of the evaluation; the audience to our criticism of Joe need not be Joe himself. In such a case, might epistemic criticism of Joe's belief serve FUNCTION? It would if it influenced others so as to discourage future wishful thinking on their part. So, before we can conclude that the function of epistemic evaluation would have us call Joe's belief rational, we must convince ourselves that calling Joe irrational will not, at least not any more than calling his belief rational would, discourage wishful thinking on the part of others.



Although an epistemic criticism would have no beneficial effects on Joe, criticisms of subjects like Joe could have different effects on a third-party, but *only if* the third-party somehow *is aware* of the pasts that these subjects have forgotten about themselves. If the third-party has no idea, the effects on him will be the same as the effect on Joe, and thus criticism of Joe again wouldn't serve FUNCTION. So, would criticizing Joe to an *in-the-know* third-party help serve FUNCTION? In particular, would it help suppress wishful thinking on the part of the audience, or others the audience in turn influences?

The view I need to attack here is that, if someone engages in wishful thinking, then even if he forgets about it later, his *surviving belief* is worth criticizing, perhaps as way to discourage others from engaging in any wishful thinking themselves. And, the problem with this view is that it again violates our old (defeasible but plausible) methodological assumption that epistemic evaluations are not used in a way that makes them serve a redundant function, a function that other uses of language serve just as well, or even much better. In this case, other assertions are clearly the effective tools for serving any purposes we might have here. If you are in the know about Joe's forgotten (lack of good) evidence, the best way for me to discourage you from engaging in, or promoting in others, any future wishful thinking is *not* for me to criticize to you Joe's *surviving* belief that coffee is healthy. A better way of promoting that outcome would be for me to *just* criticize Joe's *early* belief, his belief resulting from wishful thinking *before* he forgot his evidence, and since you are in the know, there is nothing confusing or problematic about my criticizing this to you. I could also just assert the unreliability of Joe's wishful beliefs, although here I may need to be slightly more careful for two reasons. First, a normative evaluation of wishful thinking will often be more useful than a mere assertion about reliability, for the reasons given earlier (in Section 2) about the ways in which assertions about reliability may have less influence on others. And second, it does no good if I'm misinterpreted as critically commenting on, and trying to suppress future episodes of, someone's *forgetting* his evidence. Epistemic evaluations aren't used to criticize episodes of *forgetting*.

This concludes my positive argument that a negative epistemic evaluation, criticism of Joe's belief as irrational, doesn't serve the function of epistemic evaluation, doesn't serve FUNCTION. The evaluation that serves FUNCTION is permissive, and so I conclude we serve FUNCTION by calling Joe's belief rational. Placed within my larger strategy of inferring semantic facts about the extensions of our evaluative words and concepts from their function, and thereby inferring epistemic facts about who is rational, this completes my positive argument that Joe's belief is rational.

Before I turn next to some defensive points, there is one more point of support for FUNCTION that I can now quickly add. It should now be easy enough to

see that it also doesn't serve FUNCTION to criticize a belief that was based on *good* evidence that has now been forgotten. Epistemologists on both sides of the forgotten evidence dilemma agree that *this* belief remains rational (see the authors listed at the end of Section 1.1), and this gets accepted as just intuitively true. I consider it yet another supportive point in favor of FUNCTION that it explains this widely shared intuition.

### 3.3. *What's So Bad about Laundered Beliefs?*

To fully satisfactorily resolve the problem of forgotten evidence, more needs to be said than just a positive argument for one side. The problem is a dilemma, either horn of which immediately strikes us as counter-intuitive. I have given an argument that Joe's belief is rational, but, even if that's completely right, this may still leave us with our original discomfort with the thought that Joe can get away with laundering his belief. It is one thing to learn which horn of a dilemma is true, and even to know why it is true; it's another thing to feel no intellectual discomfort making this judgment. A fully satisfactory resolution of our problem, then, would not just convince us that Joe *can* launder his belief, but would also resolve our discomfort with this judgment. We can do this if we can explain why we *misleadingly* feel tempted to reject this judgment. And there are several points that have emerged over the course of our discussion that begin to explain this.

Even as we've argued that Joe's belief is rational, we've emphasized a number of ways in which Joe exhibits cognitive defects. Joe formed his belief about coffee in an unreliable way. Unreliability is a defect, a cognitive defect worth discouraging. Joe also forgot how he formed his belief. Forgetfulness can also be a cognitive defect, one that it is often, though not always, worth discouraging. Of course, no one can remember everything, and there are certain things it's a waste of cognitive resources to bother remembering, but it's still true that forgetfulness is often harmful and worth discouraging.

These are defects of Joe's cognitive state, defects worth discouraging linguistically or by other means. If part of our language effectively functions to discourage these defects, then it may naturally require extra care not to confuse the role of that language with the role of the kind of language I've narrowly focused on here, the language I called (in Section 2.1) our epistemically evaluative practice. We certainly do have other language, language other than epistemic evaluations, that partly functions to discourage unreliability and harmful sorts of forgetfulness. "Think carefully!", "Pay attention!", "Gather up plenty of evidence!", "Double-check your conclusions!", "Avoid sloppy reasoning!", to give only some simple illustrations.

Not only are Joe's cognitive defects worth discouraging, they are worth *flagging* to others. In Section 2.2, I said that we have two broad ways of promoting

the safety of testimony. One is to encourage reliable belief-forming practices (the function of epistemically evaluative language), and another is to flag unreliable beliefs and belief-forming practices. I said that I, with many others, find Craig's proposal plausible: knowledge attributions serve the function of flagging reliable and unreliable sources of testimony for the safe consumption by others. Joe's coffee belief, even if it is true, is not knowledge.<sup>25</sup> It is thus appropriate to say Joe doesn't know that coffee is healthy, and his belief and his testimony are thereby flagged as unreliable and untrustworthy.

It is easy to confuse these defects of Joe's belief with irrationality. And it is easy to confuse the effective ways we do have of linguistically responding to these defects with the ways we use epistemic evaluations for their own purpose. But, I've argued that epistemic evaluation serves a subtly but importantly different purpose from other language, especially reliability attributions and knowledge attributions. Careful reflection on these distinctions can help loosen the grip of our initial temptation to balk at Joe's laundering of an irrational belief just by forgetting his evidence.

In fact, the whole metaphor of laundering beliefs is part of the cause of the confusion! The misleading intuition in the original dilemma is pumped by figurative talk of laundering. But, now we can better see that *laundry-ing* a belief is one thing, and a *laundry-ed* belief is another. It is appropriate to criticizing *laundry-ing*: we might do this by criticizing Joe for forgetting his past wishful thinking. But, it's not the function of epistemic evaluations to evaluate *forgetting*. They serve to evaluate the *laundry-ed* belief, the clean belief.

### 3.4. *Objection and Reply: On an Alternative Application of FUNCTION*

I've argued that a practice which has us *not* criticize Joe better serves FUNCTION than one which has us criticize him. Given my argument that our epistemically evaluative words are tailored to efficiently serve FUNCTION, this is evidence that our word 'rational', and our concept of epistemic rationality, really have Joe's belief within its extension. And, given that we do, even upon reflection, want to serve FUNCTION as best we can, we have reason to continue to use this word and this concept, and to evaluative Joe's belief as rational.

In this subsection, I'll consider a potential objection.<sup>26</sup> I want to consider an account of our epistemically evaluative practice, one which includes criticism of Joe as irrational, and which it might be argued serves FUNCTION as well as, or better than, my proposal above. I'll describe the proposed account, highlight why it may appear to serve FUNCTION better, and then argue it serves FUNCTION more poorly. So, since it is poorer, it is not likely that our actual practice, which

---

25. If true, it is a Gettier case, as Conee and Feldman (2001: p.10) observe.

26. Dan Greco suggested (but wished to not endorse) the proposal in the objection.

is tailored to serve FUNCTION efficiently, operates in this way. Nor will we want to switch to using language and concepts that do operate in the described way.

Let's name the new proposal to be considered *sneak-catcher*, and to introduce it and explain its name, let me introduce Sneaky, one more character to compare against Joe. Sneaky, like Joe, formed a belief that coffee is good for you purely on the basis of what she then knew was wishful thinking. Unlike Joe, Sneaky still remembers she formed her belief in this way (and has since acquired no new basis for this belief). Sneaky, though, now presents herself *as though she doesn't remember at all how she formed her belief*. Thus, Sneaky's belief is not exactly like Joe's, but it's difficult for people who interact with her to distinguish her case from Joe's.

Given all the facts about her, I take Sneaky's belief to be uncontroversially irrational. And, I take it that it best serves FUNCTION to criticize beliefs like Sneaky's: managing one's beliefs as she does is an unreliable method that we effectively suppress using epistemic evaluation.

The *sneak-catcher* view, now, says that we most efficiently promote reliability overall, and thus efficiently serve FUNCTION, by criticizing *both Sneaky and Joe*.

What can be said in favor of the *sneak-catcher* view? The apparent argument is this. It serves FUNCTION to criticize Sneaky. But, any rule that recommends we don't criticize Joe and do criticize Sneaky will be hard to conform to, hard to accurately implement in practice: we will frequently make errors either by letting Sneaky off the hook or by wrongly accusing Joe. Thus, to ensure we don't let Sneaky off the hook, we should adopt a rule for criticizing both, and this rule will be easy to implement without such frequent errors.

It may be helpful to note there is an analogy here to strict liability in the law. Laws imposing strict liability hold a person responsible for a harm she causes, regardless of whether or not the harm was foreseeable. Now, strict liability in the law might have a pragmatic justification: in order to catch and punish those who could have foreseen harms they caused, we find it necessary, or simply worth the costs, to implement a rule that also has us punishing others who could not have foreseen harms they caused. But, strict liability does not have the same appeal outside the legal context. To many ethicists, the *sneak-catcher* view's analogy with strict liability will immediately cast the view in an unfavorable light. As Thomas Nagel has commented, in an ethical context, "If the object of moral judgment is the *person*, then to hold him accountable for what he has done in the broader sense is akin to strict liability, which may have its legal uses but seems irrational as a moral position." (1976/79: 31); and Gideon Rosen has likewise said, 'But in morality there is no such thing as strict liability' (2004: 301).

Here is why I don't buy the above apparent argument for *sneak-catcher* (aside from a temptation to dismiss it out of hand, perhaps as Nagel and Rosen would). Compared to a rule for criticizing both Sneaky and Joe, it's no harder to imple-

ment the following *pair* of rules: 1) as long as you can't tell them apart, criticize both Sneaky and Joe, but 2) *when you are able to tell them apart*, just criticize Sneaky, not Joe. This pair of rules is no harder to jointly implement (no harder than just following (1), which is just *sneak-catcher*), and the effect of following both rules dominates following just the one in terms of serving FUNCTION. Certainly, as theorists who are considering Joe, having stipulated all the facts about his forgetting whatever evidence he did or didn't have for his belief, we can apply rule (2), and decide not to criticize him. Thus, I stand by the diagnosis that Joe's belief is rational.

#### 4. Summary

In Section 1.1, I presented a dilemma: in a case where a subject, Joe, has forgotten that he formed an initially irrational belief by wishful thinking, it seems problematic both to say his belief is now rational and to say it is still irrational. In Section 1.2, I proposed to make new headway on the problem by examining the practical function of our epistemically evaluative practice. And I stated an assumption about our words and concepts having the simplest semantics that allows them to serve their function: in particular, if it best serves the function of our epistemically evaluative practice to call a belief based on forgotten evidence (ir)rational, then that belief is in the extension of our word, '[ir]rational', and the concept of (ir)rationality our word expresses.

In Section 2.1, I clarified my assumption that our epistemically evaluative practice has a function: I mean there exists a teleological explanation for our practice of using language like '[ir]rational' or '[un]justified'. In Section 2.2, I proposed a teleological explanation, encapsulated in FUNCTION: the explanation for why we have our epistemically evaluative practice is that it serves to modify the belief-forming practices of others so as to promote reliable beliefs and testimony. In Section 2.3, I argued in support of this proposal. My strategy depended on a methodological assumption: our basic epistemically evaluative words serve a non-redundant role, one that other basic language, in particular the language of reliability, does not serve as well or better. I argued that epistemically evaluative language, with its normative character, is uniquely suited to influence the belief-forming practices of others, especially our most basic and most systematic belief-forming methods.

In Section 3.1, I argued that calling Joe's belief irrational serves FUNCTION only if it would suppress wishful thinking on his part or on the part of others. In Section 3.2, I argued that it would not suppress wishful thinking on anyone's part. Thus, I argued that calling Joe's belief rational is what best serves FUNCTION. In Section 3.3, I tried to untangle the different roles of different parts of our language, confusion of which might have misled us into feeling discomfort with

thinking that Joe's belief is rational after he forgets how he formed it. In Section 3.4, I criticized an alternative proposal on which we better serve FUNCTION by criticizing Joe.

## 5. Acknowledgements

For help during the preparation of this paper, I must thank Yuval Avnur, David James Barnett, Michelle Dyke, Dan Fogal, Dan Greco, Miriam Schoenfield, David Sosa, Dan Waxman, students in a graduate seminar and students and faculty in a reading group at the University of Texas at Austin, and two anonymous referees.

## References

- Alston, William (1986). Epistemic Circularity. *Philosophy and Phenomenological Research*, 47(1), 1–30.
- Annis, David (1980). Memory and Justification. *Philosophy and Phenomenological Research*, 40(3), 324–333.
- Berker, Selim (2013). Epistemic Teleology and the Separateness of Propositions. *The Philosophical Review*, 122(3), 337–393.
- Black, Max (1954). *Problems of Analysis*. Cornell University Press.
- Boghossian, Paul (2001). How Are Objective Epistemic Reasons Possible? *Philosophical Studies*, 106(1), 1–40.
- BonJour, Laurence (1980). Externalist Theories of Empirical Knowledge. *Midwest Studies in Philosophy*, 5, 135–150.
- Burge, Tyler (1997). Interlocution, Perception and Memory. *Philosophical Studies*, 86(1), 21–47.
- Christensen, David (1994). Conservatism in Epistemology. *Noûs*, 28(1), 69–89.
- Christensen, David (2010). Higher-Order Evidence. *Philosophy and Phenomenological Research*, 81(1), 185–215.
- Conee, Earl and Richard Feldman (2001). Internalism Defended. *American Philosophical Quarterly*, 38(1), 1–18.
- Craig, Edward (1990). *Knowledge and the State of Nature*. Oxford University Press.
- Divers, John (2010). Modal Commitments. In Bob Hale and Aviv Hoffman (Eds.), *Modality: Metaphysics, Logic and Epistemology* (189–219). Oxford University Press.
- Dogramaci, Sinan (2010). Knowledge of Validity. *Noûs*, 44(3), 403–432.
- Dogramaci, Sinan (2012). Reverse Engineering Epistemic Rationality. *Philosophy and Phenomenological Research*, 84(3), 513–530.
- Dogramaci, Sinan (in press). Communist Conventions for Deductive Reasoning. *Noûs*, 49(4), 776–799.

- Dummett, Michael (1991). *The Logical Basis of Metaphysics*. Harvard University Press.
- Evans, Jonathan St. B. T. (2003). In Two Minds: Dual Process Accounts of Reasoning. *Trends in Cognitive Science*, 7(10), 454–459.
- Field, Hartry (2000). A Priority as an Evaluative Notion. In Paul Boghossian and Christopher Peacocke (Eds.), *New Essay on the A Priori* (117–149). Oxford University Press.
- Fricker, Miranda (2007). *Epistemic Injustice: Power and the Ethics of Knowing*. Oxford University Press.
- Fricker, Miranda (in press). What's the Point of Blame? A Paradigm Based Explanation. *Noûs*.
- Gilovich, Thomas, Dale Griffin, and Daniel Kahneman (Eds.) (2002). *Heuristics and Biases: The Psychology of Intuitive Judgment*. Cambridge University Press.
- Goldman, Alvin (1999). Internalism Exposed. *The Journal of Philosophy*, 96(6), 271–293.
- Goldman, Alvin (2009). Internalism, Externalism, and the Architecture of Justification. *The Journal of Philosophy*, 106(6), 309–338
- Greco, John (2005). Justification Is Not Internal. In Mattias Steup and Ernest Sosa (Eds.), *Contemporary Debates in Epistemology* (257–269). Blackwell.
- Harman, Gilbert (1986). *Change in View*. MIT Press.
- Henderson, David (2009). Motivated Contextualism. *Philosophical Studies*, 142(1), 119–131.
- Henderson, David (2011). Gate-Keeping Contextualism. *Episteme*, 8(1), 83–98.
- Henderson, David and Terence Horgan (in press). What's the Point? In John Greco and David Henderson (Eds.), *Epistemic Evaluation: Purposeful Epistemology*.
- Hieronymi, Pamela (2006). Controlling Attitudes. *Pacific Philosophical Quarterly*, 87(1), 45–74.
- Hieronymi, Pamela (2008). Responsibility for Believing. *Synthese*, 161(3), 357–373.
- Huemer, Michael (1999). The Problem of Memory Knowledge. *Pacific Philosophical Quarterly*, 80(4), 346–357.
- Johnston, Mark (2006). Better than Mere Knowledge? The Function of Sensory Awareness. In Tamar Gendler and John Hawthorne (Eds.), *Perceptual Experience* (260–290). Oxford University Press.
- Johnston, Mark (2011). On a Neglected Epistemic Virtue. *Philosophical Issues*, 21(1), 165–218.
- Kahneman, Daniel (2003). A Perspective on Judgment and Choice: Mapping Bounded Rationality. *American Psychologist*, 58(9), 697–720.
- Kelly, Thomas (2003). Epistemic Rationality as Instrumental Rationality: A Critique. *Philosophy and Phenomenological Research*, 66(3), 612–640.

- Kment, Boris (2006a). Counterfactuals and the Analysis of Necessity. *Philosophical Perspectives*, 20(1), 237–302.
- Kment, Boris (2006b). Counterfactuals and Explanation. *Mind*, 115(458), 261–309.
- Kunda, Ziva (1990). The Case for Motivated Reasoning. *Psychological Bulletin*, 108(3), 480–498.
- Lackey, Jennifer (2008). *Learning from Words: Testimony as a Source of Knowledge*. Oxford University Press.
- Lasonen-Aarnio, Maria (2014). Higher-Order Evidence and the Limits of Defeat. *Philosophy and Phenomenological Research*, 88(2), 314–345.
- Lewis, David (1971). Immodest Inductive Methods. *Philosophy of Science*, 38(1), 54–63.
- Lewis, David (1974). Radical Interpretation. *Synthese*, 27(3-4), 331–344.
- McGrath, Matthew (2007). Memory and Epistemic Conservatism. *Synthese*, 157(1), 1–24.
- Mercier, Hugo and Dan Sperber (2011). Why Do Humans Reason? Arguments for an Argumentative Theory. *Behavioral and Brain Sciences*, 34(2), 57–74.
- Michaelian, Kourken (2013). The Evolution of Testimony: Receiver Vigilance, Speaker Honesty and the Reliability of Communication. *Episteme*, 10(1), 37–59.
- Molden, Daniel C. and E. Tory Higgins (2005). Motivated Thinking. In Holyoak and Morrison (Eds.), *The Cambridge Handbook of Thinking and Reasoning* (295–320). Cambridge University Press.
- Nagel, Thomas (1976/79). Moral Luck. In *Mortal Questions* (24–38). Cambridge University Press. (Reprinted from *Proceedings of the Aristotelian Society, Supplementary Volume 50*, 137–151.)
- Naylor, Andrew (2015). Justification and Forgetting. *Pacific Philosophical Quarterly*, 96(3), 372–391.
- Owens, David (2000). *Reason without Freedom: The Problem of Epistemic Normativity*. Routledge.
- Parsons, Charles (1990/2014). Genetic Explanation in *The Roots of Reference*. In *Philosophy of Mathematics in the Twentieth Century* (220–242). Harvard University Press. (Reprinted from *Perspectives on Quine*, 273–290, Robert Barrett and Roger Gibson, Eds., 1990, Blackwell.)
- Plantinga, Alvin (1993). *Warrant and Proper Function*. Oxford University Press.
- Pollock, John and Joseph Cruz (1999). *Contemporary Theories of Knowledge* (2nd ed.). Rowman & Littlefield. (First edition published in 1986.)
- Pronin, Emily, Daniel Lin, and Lee Ross (2002). The Bias Blind Spot: Perceptions of Bias in Self Versus Others. *Personality and Social Psychology Bulletin*, 28(3), 369–381.
- Quine, Willard V. (1986). *Philosophy of Logic* (2nd ed.). Harvard University Press. (First edition published in 1970.)



- Reynolds, Steven (2002). Testimony, Knowledge, and Epistemic Goals. *Philosophical Studies*, 110(2), 139–161.
- Rosen, Gideon (2004). Skepticism about Moral Responsibility. *Philosophical Perspectives*, 18(1), 295–313.
- Salmon, Wesley (1966). *The Foundations of Scientific Inference*. University of Pittsburgh Press.
- Schoer, Robert (2008). Memory Foundationalism and Unforgotten Carelessness. *Pacific Philosophical Quarterly*, 89(1), 74–85.
- Senor, Thomas (1993). Internalistic Foundationalism and the Justification of Memory Belief. *Synthese*, 94(3), 453–476.
- Sloman, Steven (1996). The Empirical Case for Two Systems of Reasoning. *Psychological Bulletin*, 119(1), 3–22.
- Sperber, Dan (2013). Speakers Are Honest Because Hearers are Vigilant: Reply to Kourken Michaelian. *Episteme*, 10(1), 61–71.
- Sperber, Dan and Hugo Mercier (2012). Reasoning as Social Competence. In Hélèn Landemore and Jon Elster (Eds.), *Collective Wisdom: Principles and Mechanisms* (368–392). Cambridge University Press.
- Sperber, Dan, Fabrice Clément, Christophe Heintz, Olivier Mascaro, Hugo Mercier, Gloria Origgi, and Deirdre Wilson (2010). Epistemic Vigilance. *Mind and Language*, 25(4), 359–393.
- Stalnaker, Robert (1984). *Inquiry*. MIT Press.
- Stanovich, Keith (1999). *Who Is Rational? Studies in Individual Differences in Reasoning*. Erlbaum.
- Tversky, Amos and Daniel Kahneman (1973). Availability: A Heuristic for Judging Frequency and Probability. *Cognitive Psychology*, 5(2), 207–232.
- Weatherson, Brian (2010). Do Judgments Screen Evidence? Manuscript. Retrieved from <http://www.brian.weatherson.org/JSE.pdf>
- Williams, Bernard (2002). *Truth and Truthfulness: An Essay in Genealogy*. Princeton University Press.
- Williamson, Timothy (2008). *The Philosophy of Philosophy*. Blackwell.