Chapter 7
Metacognition As Evidence for Evidentialism

Matthew Frise

Abstract Metacognition is the monitoring and controlling of cognitive processes. I examine the role of metacognition in 'ordinary retrieval cases', cases in which it is intuitive that via recollection the subject has a justified belief. Drawing on psychological research on metacognition, I argue that evidentialism has a unique, accurate prediction in each ordinary retrieval case: the subject has evidence for the proposition she justifiedly believes. But, I argue, process reliabilism has no unique, accurate predictions in these cases. I conclude that ordinary retrieval cases better support evidentialism than process reliabilism. This conclusion challenges several common assumptions. One is that non-evidentialism alone allows for a naturalized epistemology, i.e., an epistemology that is fully in accordance with scientific research and methodology. Another is that process reliabilism fares much better than evidentialism in the epistemology of memory.

Keywords Metacognition · Memory · Naturalized epistemology · Ordinary retrieval · Reliabilism

7.1 Introduction

Evidentialism roughly is the view that an attitude for a subject toward a proposition is justified just when the attitude fits the subject's total evidence.[1] Many philosophers think that a chief rival to evidentialism is process reliabilism (hereafter *reliabilism*). Reliabilism states roughly that a belief is justified just in case it results from a reliable belief formation process, that is, a process that tends to yield

[1] See Feldman and Conee (1985).

M. Frise (✉)
Department of Philosophy, Santa Clara University, Santa Clara, CA, USA

K. McCain (ed.), *Believing in Accordance with the Evidence*, Synthese Library 398, https://doi.org/10.1007/978-3-319-95993-1_7

true beliefs rather than false beliefs.[2] In this paper I defend an argument favoring evidentialism over reliabilism:

Retrieval Argument

P1. Evidentialism has a unique, accurate prediction in ordinary retrieval cases.
P2. It is not the case that reliabilism has a unique, accurate prediction in ordinary retrieval cases.
P3. If in cases *X*, H1 but not H2 has a unique, accurate prediction, then cases *X* support H1 better than H2.
C. Ordinary retrieval cases support evidentialism better than reliabilism.

Let's clarify terms. A *prediction*, here, is a proposition that a theory (at least when paired with auxiliary hypotheses) entails, and yet this theory was not designed to entail it. A *unique* prediction is a proposition that one theory (and its auxiliary hypotheses) entails but which a specified rival theory does not entail. In Sect. 7.2 I will explain exactly what an *ordinary retrieval case* is. For now, think of it as a case in which it is intuitive to any non-skeptical epistemologist that a subject, upon recollecting information related to *p*, has a belief that *p* that is memorially justified. Not all cases of recollection fit this description. In cases where the recollecting subject has forgotten a defeater for *p*, for example, it is controversial whether the subject's belief that *p* is memorially justified.[3] But *all* non-skeptical epistemologists want their theory of justification to imply that the subject's belief in an ordinary retrieval case is justified.

The conclusion of the Retrieval Argument is modest. It does not propose that, all things considered, we should endorse evidentialism over reliabilism. It proposes that certain cases count in favor of evidentialism rather than reliabilism. I will defend P1 and P2 by looking at research on the role of metacognition during memory retrieval. First (in Sect. 7.2) I discuss this role, and then (in Sects. 7.3 and 7.4) I support each premise. I do not defend P3 here, as it is uncontroversial.[4]

But why should we care about my argument, given its modest conclusion? Here are three reasons. The reasons reveal that a successful defense of even *just* P1 or P2 is significant. First, allegedly, externalism is much friendlier to a naturalized epistemology than internalism is. Internalism is the view that epistemic justification supervenes on the mental; no feature in a subject's environment affects

[2]See Goldman (1979). For two reasons, evidentialism and reliabilism are not in fact direct rivals. First, they theorize about different things. Evidentialism states conditions that justify a subject in having a doxastic attitude (propositional justification), and reliabilism states conditions in which a subject's doxastic attitude is justified (doxastic justification). With supplements, however, each does state conditions about both propositional and doxastic justification. Second, once supplemented, they can remain compatible (see Sect. 7.4). For simplicity, I take evidentialism and reliabilism to be direct rivals here.

[3]Feldman (2005: 282–3) and McGrath (2007: 4) argue that there can be memorial justification in such cases. Annis (1980: 325–6), Goldman (2009: 324), and Greco (2005: 266–8) argue otherwise.

[4]P3 follows from strong predictivism, from weak predictivism, and from the likelihood principle. See Harker (2013) and McCain (2012) for discussion of weak predictivism in epistemology.

her justification without affecting her mental life.[5] All justifying features are mental. Externalism is the denial of internalism. The sort of evidentialism I support here is internalist, while reliabilism is externalist.

According to Hilary Kornblith (2007: 51) certain data from cognitive psychology in particular threaten internalism. Even some philosophers who try to show that there is *some* affinity between internalism and naturalized epistemology, grant that there is this threat.[6] According to John Greco (2010: 61), the data threaten evidentialism specifically. Also, according to Kornblith (2007: 44), the typical manner of constructing externalist theories of justification is "thoroughly naturalistic" (cf. Kitcher (1992: 3)); externalist methodology resembles our investigation of natural kinds, in that it investigates not merely our concept of justified belief but the characteristics underlying actual beliefs that are clearly justified. Since Kornblith cites this credential on behalf of externalism, presumably he thinks internalism lacks it. Alston (2004: 50) goes so far as to claim that the "rise of externalism" is in part explained by its naturalistic methodology. If the Retrieval Argument succeeds, however, important data from cognitive psychology support a form of internalism over a leading externalist theory. What's more, externalism's naturalistic methodology may help the Retrieval Argument succeed. Philosophers in favor of naturalizing epistemology will have less reason to prefer externalism over internalism.

Second, allegedly, internalism and evidentialism fare poorly in the epistemology of memory, while externalism and reliabilism do well.[7] Joëlle Proust (2013: Chap. 9) uses data on metacognition in memory in particular to support this allegation, and her arguments have actually influenced some psychologists.[8] Other philosophers and psychologists, when discussing metacognition, simply assume that some form of externalism is correct.[9] My support for P1 helps undermine the allegation against internalism and evidentialism. Also, the Retrieval Argument suggests that philosophers and psychologists should take internalism more seriously when exploring research on metacognition, and that this research in some cases supports internalism better. Internalism in the epistemology of memory becomes safer.

Third, my support for my argument importantly develops evidentialism and reliabilism. Conee and Feldman (2008: 93) count memory as a source of evidence,

[5] Conee and Feldman (2001). Some internalists would add that all justifiers are specially accessible by their subjects. The variety of evidentialism I defend here is compatible with, but does not entail, this addition.

[6] See, e.g., Wheeler and Pereira (2008: 317). Feldman (1999), however, argues that data from cognitive psychology is much less important to epistemological theorizing than many philosophers suppose.

[7] See Bernecker (2008, 2010), Goldman (1999, 2009, 2011), Greco (2005), Plantinga (1993) and Senor (2010). Cf. Frise (2017). For replies see Frise (2015, 2018) and Conee and Feldman (2001).

[8] Proust's arguments, for example, have influenced Koriat and Adiv (2012: 1611).

[9] For philosophers, see Dokic (2014) and Michaelian (2012). For psychologists, see Reber and Unkelbach (2010).

but note that "Details about [it] and general theories about how [it works] would be extremely valuable." They say memory provides justification "only when a suitable background is in place. Exactly what constitutes that background is a difficult matter we will not attempt to resolve here. Whatever that background is, it is a matter of evidence". My defense of P1 helps complete evidentialism, theorizing about how memory works and about what this background consists partly in.

And my defense of P2 uncovers general problems for reliabilism. Reliabilism may lead to a kind of skepticism. Further, reliabilism's overall testability turns out to be surprisingly limited. Reliabilism has not in fact already gathered all the trophies in the epistemology of memory.

7.2 Metacognition in Memory

An ordinary retrieval case is one in which it is uncontroversial that a subject justifiedly believes that *p* after having a recollective experience related to *p*. Additionally, this justification is memorial rather than, say, perceptual or testimonial. There are different accounts of why there is memorial justification in these cases.[10] I remain neutral on them. Since there is memory justification outside of ordinary retrieval (e.g. for some non-occurrent beliefs), I am not commenting on memory justification *simpliciter* here.

In order to see what evidentialism and reliabilism do and don't *accurately* predict in ordinary retrieval cases, we should first see what these cases are like. Suppose Smith, a typical American adult, is asked, "Who was the first postmaster general of the United States?", and Smith thinks and has certain experiences, and then reports *p,* namely, that Benjamin Franklin was the first postmaster general of the United States. Let all of this happen in a fairly normal way, such that we find it intuitive that Smith believes that *p* justifiedly. What of interest occurred between the asking and the reporting? To answer this, we needn't merely appeal to armchair intuitions or personal experience. We can look at psychological research on metacognition in memory.

Metacognition is the monitoring and controlling of cognitive processes.[11] Cognition allows us to read the road signs outside the mind. Metacognition allows us to decipher some signs within. In an ordinary retrieval case, an information-producing cognitive mechanism (unsurprisingly) produces information, and both the *information* and its *production* are monitored. This monitoring is typically unconscious but becomes conscious in certain circumstances (when, for example,

[10] Annis (1980), Bernecker (2008), and Goldman (1999, 2009, 2011) say memory merely *preserves* the justification from the past. Audi (1995), Conee and Feldman (2011), and Huemer (1999) say recollective experience sometimes *generates* some justification.

[11] On the psychological claims below, see Koriat (2002), Koriat and Helstrup (2007), and Unkelbach (2007). Arango-Muñoz (2013a, b), Arango-Muñoz and Michaelian (2014), Michaelian (2012), Nagel (manuscript), and Proust (2013) guide my interpretation of the psychology.

there is any of a variety of difficulties in processing). The monitoring involves and gives rise to an *epistemic feeling*.[12] It's controversial just what an epistemic feeling is. At minimum, it is a phenomenal, affective, non-emotional experience with intentional content. All epistemic feelings have these features, though other mental states might have them all too. What is special about an epistemic feeling is that it gives feedback having to do with cognitive processing. Examples of such feelings include feelings of knowing, of familiarity, of uncertainty, and of forgetting. The type of epistemic feeling elicited is determined by the features detected in both the information and its production. Detecting scant or undetailed information is more likely to elicit a feeling of uncertainty, while detecting a glut of detailed information is more likely to elicit the feeling of knowing—even before that detailed information is consciously accessed.

An endorsement mechanism then evaluates the type of feeling, the features detected in monitoring, and certain of the subject's background beliefs. In light of the evaluation, the endorsement mechanism controls the information-processing. This control either terminates or permits the retrieving of information. Control can initiate a different strategy for accessing the target information (e.g., using a different heuristic, looking the information up via an external source). The endorsement mechanism controls whether the subject endorses (i.e., occurrently believes) or suspends judgment regarding the retrieved information.[13] Typically the subject's epistemic feelings determine the subject's confidence in anything that becomes endorsed. A feeling of knowing correlates with higher confidence, while a mild feeling of uncertainty does not.

Smith's ordinary retrieval, for example, begins with unconscious information-production. Monitoring this information results in his having an epistemic feeling, like the feeling of knowing. This feeling precedes his endorsing *p*, and his producing *p* consciously. Next, Smith experiences *fluent* retrieval of *p*. That is, he might experience retrieving *p* relatively quickly; or, he might experience retrieving information corroborating *p*; or, *p* might persist for a relatively long while or occur frequently in his thoughts. Or, some combination might occur. Smith will have learned to interpret (automatically and unreflectively) this experience of fluently retrieving *p* as *p*'s being familiar. As a result of monitoring, an endorsement mechanism will exert control. Smith will endorse *p* and cease his inquiry, and his confidence in *p* will be high, given the high fluency of his retrieval experience. This completes his ordinary retrieval.

[12] Alternatively dubbed a *noetic* feeling (Proust 2013) and *metacognitive* feeling (Arango-Muñoz 2013b).

[13] Michaelian (2012: 288–90) assumes that one of these propositional attitudes is thereby *formed*. But it could be that the attitude was standing and just becomes occurrent.

7.3 Evidentialism

I have described an ordinary retrieval case. What might evidentialism accurately predict here? The defense of P1 begins with answering this. But answering requires us to get clearer on what evidentialism entails in *any* case of justified belief. According to evidentialism, if the justified attitude for S toward *p* is belief, then S's evidence supports *p*. Belief is the justified attitude for S toward *p* in an ordinary retrieval case. So, evidentialism entails that S has evidence for *p*. So long as evidentialism was not designed to entail this in ordinary retrieval cases, it counts as a prediction.

But what is *evidence*, and what is it for something to be evidence *for p*? Different versions of evidentialism answer these questions differently. On the version I discuss here—*explanationist evidentialism*—S's evidence includes S's experiences. The propositions supported by the evidence are the ones that are part of the best explanation available to S for why S has that evidence. For instance, for a typical adult, a reddish visual experience typically is for her evidence that something is red. This is because, on the best explanation of her experience available to her, something is red. She need not have assessed, or even ever thought about, this explanation or any other. It just must be the best available to her. I won't defend a theory of *availability*. I'll assume simply that *p* is part of the best explanation available to S for why she has certain evidence if the following is true: S is disposed to have a seeming that *p* is part of the best answer as to why she has that evidence.[14]

Now what, if anything, does evidentialism *accurately* predict in ordinary retrieval cases? It predicts that Smith, for instance, will have evidence for *p* (i.e., that Benjamin Franklin was the first postmaster general of the United States). More generally:

> *Candidate 0*. In each ordinary retrieval case, the proposition justifiedly believed by the subject is part of the best explanation available to her for why she has her experiences.

Is this prediction accurate? Some philosophers would suggest not. Plantinga (1993: 62–4) considers our potential evidence for our memory beliefs, slipping back and forth between talking about 'present phenomena', 'phenomenal imagery', and 'beliefs about the present'. This evidence is either too rare or feeble to be what actually justifies our memory beliefs. He (1993: 188) concludes "There is nothing we can sensibly think of as evidence on the basis of which [a] memory belief is formed," because he seems to think *we have no* justifying evidence for the content of our memory beliefs.[15] Greco (2010: 61) concurs (cf. Bernecker (2010:

[14] McCain (2014: 65–70) defends the assumption about availability. Cf. Conee and Feldman (2008: 97–98).

[15] Apparently Plantinga assumes that a memory belief is based on evidence only if it is currently formed on the basis of conscious evidence. This overlooks the possibility that these memory beliefs were formed in the past and that currently they are just activated, and the possibility that their evidential bases are mental but non-conscious.

73)). If Plantinga and Greco are right, and if explanationist evidentialism correctly characterizes evidential support, then Candidate 0 is inaccurate. The subject in an ordinary retrieval case lacks evidence (from memory, at least) for her belief, since it is a memory belief.

However, Plantinga's survey of the potential evidence in ordinary retrieval cases is not exhaustive. He does not consider all 'present phenomena'. As I interpret the research on metacognition, Candidate 0 is accurate. Consider Smith. He has evidence, and it is evidence for *p*. His evidence includes (a) his epistemic feelings that bear on *p* (e.g., his feeling of knowing), (b) his experience of fluently retrieving *p*, and (c) his experience of automatically interpreting the fluently retrieved information as familiar.

Here is why (a), (b), and (c) are evidence for *p*: on the best explanation available to Smith of why these phenomena obtain, *p* is true. When asked, "who was the first postmaster general of the United States?", Smith could have experienced fluent retrieval of indefinitely many propositions other than *p* and had an associated feeling of knowing. Or, Smith could have retrieved nothing at all. On the best available explanation to Smith for why he experienced fluent retrieval or had a feeling of knowing regarding *p in particular*, Smith once learned *p* or some nontrivial support for *p*. Smith's memory supports this. As far as Smith is able to tell, what his feelings of knowing indicate is often correct and reasonable, not contentious. Also, that the feeling of knowing is a guide to the truth coheres well with Smith's other experiences, and with the fact that those experiences do not, from his perspective, tend to mislead.

A proposition that a subject fluently retrieves has likely been processed by that subject before. The fluency results from a kind of practice at processing. All else being equal, a previously processed proposition on a matter is more likely true than an incompatible unfamiliar one. The best available explanation of Smith's fluently retrieving *p* includes one or several previous representations to him of *p* as true—perhaps initially via testimony, then via further testimony, and then via recollection, and so on. There is only one true proposition about who the first postmaster general was, and indefinitely many falsehoods. Other things being equal, a proposition represented on multiple occasions as true is more likely true than false, in part since (roughly) a truth on the matter is more likely to be reencountered than a given falsehood is. Any number of falsehoods could be encountered, and so each is less likely to be *re*encountered than the truth is.[16]

And, part of the best available explanation of Smith's experience of *automatically* interpreting fluency as familiarity is that he has learned, perhaps unreflectively, this normally gets at the truth; whatever Smith fluently retrieves is likely true, and familiarity flags that truth-connection for Smith. Given what Smith can recollect and that Smith can tell that he is fairly normal and rational, he has reason to believe that his automatically interpreting fluency as familiarity results from good habituation. So on the best available explanation of Smith's experience of the automatic interpretation, *p* is true, since *p* feels familiar.

[16]Cf. Reber and Unkelbach (2010).

The best explanations available to Smith of (a), (b), and (c)—individually, but also together—include *p*. They are best because they are more parsimonious or explanatorily powerful than the alternatives omitting *p*'s truth. Here are some alternatives: Smith feels he knows any proposition that comes to mind. Smith is disposed to have feelings of knowing toward propositions he typically never learned in the past. Smith fluently retrieves propositions independently of what he has learned or had reason to believe. Smith at random automatically interprets phenomena as familiarity. He never learned, as a way of getting the truth, to interpret fluency as familiarity. As they stand, these alternative kinds of explanations are ad hoc and not very powerful. They suggest that subjects in ordinary retrieval cases are typically misremembering. They leave it mysterious to Smith why he has managed to survive, to live a normal life, to cooperate with others and agree with them about the past, and to have a highly coherent set of experiences overall. They incline us to doubt that Smith's belief is justified even though, by stipulation of his case being ordinary retrieval, it is justified. The alternative explanations can become more explanatorily powerful only by sacrificing simplicity. They can posit ad hoc reasons for his surviving, cooperating, and experiencing coherently. But the reasons bloat the explanation. The commonsensical explanations that include the truth of *p* are better. So, Smith has evidence for *p*. Again, Smith need not have worked out how *p* is part of what best explains (a), (b), and (c). This best explanation simply must be available to him.[17]

More could be said in direct support of my claim that the best available explanation of (a), (b), and (c) includes *p*, but this sketch will suffice for now. If, as I claim, evidentialism accurately predicts Candidate 0, then we are halfway to establishing P1. To establish P1 we now just need to show that this prediction is unique, i.e., that reliabilism does not share it. The next section considers reliabilism's predictions.

First, a worry. Joëlle Proust doubts that a view she calls "internalism" explains how metacognition could play a justificatory role. Yet I've claimed that metacognition plays this role on an evidentialist internalism. Proust and I pick out importantly different views with "internalism", but it's still worth deflating the doubt. She (2013, 198–200) correctly notes that a subject's environment and past largely influence whether her epistemic feelings are reliable. She (2013, 200) says: "One can thus conclude that the existence and reliability of epistemic feelings *supervene in part on* the existence and quality of the feedback provided. Therefore, the internalist

[17]For inchoate explanatory theories of memorial support, see Harman (1973: 189) and Peacocke (1986: 163–4). Jennifer Nagel (manuscript) argues that something like (c)—the interpretation of fluency as familiarity—is available to internalist accounts of the justification of "trivia beliefs". She says (manuscript: 2) a belief is a trivia belief "if and only if (1) its origin lies in testimony from a source whose identity is now unknown to the subject, and (2) the subject lacks topically related auxiliary beliefs that would suffice to support the target belief". My proposals go well beyond Nagel's. I discuss justification in ordinary retrieval cases, which often involve non-trivia beliefs. Also, Nagel does not argue that (a) or (b) helps justify, and she (manuscript: 19) thinks (c) itself justifies only "weakly". And, I state in detail why (c), on explanationist evidentialism, helps account for the relevant justification. Finally, I show that research on metacognition supports an internalist view *better* than a main externalist rival.

case for epistemic feelings as a source of epistemic intuition considerably loses in explanatory force and credibility." She seems to mean that having relevant epistemic feelings is insufficient for having justification. Rather, epistemic feelings justify only in environments where they are reliable. So, she concludes that metacognition is uncongenial to internalism.

An unstated premise here is that any justifier is reliable. That is why, on Proust's view, epistemic feelings do not justify in environments where they are unreliable. If there were reason to accept her premise, her conclusion would be hard to deny. However, it would then be unremarkable that metacognition is uncongenial to internalism. This is because, if her unstated premise were true, then internalism would be false. Internalism would tell the wrong sort of story about justification, since it omits environmental reliability constraints. There would be nothing special about internalism incorrectly explaining justification from metacognition. Now, one thing we should not do when evaluating how internalism and metacognition fit is assume that internalism is false. Proust's evaluation requires that very assumption. So, we may set it aside.

7.4 Reliabilism

I will examine some leading candidates for what reliabilism might accurately, uniquely predict in ordinary retrieval cases. We will find nothing suitable. This will sufficiently support P2. I will then examine whether reliabilism predicts Candidate 0. We will find it does not. This will complete the defense of P1. Along the way we will uncover some general problems for reliabilism. I am silent on many details of reliabilism, so that the Retrieval Argument applies to any version of it.

It may seem obvious that reliabilism predicts:

> *Candidate 1*. In each ordinary retrieval case the relevant process that forms the subject's justified belief is reliable.

The justified belief is formed by indefinitely many types of processes, and some of these processes are reliable and some aren't. The *relevant* process is the one whose reliability determines whether the particular belief is justified. But reliabilism does not predict Candidate 1. For there is a wrinkle to reliabilism.

Reliabilists distinguish belief-dependent and belief-independent belief formation processes. A belief-dependent process (e.g., an inferential process) includes beliefs among its inputs. A belief-independent process (e.g., a basic perceptual process) does not. Many belief-dependent processes are unreliable, yet they still have the virtue of being *conditionally* reliable—they satisfy the following:

> CR1. A process R is conditionally reliable *iff* R mostly produces true beliefs when all of R's belief inputs are true.[18]

[18]Comesaña (2010: 577), Goldman (2011: 278n.20), and Lyons (2013: 9) endorse CR1.

And reliabilists hold that the output of a belief-dependent conditionally reliable belief formation process is justified if the belief inputs to that process are justified. So, for example, belief in the conclusion of some reasoning is justified if all the premises are justifiedly believed and the type of reasoning typically yields true beliefs when the premises are true.

Now, in a typical adult human, if a process involves memory and metacognition in the formation of a belief that *p*, that process is belief-dependent. It has belief inputs. These include past beliefs with content relevantly similar to *p* or bearing on *p*, beliefs about how memory works, about memory experience, about epistemic feelings, about the feeling of familiarity, and so on. The belief-dependent nature of memory processing actually helped *inspire* Goldman's (1979: 13) notion of conditional reliability, shaping his original statement of reliabilism. So, reliabilism doesn't predict Candidate 1. Consider instead:

> *Candidate 2*. In each ordinary retrieval case the relevant process that forms the subject's justified belief is *conditionally* reliable.

If reliabilism predicts Candidate 2 rather than Candidate 1, it has an asset. In particular, on Candidate 2, massive *perceptual* deception needn't threaten memory justification. Since perception feeds beliefs into memory, perceptual deception can make memory unreliable. Still, memory can remain *conditionally* reliable and able to justify. Memory justification is securer if reliabilism predicts Candidate 2.[19]

If reliabilism predicts Candidate 2, it does so uniquely. But is Candidate 2 accurate? In order to answer this question we must address two others: what determines which process is relevant? And, how do we confirm that the relevant process is conditionally reliable—that it satisfies CR1? The first of these questions introduces reliabilism's dreaded generality problem: we need a principled way to identify the relevant process that forms any particular belief, so that we can test reliabilism's implications about justification in each case against our intuitive judgments.[20] While no adequate solution to this problem has been defended, perhaps one exists. Still, I point out two main difficulties with predicting and confirming Candidate 2.

Point 1: Developing a predicted interpretation of Candidate 2 is not only challenging, but also methodologically *non*-naturalistic in a way. Here is why. In order to confirm that Candidate 2 is accurate, we must interpret it as specifying a particular process type as relevant in each type of ordinary retrieval case, so that we can confirm the conditional reliability of that process. A reliabilism that solves the generality problem entails a complete interpretation of Candidate 2. And we must be able to confirm that Candidate 2, so interpreted, is accurate.

[19] Goldman's (1979: 14, 2011: 278) reliabilism predicts Candidate 2. Lyons (2009: 177) however develops an untraditional reliabilism that predicts Candidate 1 instead. Unfortunately, his view robs reliabilism of the asset I mention above. Since Lyons' (2013) reliabilism keeps with tradition, however, I draw on that work below.

[20] See Conee and Feldman (1998) and Feldman (1985).

In the interests of making a prediction, the solution must not be *designed* to entail this complete interpretation of Candidate 2. (This makes solving the generality problem even harder.) But without this design we eschew a naturalistic methodology! Recall that, according to Kornblith, externalist methodology is naturalistic in that it investigates the characteristics underlying actual, clearly justified beliefs, and then uses some of the observed characteristics to construct a theory of justification. The theory is designed to entail that justified beliefs have the observed characteristics. It follows that the theory does not predict that the beliefs have the characteristics. If the characteristics were selected via examining ordinary retrieval cases, then the theory of justification does not make predictions about these cases.

Of course, it could be that a fairly general process type is relevant in ordinary retrieval cases, and that we can identify this process by looking at cases other than ordinary retrieval. But this is unlikely. It's not as if ordinary retrieval cases are simply instances of, say, *carefully believing*. Many beliefs in ordinary retrieval cases involve automatic endorsement, leaving no room for care. Beliefs in ordinary retrieval cases seem to constitute a special class that is not fruitfully subsumed under another.

In short, a reliabilist theory that solves the generality problem *by* examining actual justified beliefs will not predict an interpretation of Candidate 2, and thus will help establish P2. A reliabilist theory that solves the generality problem *without* examining actual justified beliefs loses some naturalistic credentials. There is tension between predicting Candidate 2 and pursuing certain naturalistic methodology. This is an unsettling result for the many reliabilists who value their theory's alleged special affinity with that methodology.

Now, supposing we can identify the relevant process in each ordinary retrieval case, which belief outputs can we look at in order to determine whether that process is conditionally reliable, and so assess Candidate 2's accuracy? Reliabilism's best hope is that the data from metacognition research supports the conditional reliability of each relevant process.

Point 2: Yet the data does not support this. Here is why. According to CR1, a conditionally reliable process is one that generally produces true beliefs when all belief inputs are true. Consequently, if a token belief-dependent process has a single false belief input, its true outputs are not evidence of the process' conditional reliability. Only outputs of processes where *all* belief inputs are true could be evidence of conditional reliability. So, only those token processes could be evidence for Candidate 2's accuracy.

In order to check whether there is this evidence, it would help to have a sense of what counts as a belief input. Few reliabilists offer guidance. When Goldman (1979: 13–14) originally introduces the ideas of conditional reliability and belief-dependent processes, he gives two examples of belief inputs: a stored memory belief, and a premise in an inference. But he does not characterize inputs in general. He does say (1979: 11) that "when we say that belief is caused by a given process . . . we may interpret that to mean that it is caused by the particular inputs to the process." However, this states only that all inputs are causes of the output. It doesn't state which causal beliefs count as belief inputs.

Jack Lyons (2013: 12) characterizes belief inputs more explicitly: any belief, even a tacit belief, on which the output belief is causally or counterfactually dependent.[21] That is, suppose a belief-dependent belief formation process yields a belief that *q* for S, and the process wouldn't have if S had not (tacitly) believed *r*. S's belief that *r* counts as an input to the process that formed S's belief that *q*. Note that this counts an extraordinary number of beliefs as inputs in cases where there are any. If, for example, I didn't believe I exist, I wouldn't believe I am sitting. So my belief that I exist counts as an input to the process that formed my belief that I am sitting. And, for example, I would not believe that Ms. Tardy will be late to the party if I didn't (tacitly) believe that she is not already there, that she will be coming to the party at all, that she is still alive, that the party will continue, that I exist, etc.

Unfortunately for reliabilism, a non-trivial percent of our (tacit) beliefs are false. What's more, what we retrieve depends often on one or more particular false beliefs, namely, beliefs associated with various memory biases.[22] Given this and Lyons' extremely permissive view about which beliefs count as inputs, it is overwhelmingly likely that at least one input to any given token belief-dependent process is false. So the truth-value of the output beliefs in ordinary retrieval cases will not verify Candidate 2. On CR1, reliabilism does not imply that a conditionally reliable process with a false belief input still tends to have true outputs.

It might seem that we have *inductive* evidence of Candidate 2's accuracy. In observed ordinary retrieval cases the beliefs tend to be true, even though they typically result from a process with some false belief input. That gives us reason to suppose that in the unobserved ordinary retrieval cases—including those in which all belief inputs are true—the belief outputs tend to be true.

Perhaps it is *ordinarily* reasonable to conclude via induction that a process with mostly true outputs and with some false inputs is conditionally reliable. But the relevant process in ordinary retrieval cases is one that produces a belief by significantly altering the contents of its inputs, including its belief inputs. Memory alters these inputs considerably at three stages, and this often results in changes in truth-value (see Frise (2018) and Michaelian (2011)). Yet the alteration helps memory yield true beliefs. It's not at all clear what will happen if all the belief inputs are true. Are the output contents nonetheless adjustments of the input contents? If so, then the process may very well not tend to get the truth. Also, it could be that our false beliefs associated with our memory biases help us to get at the truth. These beliefs are typical inputs. Eliminating them may notably lower the ratio of true

[21] Cf. Lyons (2013: 28) and Conee and Feldman (1998: 26–7, n.13). I see no non-*ad hoc* reason to restrict input beliefs to those held *by the subject*. S_1's forming a belief that *p* may be causally or counterfactually dependent on S_2's belief that *q* (e.g., via testimony), and so it seems S_2's belief that *q* would count as an input to the process that formed S_1's belief that *p*. This has strange results.

[22] These beliefs concern *consistency* bias (whereby one reconstructs the past too similarly to one's view of the present), *change* bias (whereby one views oneself in the past too differently, in order to redeem an investment), *hindsight* bias (whereby one attributes present knowledge to one's past self), and *egocentric memory* bias (whereby one inflates one's present self-image by distorting one's past self-image); see Schacter (2001: Chap. 6).

outputs. So it is questionable to reason inductively about the conditional reliability of the relevant processes in ordinary retrieval. It is unclear what they would tend to produce, given all true belief inputs.

In the absence of a promising alternative view about what counts as a belief input, we have little reason to believe that Candidate 2 is accurate. Of course, we could replace CR1 with a more liberal view about conditional reliability. For example:

> CR2. A process R is conditionally reliable *iff* R mostly produces true beliefs when *at least most* of R's belief inputs are true.

On CR2, even if it is likely that some belief inputs to a token process are false, we can still confirm the conditional reliability of that process type as long as *most* belief inputs are true and the output is true. And most belief inputs in the typical ordinary retrieval case are true. Since the output is typically true, it appears that, on CR2, we have strong evidence that Candidate 2 is accurate.

CR2 seems attractive. A process that manages to be truth-conducive even while disadvantaged by false belief inputs seems at least as good as a process that is truth-conducive only when all belief inputs are true. However, CR2 does not help reliabilism. On CR2, the wrong processes will (or won't) count as conditionally reliable, and therefore capable (or incapable) of justifying. Here is just one important example.

One process that should be capable of justifying is *moderate conjunction*. This process takes five or more beliefs as inputs, but not many more, and produces a belief in the conjunction of their contents. When the inputs are S's belief that p_1, S's belief that p_2, ... S's belief that p_5, moderate conjunction produces in S a belief that (p_1 and p_2 and ... p_5). Unfortunately for reliabilism, CR2 counts moderate conjunction as conditionally unreliable. Suppose most belief inputs to moderate conjunction are true. If there are five input beliefs, at least three are true. More often than not, at least one of the remaining beliefs is false. After all, the two remaining beliefs could have several combinations of truth-values. On all but one combination, at least one belief is false. So, the output—the belief in the conjunction—will tend to be false, when most inputs are true. According to CR2 the process is conditionally unreliable, and therefore incapable of justifying. Yet, when all belief inputs are justified, moderate conjunction seems to be a paradigm of a justifying belief-dependent process of belief formation!

This may prompt us to look for an account of conditional reliability that lies between CR1 and CR2. But wherever the account lies, it faces problems. Suppose the account swings closer to CR1, and requires for conditional reliability that most outputs are true when at least 90% of belief inputs are true. As we near CR1, it becomes harder to see that actual instances of ordinary retrieval are evidence that the relevant process type is conditionally reliable. It's not clear that at least 90% of the inputs to the relevant process are true in actual cases of ordinary retrieval. So the account does not support Candidate 2's accuracy. And if the account swings closer to CR2 and selects a lower percentage, it becomes easier for moderate conjunction to fail to count as both conditionally reliable and capable of justifying, and so the account seems false. Accounts that swing toward the middle of the road and

select a percentage nearer to 75% face both problems to some extent. And accounts that concern instead the truth of all belief inputs of a certain quality—e.g., the *pertinent* ones—face a different problem. They must *predict* (entail without being so designed) which inputs have that quality, for any ordinary retrieval case. At any rate, it is significant if reliabilists must replace CR1.

In short, we have insufficient evidence that Candidate 2 is an accurate prediction. In too many ordinary retrieval cases, the process that forms the justified belief has some false belief as an input. On the best view of conditional reliability, only a process' performance, when all its belief inputs are true, matters. A general lesson here is this. To the extent that reliabilism theorizes about justification from content-modifying belief formation processes that usually have some false belief inputs, there is no clear way to confirm or to disconfirm reliabilism.

The preceding also shows that reliabilism does not even predict:

Candidate 3. Most of the justified beliefs in ordinary retrieval cases are true.

Reliabilism implies that justifying processes that involve metacognition and memory need only be conditionally reliable. So it is compatible with reliabilism that most justified beliefs in ordinary retrieval cases are false (if, e.g., most of the relevant processes producing the justified beliefs have a false belief input).

We might also consider:

Candidate 4. All belief inputs to the justified belief in an ordinary retrieval case are justified.

Reliabilism seems to predict this. On reliabilism, the output of a conditionally reliable belief-dependent process is justified when all belief inputs to that process are justified. Unfortunately for reliabilism, we have no test for Candidate 4's accuracy, not even from research on metacognition. One reason for this is our ignorance of exactly what *all* those particular belief inputs are in a given case. If Lyons' view of belief inputs is correct, in any ordinary retrieval case there are numerous (tacit) belief inputs, and we have too little information to determine that all are justified. Moreover, it seems doubtful that all the belief inputs are justified. This is because there are so many inputs, and we have a nontrivial amount of unjustified beliefs, and our beliefs associated with our memory biases appear to be regular unjustified inputs in ordinary retrieval. If this is correct, Candidate 4 seems false. And if Candidate 4 is false, reliabilism is false, since reliabilism predicts it. What's more, reliabilism leads to a kind of skepticism if Candidate 4 is false: few actual beliefs in ordinary retrieval cases are justified.

The failure of these leading candidates establishes P2. What about P1? Does reliabilism also predict Candidate 0? No defended reliabilist theory does. But one could change that. However, it is hard to see why one would, unless one simply wanted a theory with the same relevant implications that explanationist evidentialism has—a theory *designed* to entail Candidate 0. So, the theory would merely accommodate and not predict Candidate 0. Explanationist evidentialism still uniquely, accurately predicts it. P1 stands.

Reflection on Candidate 0 may raise a new doubt about P2, however. Some philosophers defend evidentialist versions of reliabilism. On these versions, all

subjects have evidence for their justified beliefs. The subject's belief is justified because it is based on certain evidence, and the process of basing that belief on this evidence is (conditionally) reliable.[23] Evidentialist reliabilism predicts:

> *Candidate 0**. In each ordinary retrieval case, the relevant process that forms the subject's justified belief is the process of basing that belief on the subject's evidence, and that process is conditionally reliable.

If Candidate 0* is accurate, then P2 is false; a form of reliabilism would have a unique accurate prediction in ordinary retrieval cases. But Candidate 0* is just an elaboration of Candidate 2, which states that in each ordinary retrieval case, the relevant process that forms the subject's justified belief is conditionally reliable. Candidate 0* specifies the relevant process. But we failed to confirm Candidate 2. The data from metacognition does not confirm the conditional reliability of any relevant process, not even the process of basing belief on the subject's evidence. Likewise, we cannot confirm Candidate 0*. So, P2 stands.

7.5 Conclusion

I conclude that ordinary retrieval cases support evidentialism better than reliabilism. This rebuts common but mistaken views about evidentialism and internalism's standing with respect to the epistemology of memory, data on metacognition, and naturalized epistemology.[24]

References

Alston, W. (2004). The 'Challenge' of externalism. In R. Shantz (Ed.), *The externalist challenge*. Berlin: de Gruyter.

Annis, D. (1980). Memory and justification. *Philosophy and Phenomenological Research, 40*(3), 324–333.

Arango-Muñoz, S. (2013a). The nature of epistemic feelings. *Philosophical Psychology, 27*(2), 193–211.

Arango-Muñoz, S. (2013b). Scaffolded memory and metacognitive feelings. *Review of Philosophy and Psychology, 4*(1), 135–152.

[23] See Comesaña (2010) and Goldman (2011). Note that, even if we could confirm a unique prediction of evidentialist reliabilism, and so the Retrieval Argument failed, we would still have a significant result: generic evidentialism (rather than explanationist evidentialism) is still better supported by ordinary retrieval cases than all non-evidentialists versions of reliabilism are. Generic evidentialism states that the justified attitude for S toward *p* is the attitude that S's evidence supports. It leaves open what counts as evidence, and leaves open how evidence supports.

[24] I thank Matthew Baddorf, Caleb Cohoe, Earl Conee, Richard Feldman, Jon Kvanvig, Kevin McCain, Kourken Michaelian, Jonathan Reibsamen, and Declan Smithies for helpful conversation and feedback on drafts of this paper.

Arango-Muñoz, S., & Michaelian, K. (2014). Epistemic feelings, epistemic emotions: Review and introduction to the focus section. *Philosophical Inquiries, 2*(1), 97–122.
Audi, R. (1995). Memorial justification. *Philosophical Topics, 23*(1), 31–45.
Bernecker, S. (2008). *The metaphysics of memory*. Dordrecht: Springer.
Bernecker, S. (2010). *Memory: A philosophical study*. Oxford: Oxford University Press.
Comesaña, J. (2010). Evidentialist reliabilism. *Noûs, 44*(4), 571–600.
Conee, E., & Feldman, R. (1998). The generality problem for Reliabilism. *Philosophical Studies, 89*, 1–29.
Conee, E., & Feldman, R. (2001). Internalism Defended. *American Philosophical Quarterly, 38*(1), 1–18.
Conee, E., & Feldman, R. (2008). Evidence. In Q. Smith (Ed.), *Epistemology: New essays*. Oxford: Oxford University Press.
Conee, E., & Feldman, R. (2011). Replies. In T. Dougherty (Ed.), *Evidentialism and its discontents*. Oxford: Oxford University Press.
Dokic, J. (2014). Feelings of (un)certainty and margins for error. *Philosophical Inquiries, 2*(1), 123–144.
Feldman, R. (1985). Reliability and justification. *The Monist, 64*(1), 59–74.
Feldman, R. (1999). Methodological naturalism in epistemology. In J. Greco & E. Sosa (Eds.), *The Blackwell guide to epistemology*. Malden: Blackwell.
Feldman, R. (2005). Justification is internal. In M. Steup & E. Sosa (Eds.), *Contemporary debates in epistemology*. Malden: Blackwell.
Feldman, R., & Conee, E. (1985). Evidentialism. *Philosophical Studies, 48*(1), 15–34.
Frise, M. (2015). Epistemology of memory. In J. Fieser & B. Dowden (Eds.), *The internet encyclopedia of philosophy*http://iep.utm.edu/epis-mem/.
Frise, M. (2017). Internalism and the problem of stored beliefs. *Erkenntnis, 82*(2), 285–304.
Frise, M. (2018). Eliminating the problem of stored beliefs. *American Philosophical Quarterly, 55*(1), 63–79.
Goldman, A. (1979). What is justified belief? In G. Pappas (Ed.), *Justification and knowledge*. Dordrecht: Reidel.
Goldman, A. (1999). Internalism exposed. *The Journal of Philosophy, 96*(6), 271–293.
Goldman, A. (2009). Internalism, externalism, and the architecture of justification. *The Journal of Philosophy, 106*(6), 309–338.
Goldman, A. (2011). Toward a synthesis of reliabilism and evidentialism? Or: Evidentialism's troubles, reliabilism's rescue package. In T. Dougherty (Ed.), *Evidentialism and its discontents*. Oxford: Oxford University Press.
Greco, J. (2005). Justification is not internal. In M. Steup & E. Sosa (Eds.), *Contemporary debates in epistemology*. Malden: Blackwell.
Harker, D. (2013). McCain on weak predictivism and external world scepticism. *Philosophia, 41*(1), 195–202.
Harman, G. (1973). *Thought*. Princeton: Princeton University Press.
Greco, J. (2010). *Achieving knowledge: A virtue-theoretic account of epistemic normativity*. Cambridge: Cambridge University Press.
Huemer, M. (1999). The problem of memory knowledge. *Pacific Philosophical Quarterly, 80*, 346–357.
Kitcher, P. (1992). The naturalists return. *Philosophical Review, 101*(1), 53–114.
Koriat, A. (2002). Metacognition research: An interim report. In T. Perfect (Ed.), *Applied metacognition*. West Nyack: Cambridge University Press.
Koriat, A., & Adiv, S. (2012). Confidence in One's social beliefs: Implications for belief justification. *Consciousness and Cognition, 21*(4), 1599–1616.
Koriat, A., & Helstrup, T. (2007). Metacognitive aspects of memory. In S. Magnussen (Ed.), *Everyday memory*. Independence: Taylor and Francis.
Kornblith, H. (2007). The naturalistic project in epistemology: Where do we go from Here? In C. Mi & R. Chen (Eds.), *Naturalized epistemology and philosophy of science* (pp. 39–59). Amsterdam: Rodopi.

Lyons, J. (2009). *Perception and basic beliefs*. Oxford: Oxford University Press.
Lyons, J. (2013). Should Reliabilists be worried about demon worlds? *Philosophy and Phenomenological Research, 86*(1), 1–40.
McCain, K. (2012). A Predictivist argument against Scepticism. *Analysis, 72*(4), 660–665.
McCain, K. (2014). *Evidentialism and epistemic justification*. New York: Routledge.
McGrath, M. (2007). Memory and epistemic conservatism. *Synthese, 157*, 1–24.
Michaelian, K. (2011). Generative memory. *Philosophical Psychology, 24*(3), 323–342.
Michaelian, K. (2012). Metacognition and endorsement. *Mind & Language, 27*(3), 284–307.
Nagel, J. (manuscript). *Factual memory, internalism, and metacognition*.
Peacocke, C. (1986). *Thoughts: An essay on content*. Malden: Blackwell.
Plantinga, A. (1993). *Warrant and proper function*. Oxford: Oxford University Press.
Proust, J. (2013). *The philosophy of metacognition: Mental agency and self-awareness*. Oxford: Oxford University Press.
Reber, R., & Unkelbach, C. (2010). The epistemic status of processing fluency as source for judgments of truth. *Review of Philosophy and Psychology, 1*(4), 563–581.
Schacter, D. L. (2001). *The seven sins of memory: How the mind forgets and remembers*. Boston: Mariner Books.
Senor, T. (2010). Memory. In E. Sosa & M. Steup (Eds.), *A companion to epistemology*. Malden: Wiley-Blackwell.
Unkelbach, C. (2007). Reversing the truth effect: Learning the interpretation of processing fluency in judgments of truth. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*(1), 219–230.
Wheeler, G., & Peirera, L. M. (2008). Methodological naturalism and epistemic internalism. *Synthese, 163*, 315–328.