

# Epistemic Feedback Loops (Or: How Not to Get Evidence)\*

Nick Hughes  
University of Oxford  
[nickhowellhughes.com](http://nickhowellhughes.com)  
[nickhowellhughes@gmail.com](mailto:nickhowellhughes@gmail.com)

ABSTRACT: Epistemologists spend a great deal of time thinking about how we should respond to our evidence. They spend far less time thinking about the ways that evidence can be acquired in the first place. This is an oversight. Some ways of acquiring evidence are better than others. Many normative epistemologies struggle to accommodate this fact. In this article I develop one that can and does. I identify a phenomenon – *epistemic feedback loops* – in which evidence acquisition has gone awry, with the result that even beliefs based on the evidence are irrational. Examples include evidence acquired under the influence of confirmation bias and evidence acquired under the influence of cognitively penetrated experiences caused by implicit bias. I then develop a theoretical framework which enables us to understand why beliefs that are the outputs of epistemic feedback loops are irrational. Finally, I argue that many popular approaches to epistemic normativity may need to be abandoned on the grounds that they cannot comfortably explain feedback loops. The scope of this last claim is broad: it includes almost all contemporary theories of justified/rational belief and of the epistemology of cognitive penetration.

## §I. Introduction

Epistemologists spend a great deal of time thinking about how we should respond to our evidence. Historically, they have spent far less time thinking about when and how evidence should be acquired in the first place.<sup>1</sup> However, this is beginning to change. Recently there has been increased interest in ‘zetetic’ epistemology – that is, the epistemology of inquiry.<sup>2</sup> The zetetic turn is a welcome development. But something has been missing from it so far. There are at least two central questions

---

\* Forthcoming in *Philosophy and Phenomenological Research*.

<sup>1</sup> This may be because evidence-gathering norms are often thought to be practical, rather than epistemic norms, and so outside the purview of epistemology. For this view, see Conee & Feldman (1985), Feldman (2000, 2005), Kelly (2003), Hedden (2015), and Cohen (2016).

<sup>2</sup> The phrase ‘zetetic epistemology’ comes from Friedman (2020).

for zetetic epistemology: ‘*When* should we acquire new evidence?’, and ‘*How* should we acquire that evidence?’ And whilst there has been a fair amount of work on the first question,<sup>3</sup> there has been much less on the second.<sup>4</sup> This is unfortunate. Some ways of acquiring evidence are better than others. As we will see, many familiar normative epistemologies struggle to accommodate this fact. This article attempts to develop one that can and does.

Section two explains why evidence acquisition matters and describes a phenomenon – *epistemic feedback loops* – in which it has gone awry, with the result that even a belief based on the evidence is irrational, or at least problematic. Section three outlines five desiderata for an adequate theory of feedback loops. Sections four to nine develop a theoretical framework – ‘dispositionalism’ – with which to understand them and show how it satisfies the desiderata. Section ten shows how the framework generalises. Section eleven draws out some consequences for normative epistemology, arguing, amongst other things, that many popular approaches to epistemic normativity may need to be abandoned on the grounds that they cannot comfortably explain feedback loops.

## §II. Epistemic Feedback Loops

To see why evidence acquisition matters, consider these three cases, the first two of which are examples of confirmation bias and the third of which is an example of cognitive penetration caused by implicit bias:

DOMINEERING CEO: Camille is the powerful and intimidating CEO of a large media company. She wishes that the company was financially stable. This causes her to unconsciously put pressure on the CFO to tell her that the company *is* financially stable. As a result of the CFO’s testimony, she forms the belief that matches her desire.

LIKEABLE LEVI: Levi fears that his Sunday league football teammates don’t like him. This causes him to unconsciously seek out evidence that

---

<sup>3</sup> For example, Good (1966) and the subsequent literature it generated; Hall & Johnson (1998); Maher (1990); Myrvold (2012); Friedman (2019, 2020, *fc*); Millson (2021); Stjernberg (2021); Thorstad (2021); Falbo (*fc*); Salow (*fc*); Carr (*ms*); and Flores and Woodard (*ms*), amongst others.

<sup>4</sup> Less, but not none. Baehr (2011), Salow (2018), Miracchi (2019), Dorst (2020), Worsnip (*fc*), and Sosa (*ms*) discuss the issue, as do epistemologists who write about cognitive penetration.

disconfirms his fear and avoid evidence that might confirm it. Having gathered a wealth of evidence that his teammates like him and none indicating that they don't, he comes to believe that he is liked after all.

RACIST NEIGHBOUR: Finn suspects that his neighbour Solomon is angry. Solomon is a young black man, and Finn's suspicion is the result of an unconscious racist implicit bias. Finn's suspicion causes him to have a perceptual experience as of Solomon's face wearing an angry expression. As a result he comes to believe that Solomon is angry. In fact, Solomon's expression is neutral.

Confirmation bias and cognitive penetration are usually theorised separately, but from the point of view of evidence acquisition they have important similarities, making a theory that explains them conjointly desirable.

What similarities? First, in each of these cases a psychological attitude of an agent (Camille's wish, Levi's fear, Finn's suspicion) causes them to unknowingly manufacture evidence in favour of a proposition *p* (The company's finances are stable, Levi is liked, Solomon is angry) which causes them to form a belief that *p*. Second, the resulting beliefs are irrational, or at least epistemically problematic.<sup>5</sup> But the problem is not with the way that Camille, Levi, and Finn *respond* to their evidence. Had the CFO's testimony been uncoerced, had Levi's investigation not been biased, and had Finn's perceptual experience not been caused by his suspicion that Solomon is angry, their beliefs would have been fine. Rather, the problem is with the way the evidence was acquired in the first place. In each case the evidence has, as I will put it, a *deviant etiology*.<sup>6</sup>

I'll call cases like these *epistemic feedback loops*. This article develops a framework which explains why they are epistemically problematic. Its key observation is simple: the ways that Camille, Levi, and Finn acquire their evidence could easily result in them acquiring misleading evidence. The challenge is to sharpen this observation and situate it within a systematic theory.

---

<sup>5</sup> I'm happy to describe them as irrational, but even readers who prefer to think of them as rational-but-nevertheless-problematic will be able to use the framework I'll develop to explain what's going wrong.

<sup>6</sup> Why stipulate that Camille, Levi, and Finn are unconscious of the mechanisms that cause them to manufacture evidence? Because if they were aware that their evidence has a deviant etiology it wouldn't be evidence in the first place. If, for example, Finn knew that it was his suspicion that Solomon is angry that caused him to have a perceptual experience as of Solomon face wearing an angry expression, then the fact that Solomon's face appears to be wearing an angry expression would not be evidence that he is angry. See Salow (2018) for discussion.

Epistemic feedback loops should not only be of interest to epistemologists. They can have serious consequences. As RACIST NEIGHBOUR suggests, cognitively penetrated experiences induced by implicit bias may reinforce racism (c.f. Siegel 2017 who discusses similar cases). Confirmation bias has been cited as an explanation for the rise of online echo chambers (Quattrociocchi *et al.* 2016). Feedback loops may even shape world-historical events: Jonathan Leader Maynard (fc) argues that Joseph Stalin's paranoid convictions were a primary cause of the Soviet Terror of 1936-1938. Suspecting internal conspiracies, Stalin ordered apparatchiks to root out perceived saboteurs from within the party. Under pressure, they coerced "confessions" from the accused which were taken at face-value by the party leadership as evidence of wrongdoing. Executions swiftly followed.

### §III. Desiderata

An adequate theory of feedback loops must satisfy a number of desiderata. First, not every feedback loop is epistemically problematic. Suppose Josie desires (perhaps unconsciously) to be more comfortable. This desire causes her to lie down on the sofa, which causes her to believe that she's comfortable. There is nothing epistemically problematic here, even though the case has the same structure as the problematic cases: a prior psychological attitude towards *p* causes the manufacture of evidence for *p* which causes a belief that *p*. Our first desideratum:

EXCEPTIONS: The theory must be able to account for the fact that some feedback loops are unproblematic and explain the differences between these and problematic feedback loops.

Second, although they all share the same structure at a certain level of abstraction, there are differences between DOMINEERING CEO, LIKEABLE LEVI, and RACIST NEIGHBOUR. One difference is between Camille and Finn, on the one hand, and Levi, on the other. Camille and Finn's prior attitude towards *p* leads directly to the *creation* of evidence for *p*. Whilst Levi might also directly create some of his evidence (by fishing for compliments, for example), he also engages in a process better described as *selecting* evidence, by filtering out unwanted sources. Another difference is in the kind of evidence that is manufactured. Finn manufactures visual-perceptual evidence for *p*. Camille manufactures testimonial evidence for *p*. Levi most likely manufactures a combination of the two. A third difference is that Finn manufactures

evidence that is in some sense internal to his mind, whereas Camille and Levi manufacture evidence in the external world. A fourth difference is in the prior attitudes themselves. It is Camille's *wish* that causes her to manufacture evidence, it is Levi's *fear*, and it is Finn's *suspicion*. Epistemic feedback loops can take many different forms. An adequate theory must be able to accommodate all the forms they can take. This is our second desideratum:

VARIETY: An adequate theory must be able to accommodate the various forms that feedback loops can take.

Third, the phenomenon is not limited to sophisticated, reflective, agents like Camille, Levi, and Finn. Consider:

WOLSEY: Wolsey the dog loves sausages and hopes to find one. This causes him to perceptually experience a stick as a sausage, which causes him to believe that it's a sausage.

Wolsey's belief is epistemically problematic in much the same way as Finn's (though we might hesitate to call it 'irrational' – more on this shortly). An adequate theory should be able to explain why. Hence, our third desideratum:

INCLUSIVITY: The theory must be able to explain why feedback loops are epistemically problematic even when they arise in unsophisticated agents.

Fourth, loop cases (as I'll call them) are importantly different to bad cases of the kind epistemologists are familiar with – cases in which an agent is envatted, deceived by an evil demon, lied to by an apparently credible person, tricked by a hidden light source into believing that a white table is red, and so on. Although the beliefs of agents in these cases are false, they are rationally unimpeachable, unlike Camille, Levi, and Finn's beliefs. This gives us a fourth desideratum:

NUANCE: The theory must be able to explain the differences between loop cases and familiar bad cases.

Fifth, although beliefs resulting from feedback loops are problematic, they are liable to give rise to ambivalence. Camille, Levi, and Finn aren't comfortably lumped in with agents who fail to *respond* correctly to their evidence. Compare them with X, a patient suffering from Othello Syndrome described in a case study by Pal *et al.* (2012). Amongst other things, X believed that pictures of fruit on his wife's social

media page indicated that she was having an affair. Although Camille, Levi, and Finn acquire their evidence in a way that makes their beliefs problematic, they are, unlike X, doing well in at least one way; *given* the evidence they have, it would be surprising if they didn't believe what they do. Indeed, were they to suspend judgement on p or disbelieve that p with this evidence in hand, they would seem to exhibit another kind of rational failing. Our theory should explain why. This gives us a fifth desideratum:

AMBIVALENCE: The theory should explain why beliefs produced by feedback loops are okay on a certain dimension of evaluation.

We have five desiderata: EXCEPTIONS, VARIETY, INCLUSIVITY, NUANCE and AMBIVALENCE. Is there a theory that satisfies all of them? I think there is. In the next sections, I'll present and defend it.

## §IV. Functional and Dysfunctional Cognition

### §4.1. Setting the Stages

My starting assumption is that a core function of cognition is to produce knowledge which the agent can use in action. I'll develop a simple theoretical framework to explain feedback loops by reference to cognition working well or badly relative to this purpose with respect to a given proposition p.<sup>7</sup> The framework builds on ideas developed by Maria Lasonen-Aarnio (2010, 2020, 2021, fc1, fc2) and recently put to work by Tim Williamson (fc1, fc2) and myself (Hughes fc1, fc2). I'll call it 'dispositionalism'.

The framework makes use of two distinctions.

The first is between cognition working well or badly *locally* and cognition working well or badly *broadly*. Local evaluations focus on the results cognition produces in the world in which it takes place. Broad evaluations focus on the results it produces in a range of counterfactual worlds.

---

<sup>7</sup> It is, I believe, possible to develop a version of the framework on the starting assumption that a core function of cognition is merely to produce *true beliefs* (rather than knowledge), though I will not do that here.

The second distinction is between cognition working well or badly when the agent *acquires* evidence and cognition working well or badly when the agent *responds* to their evidence. I'll call these the *acquisition stage* and the *response stage* respectively.

I'll say that cognition is *functional* iff it is working well and *dysfunctional* iff it is working badly.

Since the framework aims to explain commonalities between Camille, Levi, and Finn and less sophisticated agents like Wolsey the dog it deliberately avoids employing the anthropocentric language, familiar to epistemologists, of justifications, excuses, responsibility, reasons, rationality, reasonableness, virtue, oughts, duties, requirements, praiseworthiness, blameworthiness, and so on. Such concepts are only comfortably applied to sophisticated, self-conscious, reflective, adult humans.<sup>8</sup> Later we will see how the framework can be used to shed new light on some of them.

I'll make some simplifying assumptions. First, I'll assume that, for every p, an agent either believes that p, suspends judgement on p, or disbelieves that p. Second, I'll assume that an agent believes that p iff they don't suspend judgement on p or disbelieve that p, disbelieves that p iff they don't suspend judgement on p or believe that p, and suspends judgement on p iff they don't believe that p or disbelieve that p. Third, I'll assume that an agent disbelieves that p iff they believe that not-p. Fourth, for the sake of brevity I'll only focus on the coarse-grained attitudes {belief, suspension, disbelief}, ignoring more fine-grained attitudes like credences.<sup>9</sup> The first two assumptions are arguably unrealistic. It might be possible to take no attitude at all towards a proposition. It might be possible to simultaneously take multiple attitudes towards a single proposition (for example, under different modes of presentation). Suspension might be an umbrella concept covering multiple distinct attitudes (Lord & Sylvan *fc*). I'll ignore these complications because nothing turns on them. Accommodating them in the framework would involve needlessly bringing in distracting complexities.

---

<sup>8</sup> See Hughes (*fc2*) for discussion.

<sup>9</sup> Later on (footnote 28) I'll briefly describe some ways in which the framework might be adapted to handle fine-grained attitudes.

## §V. The Acquisition Stage

Starting with the acquisition stage, the two distinctions described above are applied as follows:

- Cognition is *locally functional* with regards to  $p$  at the acquisition stage iff the agent does not acquire misleading evidence about  $p$ .
- Cognition is *locally dysfunctional* with regards to  $p$  at the acquisition stage iff the agent acquires misleading evidence about  $p$ .
- Cognition is *broadly functional* with regards to  $p$  at the acquisition stage iff the way the agent acquires evidence about  $p$  would not normally yield misleading evidence about  $p$ .
- Cognition is *broadly dysfunctional* with regards to  $p$  at the acquisition stage iff it would not be abnormal for the way the agent acquires evidence about  $p$  to yield misleading evidence about  $p$ .

This part of the framework makes use of two concepts – misleading evidence and normality – which require further comment. I'll explain them in turn.

### §5.1. Misleading Evidence

#### §5.1.1. Two Conceptions of Misleading Evidence

What is misleading evidence? We can distinguish between at least two explications of the concept, both of which are perfectly legitimate, but only one of which is relevant for our purposes.

According to the first explication (the irrelevant one), a piece of evidence  $e$  is misleading with respect to  $p$  just in case  $p$  is false and the probability of  $p$  is higher conditional on  $e$  than the unconditional probability of  $p$ . Formally:

- $(P(p|e) > P(p)) \wedge \sim p$



This explication focuses on individual pieces of evidence – a single proposition known to be true, for instance. The second explication focuses on one’s total body of evidence  $E$ , comprised of all of the propositions that are part of one’s evidence. We are interested in whether  $E$  is misleading, not whether  $e$  is. Virtually any way of acquiring evidence about  $p$  will yield some misleading evidence  $e$  bearing on  $p$  in some normal circumstances. Seeing that everyone in the room is under two meters tall is *some* evidence  $e$  for the false proposition that everyone in the world is under two meters tall ( $= p$ ) – it raises the probability of  $p$ , if only very slightly. But nothing abnormal need happen for one to see that everyone in the room is under two meters tall.

### §5.1.2. *Delusive Evidence*

When is one’s total evidence  $E$  misleading with respect to  $p$ ? A natural idea is that  $E$  is misleading with respect to  $p$  iff  $E$  leads an agent into believing that  $p$  when not- $p$ . But which agent? The answer cannot simply be the agent who possesses  $E$ . The presence of images of fruit on his wife’s social media page led  $X$  to believe that she was having an affair. It does not follow that his evidence was misleading with respect to  $p$  ( $= X$ ’s wife is having an affair); the problem was with him, not with his evidence. A more plausible idea is that  $E$  is misleading with respect to  $p$  iff a *rational* response to  $E$  yields a false belief that  $p$ .

This is close to what I have in mind, but it isn’t *quite* right for our purposes. It would make sense if we took the function of cognition to be to produce true beliefs. But my starting assumption is that its function is to produce knowledge, not merely true beliefs. If so, a non-knowlegeable belief is as bad as a false belief. So it’s more natural within the current framework to say that  $E$  is misleading with respect to  $p$  iff a rational response to  $E$  yields a non-knowlegeable belief that  $p$  (this could be a false belief, of course). In §6.2 we will see how this idea can be de-anthropocentred and made more precise – a rational response is a *broadly functional* response.

This way of conceptualising misleading evidence is unconventional insofar as it delivers the result that, for instance, an agent in a Gettier case has misleading evidence. ‘Misleading evidence’ is usually taken to mean: misleading with respect to the truth-value of  $p$ . But we will think of it as meaning: misleading with respect to whether one should believe that  $p$ . If beliefs should be knowledge, then evidence that will lead to non-knowlegeable beliefs is evidence that should be avoided.

Since this is an unconventional way of thinking about misleading evidence, it will be useful to have a name for it, in order to avoid confusion with more traditional truth-centric ways of thinking. So from here on I'll call it *delusive evidence*:

DELUSIVE EVIDENCE: One's total evidence  $E$  is delusive with respect to  $p$  if and only if a broadly functional (i.e. rational) response to  $E$  yields a non-knowledgeable belief that  $p$ .

In line with this, the conditions on functional and dysfunctional acquisition described above can be restated as follows:

- Cognition is locally functional with regards to  $p$  at the acquisition stage iff the agent does not acquire delusive evidence about  $p$ .
- Cognition is locally dysfunctional with regards to  $p$  at the acquisition stage iff the agent acquires delusive evidence about  $p$ .
- Cognition is broadly functional with regards to  $p$  at the acquisition stage iff the way the agent acquires evidence about  $p$  would not normally yield delusive evidence about  $p$ .
- Cognition is broadly dysfunctional with regards to  $p$  at the acquisition stage iff it would not be abnormal for the way the agent acquires evidence about  $p$  to yield delusive evidence about  $p$ .<sup>10</sup>

## §5.2. Normality

The framework also makes use of the concept of normality. What does it mean to say that a way of acquiring evidence would not normally yield delusive evidence, or that it would not be abnormal for it to yield delusive evidence?

---

<sup>10</sup> It might appear, at first glance, that these definitions are circular, since they define broadly functional and dysfunctional acquisition in terms of its relationship to delusive evidence, and delusive evidence is defined in terms of a broadly functional response (thanks to a referee for pointing this out). However, there is no circularity. Circularity would only arise if the definition of a broadly functional *response* made use of the concept of broadly functional or dysfunctional *acquisition*. But as we will see shortly, it does not. (C.f. footnote 22).

### §5.2.1. Normal Circumstances

'Normally' is being used here as a restricted necessity operator. To say that a way of acquiring evidence would not normally yield delusive evidence is to say that in *normal circumstances* it does not yield delusive evidence. To say that it would not be abnormal for a way of acquiring evidence to yield delusive evidence is to say that even in some normal circumstances it yields delusive evidence.

Normality and abnormality are indexed to *ways*. So to say that a way *w* of acquiring evidence would not yield delusive evidence in normal circumstances is to say that *w* would not yield delusive evidence *in circumstances normal for w*.

What are normal circumstances? The question is best answered ostensively, by looking at examples.

### §5.2.2. Normality in Good/Bad Case Pairs

So, consider a good/bad case pair of the familiar kind. In the good case Annie is walking down rue d'Aboukir in Paris, sees that number 44 is a typical Parisian apartment building, and so believes, correctly, that *p* (= 44 is an apartment building). Cognition is locally functional – Annie does not acquire delusive evidence about *p*. Filling in the details in the natural way, it is also broadly functional. The way Annie acquires her evidence – roughly, through a disinterested openness to receiving information from her environment through visual perception<sup>11</sup> – will yield delusive evidence about *p* only in *abnormal* circumstances, such as those in which 44 is actually, say, a giant ventilation shaft hidden behind an ingenious façade giving it the appearance of being a typical Parisian apartment building.

In the bad case the set-up is exactly the same except that 44 is a giant ventilation shaft hidden behind an ingenious façade giving it the appearance of being a typical Parisian apartment building.<sup>12</sup> As before, Annie believes that *p*, only this time falsely and hence non-knowledgeably. Cognition is locally dysfunctional, as Annie acquires delusive evidence. But it is still *broadly functional*. Although Annie acquires delusive evidence, the way she acquires it would not normally yield delusive evidence. It only does so because she is in abnormal circumstances.

---

<sup>11</sup> Disinterested as in impartial, not *un*interested as in indifferent.

<sup>12</sup> As it in fact is. Have a look on Google Earth.

What goes for Annie in the bad case also goes for more radical bad cases in which the agent is not in touch with their environment, such as those involving envattment or the machinations of evil epistemic demons. Suppose that you believe on the basis of your perceptual experience that LeBron James has just hit a game-winning three-pointer. Alas, you're wrong. You're not sitting courtside at Staples Centre, you've been envatted. Even radically bad cases like this are set up on the assumption that the agent acquires their evidence in a disinterested and open way, so that were they reconnected with the environment they would not acquire delusive evidence. Indeed, it is precisely this feature of the agent's cognition that the evil demon or mad scientist exploits as part of their strategy to get the agent to believe falsehoods.

### §5.2.3. *Normality in Loop Cases*

By contrast, the situation is quite different in problematic loop cases. Here cognition is *broadly dysfunctional*. Unlike Annie, Camille, Levi, and Finn are *not* disinterestedly open to receiving information from the environment. Instead they manufacture evidence that harmonises with a prior psychological state. It wouldn't be at all abnormal for the ways they acquire their evidence to yield delusive evidence – most obviously by yielding evidence that will cause them to falsely believe that p. Nothing weird or out-of-the-ordinary would have to happen for this state of affairs to obtain. The reason why is clear: in virtue of being biased towards harmony with a prior psychological state, their ways of acquiring evidence are *indifferent* and *insensitive* to their environments.

### §5.2.4. *Ways of Acquiring Evidence*

I've described two ways of acquiring evidence. The first is through a disinterested openness to one's environment. The second is by manufacturing evidence that harmonises with a prior psychological state. This individuation of ways of acquiring evidence is obviously very coarse-grained. More fine-grained individuations are possible. We could, for instance, distinguish between ways of acquiring evidence that yield visual-perceptual evidence and ways that yield testimonial evidence. We could also distinguish between ways of acquiring evidence that fall under the umbrella description of being disinterestedly open to one's environment. Equally clearly, this is not a complete taxonomy of ways of acquiring evidence. Nevertheless, we have enough here for the purposes of explaining feedback loops, so there is no need to go more fine-grained.

### §5.2.5. *What Are Normal Circumstances?*

Can we say something general about normal and abnormal circumstances, beyond pointing to intuitive examples? As satisfying as it would be to have a reductive analysis, we have no right to expect one, for there is no reason to think that normality can be defined in terms of more basic concepts. Nevertheless, some things can be said. What counts as normal for our purposes is not a matter of statistical frequency. It would not be abnormal for this week's winning lottery numbers to be [04, 10, 18, 19, 45, 36] even though these numbers will win only very infrequently.<sup>13</sup> It would also be a mistake to take normal circumstances for performing a task (in our case, acquiring evidence) to be circumstances that are favourable to the successful execution of that task (e.g. acquiring non-delusive evidence).<sup>14</sup> One reason for this is because doing so would give us the implausible result that cognition is broadly functional *no matter how* one acquires one's evidence. Another is because some favourable circumstances are intuitively highly abnormal. For instance, favourable circumstances for acquiring evidence in the way that Levi does include those in which a guardian angel changes the world to fit with his evidence.

Bob Beddor and Carlotta Pavese (2020) suggest that normal circumstances for performing a task are, roughly, circumstances in which it would be fair to evaluate the performance.<sup>15</sup> This seems to fit with what we've seen so far. Intuitively, it is not fair to negatively evaluate Annie's way of acquiring evidence based on how it performs in the bad case; she was unlucky. By contrast, it *is* fair to negatively evaluate Camille, Levi, and Finn's ways of acquiring evidence based on how they perform in cases in which *p* is false.

Martin Smith (2010, 2016) argues that normal circumstances are those that require no special explanation. Adapting this idea for our purposes, we might say that normal circumstances for way *w* of acquiring evidence to have some outcome *o* (e.g. yielding delusive evidence) are circumstances in which *w* resulting in *o* doesn't require a special explanation, and that abnormal circumstances for *w* resulting in *o* are those in which *w* resulting in *o* does require a special explanation. Again, this seems to fit with what we've seen. Were Annie to end up with delusive evidence,

---

<sup>13</sup> C.f. Smith (2010, 2016)

<sup>14</sup> C.f. Beddor & Pavese (2020).

<sup>15</sup> Beddor and Pavese wisely refrain from proposing this as a non-circular analysis of normal circumstances; it may be that judgements about whether it is fair to evaluate a performance in such-and-such circumstances are guided by tacit judgements about whether those circumstances are normal or abnormal for performances of the relevant kind.

given the way she acquires it, this would cry out for an explanation – for instance, that she was taken in by an ingenious façade. By contrast, no special explanation is called for in order for us to understand what’s happened when Camille, Levi, and Finn end up with delusive evidence, given how they acquire it.<sup>16</sup>

I won’t take a stance on these proposals here. Nor will I offer my own. Instead I propose to treat normality as primitive and let judgements about cases be our guide.<sup>17</sup>

## §VI. The Response Stage

### §6.1. Local Response

Now for the response stage. In keeping with the assumption that a core function of cognition is to produce knowledge, cognition is locally functional with regards to p at the response stage iff:

1. The agent knows that p, or:
2. The agent knows that not-p, or:
3. The agent suspends judgement on p and the agent is not in a position to know that p and not in a position to know that not-p.

Cognition is locally dysfunctional with regards to p at the response stage iff:

1. The agent believes that p without knowing that p, or:
2. The agent disbelieves that p without knowing that not-p, or:
3. The agent suspends judgement on p and the agent is in a position to know that p or is in a position to know that not-p.

---

<sup>16</sup> Again, we need not view this as an attempt to offer a non-circular analysis.

<sup>17</sup> For other ways of thinking about normality, see Millikan (1984), Nickel (2016), and Graham (2017), amongst others.

### §6.1.1. *Being in a Position to Know*

This part of the framework makes use of the concept of ‘being in a position to know’. When is an agent in a position to know that  $p$ ? The question to ask is: ‘If the agent believed that  $p$  at  $t$ , would they know that  $p$  at  $t$ ?’. If the answer is ‘yes’ they are in a position to know that  $p$  at  $t$ . If the answer is ‘no’, they are not.<sup>18</sup> For illustration, take the proposition ‘Angela Merkel’s father was not over 5 meters tall’. Presumably you believe it, and rightly so: that belief is knowledge. By contrast, take the proposition ‘There were 2991 Tigers in India at noon on July 7<sup>th</sup> 2020’. Presumably you suspend judgement on it. Rightly so: you’re not in a position to know if it’s true.<sup>19</sup>

Cognition would have been locally dysfunctional if you’d believed that there were 2991 Tigers in India at noon on July 7<sup>th</sup> 2020, since you are not in a position to know whether it is true or false. It would also have been locally dysfunctional if you’d suspended judgement on whether Angela Merkel’s father was over 5 meters tall, since you are in a position to know that he wasn’t. A consequence is that suspension of judgment is not a safe harbour which is always epistemically okay, as it is sometimes thought. I think this is the right result. If the function of cognition is to produce knowledge which the agent can use in action, it is not working well if it is resistant to knowledge.<sup>20</sup>

We should be careful interpreting ‘in a position to...’ in the statement ‘ $S$  is in a position to know that  $p$ ’. On a very liberal interpretation, one is in a position to know everything that is entailed by one’s evidence, including all logical and mathematical truths, many of which are highly non-obvious. But it is surely wrong to say that the average person is cognitively dysfunctional when they suspend judgement on whether the Riemann Hypothesis is correct, even though the answer is entailed by their evidence. This observation motivates a more restrictive interpretation, to the effect that unless one is *able to reliably infer  $p$*  from one’s evidence, one is not in a position to know that  $p$ . The average person is not able to reliably infer the truth-value of the Riemann Hypothesis from their evidence – were they to form a belief about it, they would be guessing.

---

<sup>18</sup> The conditional ‘if  $S$  believed that  $p$ , they would know that  $p$ ’ is a useful heuristic for the purposes of latching on to the intended interpretation of ‘in a position to know’, but it is fallible, and so it shouldn’t be used in an analysis. We can contrive situations in which  $S$  is, intuitively, in a position to know that  $p$ , but were  $S$  to believe that  $p$ ,  $p$  would be rendered false and hence not knowable.

<sup>19</sup> As these examples suggest, our interest is in *dispositional* belief, rather than occurrent, conscious, judgement. Prior to reading this paragraph, you had probably never entertained the proposition ‘Angela Merkel’s father was not over 5 meters tall’. Nevertheless, you believed it, dispositionally.

<sup>20</sup> C.f. Williamson (fc2).

What is the relevant reading of 'able'? We should let judgements about cases be our guide. If S is intuitively doing okay (locally) in suspending judgement on p, then she was not in a position to know whether p. If S is intuitively not doing okay (locally) in suspending judgement on p, then she was in a position to know whether p. Intuitively, you're doing okay when you suspend judgement on whether the Riemann Hypothesis is correct. You're *not* doing okay if you suspend judgement on whether Angela Merkel's father was over 5 meters tall. Of course, there are bound to be borderline cases in which intuitions are murky, but precision should not be attempted where it is unavailable.

## §6.2. Broad Response

As the biconditionals above indicate, whether cognition is locally functional or dysfunctional at the response stage is a matter of the results it produces in the world in which cognition takes place. Whether it is *broadly* functional or dysfunctional at the response stage depends, as it does at the acquisition stage, on the results it produces in a range of counterfactual worlds. We focus on the results it produces when the evidence is acquired in a *non-deviant* way.

Cognition is broadly functional with regards to p at the response stage iff:

1. The agent believes that p and the way the agent responds to the evidence for p manifests a disposition to believe that p iff they are in a position to know that p, or:
2. The agent disbelieves that p and the way the agent responds to the evidence for p manifests a disposition to disbelieve that p iff they are in a position to know that not-p, or:
3. The agent suspends judgement on p and the way the agent responds to the evidence for p manifests a disposition to suspend judgement on p iff they are not in a position to know that p and not in a position to know that not-p.

Cognition is broadly dysfunctional with regards to p at the response stage iff:



1. The agent believes that p and the way the agent responds to the evidence for p does not manifest a disposition to believe that p iff they are in a position to know that p, or:
2. The agent disbelieves that p and the way the agent responds to the evidence for p does not manifest a disposition to disbelieve that p iff they are in a position to know that not-p, or:
3. The agent suspends judgement on p and the way the agent responds to the evidence for p does not manifest a disposition to suspend judgement on p iff they are not in a position to know that p and not in a position to know that not-p.

### §6.2.1. Knowledge-Conducive/Resistant Dispositions

The idea that doxastic states should be evaluated by looking at the dispositions the agent manifests relative to knowledge is developed in detail by Lasonen-Aarnio (2010, 2020, 2021, fc1, fc2). Following Lasonen-Aarnio, I'll call the functional dispositions 'knowledge-conducive'. I'll call the dysfunctional dispositions 'knowledge-resistant'.

What does it mean to say that an agent manifests a disposition to believe that p iff they are in a position to know that p? A plausible answer takes its lead from our definition of broadly functional acquisition, in which evidence acquisition is broadly functional just in case the way the agent acquires evidence does not yield delusive evidence in normal circumstances. Following on from that idea, we can say that an agent manifests a disposition to believe that p iff they are in a position to know that p just in case *in normal circumstances* the way they respond to their evidence results in them believing that p iff they are in a position to know that p.<sup>21,22,23</sup> Again, it will

---

<sup>21</sup> C.f. Lasonen-Aarnio (2010, fc1, fc2)

<sup>22</sup> Earlier (footnote 10) I flagged a worry about circularity. Since the framework defines broadly functional and dysfunctional acquisition partly in terms of a broadly functional response, it might look as though the definitions of broadly functional and dysfunctional cognition are circular. We are now in a position to see that they are not. Were the concepts of broadly functional and dysfunctional *response* being defined by reference to broadly functional or dysfunctional *acquisition*, there would be circularity. But they are not. Rather, they are defined independently, in terms of whether, in normal circumstances, the way the agent responds to their evidence results in them believing that p iff they are in a position to know that p.

<sup>23</sup> Suppose that p is necessarily false. In that case, there are no circumstances in which one is in a position to know that p. *Ipsa facto* there are no normal circumstances in which one is in a position to know that p. Hence, if p is necessarily false, cognition can never be broadly functional at the response stage when one believes that p. But that is implausible. Suppose you are unwittingly using a faulty calculator. It

be helpful to look at some examples, starting with variations on the good and bad cases.

### §6.2.2. *Knowledge-Conducive Dispositions in the Bad Case*

So, think again about Annie in the bad case, walking down rue d'Aboukir. She believes that 44 is an apartment building ( $= p$ ). In fact it is an ingenious façade hiding a ventilation shaft. Although cognition is locally dysfunctional, since her belief is not knowledge, on the natural way of filling in the details of the case it is broadly functional: the way Annie responds to her evidence would have resulted in her knowing that  $p$  in normal circumstances. The only reason she believes without knowing is because she is in abnormal circumstances. Hence, she manifests a knowledge-conducive disposition. The same goes in radically bad cases. You believe on the basis of your perceptual experience that LeBron James has just made a game-winning three-pointer, but you're wrong – you've been envatted. You believe that  $p$  without knowing that  $p$ , but the way you respond to your evidence would have resulted in you knowing that  $p$  in normal circumstances. Hence, you manifest a knowledge-conducive disposition.

### §6.2.3. *Knowledge-Resistant Dispositions in the Good and Bad Case*

Now imagine that in the bad case Annie instead refuses to believe her eyes and inexplicably suspends judgement on whether 44 is an apartment building ( $= p$ ). Cognition is now locally functional, since she's not in a position to know that  $p$  and not in a position to know that not- $p$ . However, it is broadly *dysfunctional*: in normal circumstances the disposition Annie manifests would lead to her suspending judgement even when she is in a position to know. Hence, she is *knowledge-resistant*.

---

incorrectly tells you that  $744 \times 962 = 714728$ . Even though this is necessarily false, it is surely rational for you to believe it, given that you have no reason to doubt the reliability of the calculator. We can solve this problem by tweaking the framework. Instead of saying that cognition is broadly functional with regards to  $p$  iff, in normal circumstances, the way the agent responds to the evidence for  $p$  results in them believing that  $p$  iff they know that  $p$ , we can say that it is broadly functional with regards to  $p$  iff, in normal circumstances, the way the agent responds to the evidence for  $p$  results in them believing that  $p$  or a relevantly similar proposition  $p^*$  iff they know that  $p$  or  $p^*$ . This gets us the result that cognition is broadly functional when the agent believes that  $744 \times 962 = 714728$  using an apparently reliable (but actually faulty) calculator, because the way the agent responds to the evidence yields a knowledgeable belief about a relevantly similar proposition  $p^*$  (namely, the necessary truth that  $744 \times 963 = 715728$ ) in normal circumstances, since normal circumstances are those in which the calculator is not faulty. This tweak can also be used to deal with the problem as it arises for suspension of judgement and disbelief. It is similar to the way safety theories of knowledge overcome problems concerning necessary truths. Thanks to a referee for encouraging me to discuss this issue.

The same goes for you if you suspend judgement on whether LeBron just made a game-winning three-pointer.

Next, consider the good case again, in which 44 is an apartment building (= p). Annie believes that p and this belief is knowledge. Now suppose that a prankster resident falsely tells her that it is a façade hiding a ventilation shaft. Annie refuses to believe him, despite having no reason to think that he's lying. Again, cognition is broadly dysfunctional, as the way Annie responds to her evidence for p does not manifest a disposition to believe that p iff she is in a position to know that p: it would not be abnormal for Annie's dogmatism to cause her to believe that p even when she doesn't know that p.<sup>24</sup> She is knowledge-resistant.

#### §6.2.4. Knowledge-Conducive/Resistant Dispositions in Loop Cases

We saw that Camille, Levi, and Finn's cognition is broadly *dysfunctional* at the acquisition stage. How do they fare at the response stage? Here cognition is broadly *functional* when they believe that p. Recall, Camille has been told by the CFO that the company is financially stable, Levi has received lots of compliments and no criticism, and Finn has had a perceptual experience as of Solomon's face wearing an angry expression. It's no surprise that they believe what they do, given this evidence. As noted earlier, the problem is not with how they respond to their evidence, but with how it was acquired. Whilst there are normal circumstances in which they believe that p without knowing that p, owing to the deviant etiology of their evidence, the way they respond to their evidence manifests a disposition to believe that p iff they are in a position to know that p *when the evidence is acquired in a non-deviant way*: were the evidence *not* to have a deviant etiology, the way they respond to it would normally result in them knowing that p. By contrast, cognition would be broadly *dysfunctional* at the response stage were they to suspend judgement on p. Like Annie when she inexplicably suspends judgement on whether 44 is an apartment building, they would manifest a disposition to suspend judgement even when they are in a position to know. Hence, they would be knowledge-resistant.

---

<sup>24</sup> C.f. Lasonen-Aarnio (fc2)

## §VII. Applying the Framework

### §7.1. Good/Bad Case Pairs

This framework sheds light on epistemic feedback loops. Before we see exactly what it says about them, it will be useful to summarise what it says about good/bad case pairs in order to highlight the similarities and differences between these and loop cases.

The good and bad cases are represented in table 1. The belief rows tell us whether cognition is functional or dysfunctional when the agent believes that  $p$ , the suspension rows tell us whether it is functional or dysfunctional when the agent suspends judgement on  $p$ , the disbelief rows tell us whether it is functional or dysfunctional when the agent disbelieves that  $p$ . A green cell with an ' $F$ ' means cognition is functional. A red cell with a ' $DF$ ' means it is dysfunctional.

Table 1	Doxastic attitude	Local acquisition	Broad acquisition	Local response	Broad response
Good case	Belief	F	F	F	F
	Suspension			DF	DF
	Disbelief			DF	DF
Bad case	Belief	DF	F	DF	F
	Suspension			F	DF
	Disbelief			DF	DF

In several recent papers, I have argued that the bad case is an *epistemic dilemma* – a situation in which one is doomed to do wrong whichever doxastic attitude one adopts towards *p* (Hughes 2019, fc2, fc3, fc4). The table reflects this idea. In the bad case there is no way for cognition to be functional both locally and broadly at the response stage.<sup>25</sup> If the agent suspends judgement on *p*, then whilst cognition is locally functional, it is broadly dysfunctional. If the agent believes that *p*, then whilst cognition is broadly functional, it is locally dysfunctional. If the agent disbelieves that *p*, cognition is both locally and broadly dysfunctional. As we'll see, feedback loops are also epistemic dilemmas, though in a different way.

## §7.2. Loop Cases

How should we fill in a table like the one above for loop cases? It depends. Is it possible for Camille, Levi, and Finn to come to *know* that *p*, given the deviant etiology of their beliefs? Intuitively, it is not – or at least, so it seems to me. Normality-theoretic approaches to knowledge (e.g., Goodman & Salow 2018, Beddor & Pavese 2020) deliver this verdict. According to these views, an agent knows that *p* only if the way the agent forms their belief results in a true belief in all normal circumstances. Since there are normal circumstances in which Camille, Levi, and Finn falsely believe that *p*, it follows that they don't know that *p*.

However, according to minimalist similarity-theoretic approaches to knowledge, which take safety from error in similar worlds to be necessary *and sufficient* for knowing (Lasonen-Aarnio 2010, Byrne 2018, Hirvela *fc*), it seems to be in principle possible for an agent in a loop cases to come to know that *p*, as there may be no nearby world in which they falsely believe that *p*. Perhaps, for instance, Levi is such a fantastic guy that things would have to be very different for him to believe that he is liked when he isn't.

The dispositionalist framework is able to capture the distinctive epistemic properties of feedback loops on either assumption. For this reason, I'll remain neutral on the knowledge question and simply show how things look given the respective assumptions.

---

<sup>25</sup> Williamson (fc2) echoes this point.

### §7.2.1. No Knowledge

DOMINEERING CEO, LIKEABLE LEVI, and RACIST NEIGHBOUR don't specify whether p is true (i.e. whether the company is financially stable, Levi is liked, and Solomon is angry). Obviously in versions of the cases in which p is false Camille, Levi, and Finn can't come to know that p. But how do things look on the assumption that they can't come to know that p even if p is true (perhaps because they falsely believe that p in some normal circumstances)?

In that case, cognition is:

- *Locally dysfunctional at the acquisition stage*, as they acquire delusive evidence about p.
- *Locally dysfunctional at the response stage when they believe that p*, as they don't know that p.
- *Broadly dysfunctional at the acquisition stage*, as the way they acquire evidence yields delusive evidence about p in some normal circumstances.
- *Broadly functional at the response stage when they believe that p*, as they manifest a knowledge-conducive disposition.

Table 2 represents the situation and compares it with the bad case:

Table 2	Doxastic attitude	Local acquisition	Broad acquisition	Local response	Broad response
Feedback case without knowledge	Belief	DF	DF	DF	F
	Suspension			F	DF
	Disbelief			DF	DF
Bad case	Belief	DF	F	DF	F
	Suspension			F	DF
	Disbelief			DF	DF

Notice the contrast. None of Annie, Camille, Levi, or Finn know that p. However, the framework identifies an additional problem with Camille, Levi, and Finn that Annie doesn't have – cognition is broadly dysfunctional at the acquisition stage.

### §7.2.2. Knowledge

How do things look on the assumption that knowledge is possible in loop cases? Of course, if p is false then Camille, Levi, and Finn can't come to know that p. In those versions of the cases the situation will be the same as above. But what about versions of the cases in which p is true and they each successfully come to know that p?

In these cases cognition is:

- *Locally functional at the acquisition stage*, as they do not acquire delusive evidence about p

- *Locally functional at the response stage when they believe that p, as they believe knowledgeably.*
- *Broadly dysfunctional at the acquisition stage, as the way they acquire their evidence for p yields delusive evidence about p in some normal circumstances.*
- *Broadly functional at the response stage when they believe that p, as the way they respond to the evidence manifests a knowledge-conducive disposition.*

Table 3 represents the situation and compares it with the good case:

Table 3	Doxastic attitude	Local acquisition	Broad acquisition	Local response	Broad response
Feedback case with knowledge	Belief	F	DF	F	F
	Suspension			DF	DF
	Disbelief			DF	DF
Good case	Belief	F	F	F	F
	Suspension			DF	DF
	Disbelief			DF	DF

Again, notice the contrast. Although Annie, Camille, Levi, and Finn all know that p, the framework identifies a problem with Camille, Levi, and Finn that Annie doesn't have – cognition is broadly dysfunctional at the acquisition stage.



### §7.3. Dilemmas

On either assumption, a consequence is that feedback loops are also epistemic dilemmas – this time at the level of broad cognition. Since cognition is broadly dysfunctional at the acquisition stage whatever doxastic attitude the agent adopts, it is dysfunctional *overall*, even though it is broadly functional at the response stage when the agent believes that *p*.

There would be no such dilemma if Camille, Levi, and Finn had not engaged in broadly dysfunctional evidence acquisition in the first place, and it is plausible that there is a norm prohibiting broadly dysfunctional acquisition. Call it the ‘Acquisition Norm’:

ACQUISITION NORM: One must not engage in broadly dysfunctional evidence acquisition.

Hence, these are dilemmas *perplexity secundum quid* – they only arise because agents in loop cases have already violated the Acquisition Norm. However, given that they now have their evidence they must respond to it – the evidence cannot be ‘undone’, so to speak. In this respect feedback loops are different to situations in which an agent dysfunctionally believes that *p*, knows that if *p* then *q*, and is on the cusp of inferring that *q*. This agent has an escape route – the thing for them to do isn’t to believe that *q*, but to give up their belief that *p*.

## §VIII. Rationality & Oughts

In §2 I described Camille, Levi, and Finn’s beliefs as ‘irrational’. Is there a natural way of mapping the concept of rationality on to the dispositionalist framework, which does not make use of it, but rather makes use of the concepts of functional and dysfunctional cognition? I think there is. A natural proposal is that a doxastic attitude is rational iff cognition is broadly functional at *both* the acquisition and response stage, and irrational otherwise. This gives us the result that the beliefs of agents in familiar good and bad cases, like Annie, are rational, as are those of agents envatted and tricked by evil epistemic demons, but that the beliefs of agents in loop cases are irrational, even if they are knowledgeable (assuming that they *can* be

knowledgeable).<sup>26</sup> It also gives us the result, which may seem undesirable to some, that feedback loops are rational dilemmas.

On the plausible assumption that one ought to be rational and ought not to be irrational, we can also derive deontic conclusions. Agents like Annie in good and bad cases ought to believe that *p*, as ought agents who are envatted or tricked by evil epistemic demons. Agents in loop cases ought not to believe that *p* and face deontic dilemmas.

## §IX. The Desiderata

How does dispositionalism fare on the five desiderata described in §3?

### §9.1. EXCEPTIONS

Let's start with EXCEPTIONS. Consider Josie again. She desires to be comfortable. This desire causes her to lie down on the sofa, which causes her to believe that she is comfortable. Josie acquires evidence for *p* by *making p* the case. There are no normal circumstances in which this way of acquiring evidence yields delusive evidence. As a result, although this case has the same superficial structure as problematic loop cases, the dispositionalist framework judges it quite differently. Unlike with Camille, Levi, and Finn, Josie's cognition is broadly functional at the acquisition stage. Since her response to the evidence is also broadly functional, the theory tells us that Josie's belief is rational, unlike Camille, Levi, and Finn's. This is the right result. Hence, the EXCEPTIONS desideratum is satisfied.

### §9.2. VARIETY

What about VARIETY? Dispositionalism says that cognition is broadly dysfunctional at the acquisition stage iff the *way* the agent acquires evidence yields delusive evidence in some normal circumstances. This diagnosis is indifferent to whether the evidence is created (as in RACIST NEIGHBOUR and DOMINEERING CEO) or selected (as in LIKEABLE LEVI); whether the prior attitude towards *p* is fear, desire, suspicion, or

---

<sup>26</sup> What about Wolsey? We might be reluctant to call his beliefs 'rational' or 'irrational' (see Hughes fc2 for discussion). If so, we can simply note that his cognition is broadly dysfunctional at the acquisition stage and broadly functional at the response stage.

something else; whether the evidence is testimonial, visual-perceptual, or of some other form; whether the evidence is internal or external to the mind; and whether the case is one of confirmation bias or cognitive penetration. Despite these differences, in each case dispositionalism identifies a common problem: the way in which the evidence was acquired yields delusive evidence in some normal circumstances. Hence, the VARIETY desideratum is satisfied.

### §9.3. INCLUSIVITY

What about INCLUSIVITY? Recall, Wolsey's belief is just as epistemically problematic as Finn's. Since dispositionalism makes no appeal to properties found only in sophisticated agents, it is capable of explaining what's going wrong with Wolsey. The way in which he acquires evidence yields delusive evidence in some normal circumstances. Hence, the INCLUSIVITY desideratum is satisfied.

### §9.4. NUANCE

What about NUANCE? We've already seen that there is a difference between Annie in the bad case and our feedback loop agents. Even though Annie's cognition is locally dysfunctional at the acquisition stage, it is broadly functional – Annie would not acquire delusive evidence in normal circumstances. This explains why it's rational for her to believe that *p*, but not rational for Camille, Levi, and Finn to believe that *p*. What goes for Annie also goes for radically bad cases such as those involving envattment and the machinations of evil epistemic demons – here too cognition is broadly functional at the stage of evidence acquisition. Hence, dispositionalism explains the differences between bad cases and loop cases. It also recognises similarities between them – cognition is broadly functional at the response stage with each of Annie, Camille, Levi, and Finn when they believe that *p*. Hence, the NUANCE desideratum is satisfied.

### §9.5. AMBIVALENCE

What about AMBIVALENCE? As noted, in one way cognition is functional with Camille, Levi, and Finn when they believe that *p* – it is broadly functional at the response stage. It would *not* be functional if they were to suspend judgement on *p* or disbelieve that *p*. If we were to focus only on the response stage and ignore the way in which their evidence was acquired, Camille, Levi, and Finn would seem to be paragons of rationality. By contrast, patient X is broadly *dysfunctional* at the

response stage. His evidence, we may suppose, is not acquired in a deviant way. Rather, he fails to *respond* to it correctly – images of fruit on your wife’s social media page are not evidence that she is having an affair. So dispositionalism identifies a significant difference between Camille, Levi, and Finn, on the one hand, and patient X on the other. Hence, it satisfies the AMBIVALENCE desideratum.

## §X. Generalising

The dispositionalist framework also generalises, giving us the correct verdicts about another class of cases.<sup>27</sup> A characteristic feature of feedback loops is that a *prior attitude towards p* causes the agent to manufacture evidence in favour of p. Feedback loops are a striking example of bad evidence acquisition. But there is no reason to think that the only way one can engage in bad acquisition is by manufacturing evidence in accordance with a prior attitude. We can imagine cases in which an agent has no prior attitude towards p, but unwittingly manufactures or selects evidence in favour of it nonetheless.

To see this, consider Larry. Larry is a detective. Unfortunately, he’s a very lazy detective. Worse still, he doesn’t know it – he thinks he’s a model of diligence. Investigating a recent murder, Larry gathers up what he believes to be all the relevant available evidence concerning the identity of the murderer. This evidence points to Ryan being the man who did the deed. In fact, however, Larry doesn’t have anything like all the relevant available evidence. Unbeknownst to him, his laziness caused him to gather only the evidence that was most easily available, which just so happens to indict Ryan.

Even if Larry believes that Ryan is the murderer, thereby responding rationally to the evidence he possesses, it is clear that his belief is epistemically defective. Dispositionalism predicts this. Although Larry’s cognition is broadly functional at the response stage, it is broadly *dysfunctional* at the acquisition stage, since the way that he acquires his evidence yields delusive evidence about p in some normal circumstances. This is true even though (we may suppose) Larry had no prior attitude towards the proposition that Ryan is the murderer. Many other similar cases could be constructed. Thus, the dispositionalist framework generalises to explain

---

<sup>27</sup> Thanks to a referee for prompting me to discuss this.

even certain cases of bad evidence acquisition that lack some of the characteristic features of feedback loops.

## §XI. Consequences for Normative Epistemology

This completes my account of feedback loops.<sup>28</sup> In this section I'll draw out some consequences for normative epistemology.

Many theories of rational belief are unable to smoothly explain feedback loops. Here I'll give two examples from the recent literature, before extracting a general lesson. In each case the problem is the same. The theories are designed to deal with the response stage of cognition but ignore the acquisition stage. Patching them up by employing some of the tools of dispositionalism to handle the acquisition stage results in hybrid theories that are less simple, less parsimonious, and less elegant than dispositionalism. Since dispositionalism is able to smoothly explain feedback loops, there is a *pro tanto* reason to prefer it to these competing views.

---

<sup>28</sup> I have only presented an account of feedback loops for the coarse-grained attitudes {belief, suspension, disbelief}. Presumably feedback loops can also produce more fine-grained attitudes like credences. It would be interesting to see how the framework might be adapted to handle such cases (thanks to a referee for encouraging me to think about this). Unfortunately, there is no space to properly explore this issue here – it is a question for future work. Nevertheless, I will briefly mention a couple of options. Sarah Moss (2018) argues that credences can be knowledge, even if they are not truth-apt. She proposes a knowledge norm for credences: one ought to have credence  $n$  in  $p$  iff  $n$  is knowledge. If we accept these ideas, extending the framework to handle credences should be a fairly straightforward matter. Cognition will be locally dysfunctional at the acquisition stage just in case one acquires *delusive\* evidence* for  $p$ , where *delusive\* evidence* is a body of evidence  $E$  such that a rational response to  $E$  results in a non-knowledgeable credence  $n$  in  $p$ . Otherwise, it will be locally functional. Cognition will be broadly dysfunctional at the acquisition stage just in case it would not be abnormal for the way the agent acquires evidence for  $p$  to yield *delusive\* evidence* for  $p$ . Otherwise, it will be broadly functional. Cognition will be locally dysfunctional at the response stage just in case the agent has a non-knowledgeable credence in  $p$  or has an imprecise credence in  $p$  despite being in a position to have a knowledgeable precise credence in  $p$  (imprecise credences being the credal analog to suspending judgement). Otherwise it will be locally functional. It will be broadly dysfunctional at the response stage just in case the way the agent responds to the evidence does not manifest a disposition to have a knowledgeable credence in  $p$ , and broadly functional otherwise. Of course, the idea that credences can be knowledge is controversial. A natural choice for those who reject it is to try to develop an *accuracy-centric* version of the framework, akin to a truth-centric version of the framework for coarse-grained attitudes. I believe this can be done, but working out the details is beyond the scope of this paper.

### §11.1. Lord's Reasons-First View

Errol Lord (2018) argues that one rationally believes that  $p$  iff one believes that  $p$  for a sufficient objective reason  $r$  to do so and manifests know-how to use  $r$  as a reason to believe that  $p$ .<sup>29</sup> He maintains that only facts are objective reasons. In order to accommodate the observation that it's rational for Annie to believe that 44 is an apartment building in the bad case, despite it not being a fact that it is an apartment building, he maintains that the fact that it perceptually appears to Annie that  $p$  is a sufficient objective reason for her to believe that  $p$ .

However, just as it perceptually appears to Annie that 44 is an apartment building, so too it perceptually appears to Finn that Solomon is angry. Yet whereas Annie's belief is rational, Finn's belief is irrational (or, at least, problematic in a way that Annie's isn't). Nothing in Lord's account provides him with the resources to explain this difference. The upshot is that his theory cannot handle feedback loops. Now, admittedly, Lord's theory isn't *designed* to deal with such cases – he is interested in what I have been calling the 'response stage' of cognition. Nevertheless, once we step back and take a broader view of cognition, this is, if not a decisive reason to reject his view, at least a strike against it.<sup>30</sup>

An obvious response is that Lord can simply help himself to the idea of cognition being broadly functional or dysfunctional at the acquisition stage in order to deal with the problem. That's true, so we don't have a knock-down objection to his view here. But it is a move that comes with a cost. The resulting view would be that one rationally believes that  $p$  iff:

- (a.) One believes that  $p$  for a sufficient objective reason  $r$  to do so and manifests know-how to use  $r$  as a reason to believe that  $p$ , *and*
- (b.) One acquires  $r$  in a broadly functional way.

Unlike dispositionalism, this is a hybrid view. It uses one theoretical framework to deal with the acquisition stage – employing the concept of the manifestation of ways of acquiring evidence in normal circumstances – and an entirely different theoretical framework to deal with the response stage – employing the concepts of believing for a sufficient objective reason  $r$  and manifesting know-how to use  $r$  as a reason to

---

<sup>29</sup> Kiesewetter (2017) holds a similar view.

<sup>30</sup> Miracchi (2019) also argues that Lord's view cannot handle cases where the agent acquires their evidence in a bad way, though her approach is different to mine.

believe that *p*. Whilst we shouldn't reject hybrid views out of hand, simplicity, parsimony, and elegance speak in favour of accepting the dispositionalist view, since it gets the job done with fewer resources. There is, then, at least a *pro tanto* reason to prefer dispositionalism to Lord's Reasons-First view.

### §11.2. Dutant & Littlejohn's Probable Knowledge View

Dutant and Littlejohn (fc) argue that it is rational for one to believe that *p* iff it is sufficiently probable on one's evidence that one knows that *p*. Is it sufficiently probable on Annie's evidence in the bad case that she knows that *p*? Dutant and Littlejohn don't say (they focus on how their view handles lottery and preface cases). Presumably it is. If it wasn't, the view would be a non-starter – any plausible theory of rationality must deliver the verdict that Annie's belief is rational. Is it probable on Camille, Levi, and Finn's evidence that *they* know that *p*? Again, the answer is presumably 'yes' – at least according to Dutant and Littlejohn's view. They make no distinction between functional and dysfunctional evidence acquisition – like Lord, they are only interested in the response stage – and Camille, Levi, and Finn's beliefs would be rational were their evidence not to have a deviant etiology. But this means that their view faces the same problem as Lord's: as things stand, it cannot explain why Annie's belief in the bad case is rational but Camille, Levi, and Finn's beliefs are irrational (or, at a minimum, problematic).

Of course, the problem is not insurmountable. Like Lord, Dutant and Littlejohn are free to help themselves to the idea of cognition being broadly functional or dysfunctional at the acquisition stage in order to deal with the problem. But, as before, there is a price to be paid, for the resulting hybrid theory is less simple, less parsimonious, and less elegant than dispositionalism.

### §11.3. A Lesson

A broader lesson can be taken from the above. Normative epistemology needs to account for the fact that some ways of acquiring evidence are bad and others are okay. The best way to do so, I suggest, is to look at the outcomes that different ways of acquiring evidence produce in normal circumstances. Since a simpler theory is, *ceteris paribus*, a better theory, this gives us a reason to adopt the same approach to evaluating ways of responding to evidence – looking at the outcomes that different ways of responding to evidence produce in normal circumstances. This puts pressure on theories of the response stage that don't take this approach. There are

many of them. They include (at least) Evidentialism in its many guises (e.g. Conee & Feldman 2004, Smith 2016); Coherentism (e.g. Lehrer 1990); Phenomenal Conservatism (e.g. Huemer 2001); Dogmatism (e.g. Pryor 2000);<sup>31</sup> Simple Process Reliabilism (e.g. Goldman 1979); and many Knowledge-First theories (e.g. Bird 2007, Reynolds 2011, Ichikawa 2014). This list is open-ended.

#### §11.4. Existing Theories of Cognitive Penetration

There has been some work on the epistemology of cognitive penetration. Since the dispositionalist framework aims to explain what's going wrong in cases of cognitive penetration, we should compare it to existing theories. There are reasons to think that dispositionalism usurps these theories.

##### §11.4.1. *The Generalisation Problem*

Let me explain. Theories of cognitive penetration focus, naturally enough, on cases like RACIST NEIGHBOUR. But the phenomenon of epistemic feedback is broader than that – cases of confirmation bias like DOMINEERING CEO and LIKEABLE LEVI don't involve cognitive penetration. A problem that many attempts to account for cognitive penetration would face were we to treat them as general theories of epistemic feedback loops (though of course they are not offered as such) is that they fail to generalise in a way that explains the broader phenomenon.

For example, consider Matthew McGrath's (2013a, 2013b) 'quasi-inferential' account of cognitive penetration. McGrath argues that beliefs like Finn's are unjustified because the higher-level penetrated perceptual experience (Finn's experience as of Solomon's face wearing an angry expression) on which the belief that *p* is based is unjustifiably quasi-inferred from lower-level perceptual experiences such as shape, size, colour, texture, etc. This process, McGrath argues, is analogous to leaping to a conclusion from insufficient evidence.

Were we to treat this as a general theory of what's going wrong in epistemic feedback loops, it would face the generalisation problem. It doesn't explain what's going wrong in loop cases involving confirmation bias, and so it wouldn't satisfy the VARIETY desideratum. Camille does not believe that the company is financially stable as a result of a quasi-inference from lower- to higher-level perceptual

---

<sup>31</sup> I'm not the first to point out that cases like RACIST NEIGHBOUR cause problems for Phenomenal Conservatism and Dogmatism. See Lyons (2011), Siegel (2012, 2017), and McGrath (2013).



experiences, or, for that matter, as a result of *any* kind of inference akin to leaping to conclusions. She believes it because the CFO told her so (albeit under pressure). Similarly, Levi doesn't believe that he's liked because of a dodgy inference. He believes it because he's had lots of compliments and no criticism. Whatever its merits as a theory of cognitive penetration, McGrath's account would be inadequate as a theory of feedback loops.

Were we to treat them as theories of feedback loops, many theories of cognitive penetration would face the same problem as McGrath's – failing to generalise to explain the broader phenomenon. It would be a problem for Chudnoff's (2020) 'Presentational Conservatism', Siegel's (2017) 'Rationality of Perception' thesis, Brogaard's (2013) 'Sensible Dogmatism', Teng's (2016) 'Imagining Account' and Vahid's (2014) account appealing to the sensorimotor theory of perception, amongst others.

One might think that a theory of cognitive penetration like one of the above can happily co-exist alongside a dispositionalist theory of feedback loops, with the former providing an account of the epistemology of cognitive penetration specifically, and dispositionalism providing an account of feedback loops more generally. But, in fact, it isn't clear that they can. It is a general methodological point that if *A* and *B* are instantiations of the same phenomenon, and one theory (call it 'AB-Theory') successfully explains *both A and B*, whereas another ('B-Theory') could *at most* explain *B*, then B-Theory is either redundant or false. It is redundant if it is compatible with AB-Theory. It is false if it is incompatible with AB-Theory. Either way, we no longer have any use for B-Theory, for it is superseded by AB-Theory. On the account I have offered here, the theories of cognitive penetration above are B-Theories, and dispositionalism is the AB-Theory.<sup>32</sup>

---

<sup>32</sup> Thanks to Bob Beddor, Kevin Dorst, Daniel Drucker, Adam Etinson, David Faraci, Giada Frantantonio, Rachel Fraser, Alex Grzankowski, Mike Hannon, Louise Hanson, Max Hayward, Jaakko Hirvela, Caroline Krager, Maria Lasonen-Aarnio, Clayton Littlejohn, Robin McKenna, Lisa Miracchi, Thi Nguyen, Andrew Peet, Bernhard Salow, Rob Simpson, Tim Williamson, Elise Woodard, two anonymous referees, and audiences at the University of Helsinki, the University of Glasgow, and the Inquiry Network Group.

## §XII. References

- Baehr, J.** 2011. "Evidentialism, Vice, and Virtue" in Dougherty, T. (ed.) *Evidentialism and its Discontents*. Oxford: Oxford University Press
- Beddor, B. & Pavese, C.** 2020. "Modal Virtue Epistemology" *Philosophical and Phenomenological Research* 101 (1): 61-79
- Bird, A.** 2007. "Justified Judging" *Philosophy and Phenomenological Research* 74 (1): 81-101
- Brogaard, B.** 2013. "Phenomenal Seemings and Sensible Dogmatism" Tucker, C (ed.) *Seemings and Justification*. Oxford: Oxford University Press.
- Byrne, A.** 2018. *Transparency and Self-Knowledge*. Oxford: Oxford University Press
- Carr, J.** ms. "Accuracy, Inquiry, and Belief"
- Cohen, S.** 1984. "Justification and Truth" *Philosophical Studies* 46 (3): 279-295
- Cohen, S.** 2016. "Theorizing About the Epistemic" *Inquiry* 59 (7-8): 839-857
- Conee, E. & Feldman, R.** 1985. "Evidentialism" *Philosophical Studies* 48 (1): 15-34
- Conee, E. & Feldman, R.** 2004. *Evidentialism: Essays in Epistemology*. Oxford: Oxford University Press
- Chudnoff, E.** 2020. "Experience and Epistemic Structure: Can Cognitive Penetration Result in Epistemic Downgrade" Nes, A. & Chan, T. (eds.) *Inference and Consciousness*. Routledge
- Dorst, K.** 2020. "Evidence: A Guide for the Uncertain" *Philosophy and Phenomenological Research* 100 (3): 586-632
- Dutant, J. & Littlejohn, C.** forthcoming. "Defeaters as Indicators of Ignorance" Simion, M. & Brown, J. (eds.) *Reasons, Justification, and Defeat*. Oxford: Oxford University Press
- Good, I.J.** 1966. "On the Principle of Total Evidence" *British Journal for the Philosophy of Science* 17 (4): 319-321
- Falbo, A.** forthcoming. "Inquiry and Confirmation" *Analysis*
- Feldman, R.** 2000. "The Ethics of Belief" *Philosophy and Phenomenological Research* 60 (3): 667-695
- Feldman, R.** 2005. "Epistemological Duties" Moser, P. & Feldman, R. (eds.) *The Oxford Handbook of Epistemology*: 253-264. Oxford University Press
- Flores, C. & Woodard, E.** manuscript. "Epistemic Vigilance: In Defense of Epistemic Norms on Evidence-Gathering"
- Friedman, J.** 2019. "Checking Again" *Philosophical Issues* 29 (1): 84-96
- Friedman, J.** 2020. "The Epistemic and the Zetetic" *The Philosophical Review* 129 (4): 501-536

- Friedman, J.** forthcoming. "Zetetic Epistemology" Reed, B. & Flowerree, AK. (eds.) *Towards an Expansive Epistemology: Norms, Action, and the Social Sphere*. Routledge.
- Goldman, A.** 1979. "What Is Justified Belief?" Pappas, G. (ed.) *Justification and Knowledge*. Dordrecht
- Goldman, A.** 1986 *Epistemology and Cognition*. Cambridge, MA: Harvard University Press
- Goodman, J. & Salow, B.** 2018. "Taking a Chance on KK" *Philosophical Studies* 175 (1): 183-196
- Graham, P.** 2017. "Normal Circumstances Reliabilism: Goldman on Reliability and Justified Belief" *Philosophical Topics* 45 (1): 33-61
- Greco, J.** 2005. "Justification is not Internal" in *Contemporary Debates in Epistemology*. Wiley-Blackwell.
- Hall, R. & Johnson, C.** 1998. "The Epistemic Duty to Seek More Evidence" *American Philosophical Quarterly* 35 (2): 129-139
- Hedden, B.** 2015. "Options and Diachronic Tragedy" *Philosophy and Phenomenological Research* 90 (2): 423-451
- Hirvela, J.** forthcoming. "Global Safety – How to Deal With Necessary Truths" *Synthese*
- Huemer, M.** 2001. *Skepticism and the Veil of Perception*. Rowman & Littlefield
- Hughes, N.** 2019. "Dilemmic Epistemology" *Synthese* 196: 4059-4090
- Hughes, N.** forthcoming-1. "Evidence and Bias" Lasonen-Aarnio, M. & Littlejohn, C. (eds.) *The Routledge Handbook of the Philosophy of Evidence*
- Hughes, N.** forthcoming-2. "Epistemic Dilemmas Defended" Hughes, N. (ed.) *Essays on Epistemic Dilemmas*. Oxford: Oxford University Press
- Hughes, N.** forthcoming-3. "Who's Afraid Of Epistemic Dilemmas?" Stapleford, S., Steup, M. & McCain, K. (eds.) *Epistemic Dilemmas: New Arguments, New Angles*
- Hughes, N.** forthcoming-4. "Epistemology Without Guidance" *Philosophical Studies*
- Ichikawa, J.** "Justification Is Potential Knowledge" *Canadian Journal of Philosophy* 44 (2): 184-206
- Kiesewetter, B.** 2017. *The Normativity of Rationality*. Oxford: Oxford University Press
- Kelly, T.** 2003. "Epistemic Rationality as Instrumental Rationality: A Critique." *Philosophy and Phenomenological Research* 66 (3): 612-640
- Lasonen-Aarnio, M.** 2010. "Unreasonable Knowledge" *Philosophical Perspectives*. 24: 1-21
- Lasonen-Aarnio, M.** 2020. "Enkrasia or Evidentialism? Learning to Love Mismatch" *Philosophical Studies* 177 (3): 597-632
- Lasonen-Aarnio, M.** 2021. "Dispositional Evaluations and Defeat" Brown, J. & Simion, M. (eds.) *Reasons, Justification, and Defeat*. Oxford: Oxford University Press

- Lasonen-Aarnio, M.** forthcoming-1. "Virtuous Failure and Victims of Deceit" Dutant, J. & Dorsch, F. (eds.) *The New Evil Demon Problem*. Oxford University Press
- Lasonen-Aarnio, M.** forthcoming-2. "Perspectives and Good Dispositions" *Philosophy and Phenomenological Research*
- Lehrer, K.** 1990. *Theory of Knowledge*. Westview Press
- Littlejohn, C.** 2013, "The Russellian Retreat" *Proceedings of the Aristotelian Society* 113 (3): 293–320
- Lord, C., Ross, L., & Lepper, M.** 1979. 'Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence'. *Journal of Personality and Social Psychology*, 37(11): 2098–2109
- Lord, E.** 2018. *The Importance of Being Rational*. Oxford University Press
- Lord, E. & Sylvan, K.** forthcoming. "Beginning in Wonder: Suspensive Attitudes and Epistemic Dilemmas" Hughes, N. (ed.) *Epistemic Dilemmas*. Oxford University Press
- Lyons, J.** 2011. "Circularity, Reliability, and the Cognitive Penetrability of Perception" *Philosophical Issues* 21: 289-311
- Maher, P.** 1990. "Why Scientists Gather Evidence" *British Journal for the Philosophy of Science* 41 (1): 103-119
- Maynard, J.L.** forthcoming. *Ideology and Mass Killing*. Oxford University Press
- McGrath, M.** 2013. "Phenomenal Conservatism and Cognitive Penetration: The 'Bad Basis' Counterexamples" Tucker, C. (ed.) *Seemings and Justification*. Oxford: Oxford University Press
- Millikan, R.** 1984. *Language, Thought, and Other Biological Categories*. MIT Press
- Millson, J.** 2021. "Seeking Confirmation: A Puzzle for Norms of Inquiry" *Analysis* 80 (4): 683-692
- Miracchi, L.** 2019. "When Evidence Isn't Enough: Suspension, Evidentialism, and Knowledge-first Virtue Epistemology" *Episteme* 16: 413-437
- Moss, S.** *Probabilistic Knowledge*. Oxford University Press.
- Myrvold, W.** 2012. "Epistemic Values and the Value of Learning" *Synthese* 187 (2): 547-568
- Nickel, B.** *Between Logic and the World: An Integrated Theory of Generics*. Oxford University Press.
- Pal, K. et al.** 2012. "Othello Syndrome Secondary to Ropinirole: A Case Study" *Case Reports in Psychiatry* 353021
- Pryor, J.** 2000. "The Skeptic and the Dogmatist" *Nous* 34 (4): 517-549
- Quattrociocchi, W., Scala, A., & Sunstein, C.** 2016 "Echo Chambers on Facebook" Available at <http://dx.doi.org/10.2139/ssrn.2795110>
- Reynolds, S.** 2011. "Justification as the Appearance of Knowledge" *Philosophical Studies* 163: 367-383

- Salow, B.** 2018. "The Externalist's Guide to Fishing For Compliments" *Mind*, 127 (507): 691-728
- Salow, B.** forthcoming. "The Value of Evidence" Littlejohn, C. & Lasonen-Aarnio, M. (eds.) *The Oxford Handbook of the Philosophy of Evidence*. Oxford University Press.
- Siegel, S.** 2012. "Cognitive Penetrability and Perceptual Justification" *Nous* 46 (2): 201-222
- Siegel, S.** 2017. *The Rationality of Perception*. Oxford University Press
- Smith, M.** 2010. "What Else Justification Could Be" *Nous* 44 (1): 10-31
- Smith, M.** 2016. *Between Probability and Certainty*. Oxford University Press
- Sosa, E.** manuscript. "Epistemic Explanations"
- Stjernberg, F.** 2021. "Epistemic Vices, Critical and Zetetic" *Science and Proven Experience, Vetenskap Och Beprovad Erfarenhet*
- Teng, L.** 2016. "Cognitive Penetration, Imagining, and the Downgrade Thesis" *Philosophical Topics* 44 (2): 405-426
- Thorstad, D.** 2021. "Inquiry and the Epistemic" *Philosophical Studies*
- Vahid, H.** 2014. "Cognitive Penetration, the Downgrade Principle, and Extended Cognition" *Philosophical Issues* 24 (1): 439-459
- Williamson, T.** forthcoming-1. "Justifications, excuses, and sceptical scenarios" Dutant, J. & Dorsch, F. (eds.) *The New Evil Demon Problem*. Oxford University Press
- Williamson, T.** forthcoming-2. "Epistemological Ambivalence" Hughes, N. (ed.) *Epistemic Dilemmas*. Oxford University Press
- Worsnip, A.** forthcoming. "The Obligation to Diversify One's Sources: Against Epistemic Partisanship in the Consumption of News Media" Fox, C. & Saunders, J. (eds.) *Media Ethics: Free Speech and the Requirements of Democracy*. Routledge.