# For *The Blackwell Companion to Epistemology*, Third Edition

Kurt Sylvan, Matthias Steup, Ernest Sosa, and Jonathan Dancy (eds.)

**Defeat**

Mona Simion

### 1. Introduction

Investigations into the nature and normativity of defeat come with high theoretical stakes: The notion of defeat has been central to a wide range of different philosophical debates. Unfortunately, not many systematic, full accounts of defeat have been offered.[[1]](#endnote-1) This chapter does two things: first, it surveys classical and recent accounts of defeat; second, it defends a knowledge-first functionalist view of defeat, and it shows that it is superior to both its internalist and its externalist competition on both extensional and theoretical grounds. On this account, defeaters are ignorance indicators: they are facts that one is in a position to know and that reduce one’s evidential probability that *p* is the case.

**2. Terminological Issues and Data**

Traditionally, defeat is taxonomized as coming in two broad categories: psychological (or doxastic) defeat, and normative defeat. One important problem with this way of taxonomizing is that the terminology is often ambiguous about whether it maps onto the target of defeat or the nature of defeat.

One (popular) way to read the distinction is as pertaining to the nature of defeat: on this reading, psychological defeaters are mental states (beliefs, seemings), while normative defeaters lie outside of cognizer’s ken: often, these are unpacked as facts/propositions that the cognizer should have believed/known (Goldberg 2016, Lackey 2008). There are two important problems with this way to taxonomize defeat by its nature: first, insofar as the target of defeat is left unspecified (are we talking about justification defeat or about mere belief defeat?), it may trigger the false implicature that mere unjustified beliefs may act as justification defeaters. The reason for this is that the contrast class is normative defeaters: insofar as psychological defeaters are contrasted with normative defeaters, it is reasonable to assume that what is under discussion is *mere* psychological defeat that carries no normative epistemic weight. Mere psychological defeaters, however – e.g. unjustified beliefs – cannot defeat justification: justification defeat is a normative phenomenon, and non-normative considerations do not carry weight against normative considerations. My wish to own your bag does not defeat the impermissibility of stealing it, no matter how strong it may be – at best, it may defeat my resolve not to steal it. Similarly, my biases against women cannot defeat the justification I get from their testimony – at best, they may defeat my testimonial belief. Belief loss, however, is not justification loss.

A closely related problem has to do with the incompleteness of such taxonomizing: if we read psychological defeaters as *merely psychological –* i.e. not carrying epistemic normative weight – and normative defeaters as facts lying outside of the co~~n~~gnizer’s ken, one crucial category of defeat falls out of our taxonomy: defeat by justified beliefs. These are clearly normative defeaters in the sense that they carry normative weight, but they are also psychologically instantiated (as opposed to what ‘normative defeat’ discussions regularly track, i.e. facts the cognizer should have been aware of).

 I aim to precisify the terminology as follows: when it comes to the *target of defeat*, I will distinguish between:

**Mere doxastic defeat**: defeat that leads to belief loss or decrease in level of confidence without affecting justification.

And

**Justification defeat**: defeat that leads to lowering of the level of justification for the target proposition.

Henceforth, unless otherwise specified, when I talk about defeat what I mean is justification defeat.

Note that this target-centric taxonomy leaves it open whether a particular justification defeater is an *internal defeater* – i.e. instantiated psychologically – or an *external defeater* – a defeater that the cogniser should have registered psychologically but did not.

**Internal defeat**: defeat that is psychologically instantiated.

**External defeat**: defeat that is not psychologically instantiated.

Terminology cleared up, we are now in a position to ask important and substantive philosophical questions having to do with the nature of defeat and its normative source.

To answer these questions, it will be useful to consider some toy cases that instantiate paradigmatic instances of justification defeat.

**Climate Change**: Ann is a climate change sceptic: after having listened carefully to all extant expert testimony on the matter that suggests that anthropogenic climate change is happening, she still holds on to her unshaken belief that climate change is not happening.

Ann has plenty of justification defeaters to her anti-climate-change belief, and nevertheless dogmatically holds on to it. Furthermore, her defeaters are internal defeaters, in that they are psychologically registered. In contrast, consider the case of Bill, the sexist scientist:

**Sexist Scientist:** Bill is a sexist scientist: he regularly dismisses testimony from expert women scientists due to his bias. Indeed, he is so convinced they are not worth listening to that, when women talk about scientific matters, he just zones out. One day his colleague Mary, a renowned expert on the relevant topic, tells him that the experiment that he ran and that shows that p is the case was seriously flawed, and she presents him with conclusive evidence to this effect. As per usual, Bill doesn’t pay any attention to what Mary is saying, and continues believing that p.

The case of Bill is one of external justification defeat. Bill’s justification to believe that p is defeated by Mary’s testimony, but the defeater is not psychologically registered by Bill.

 Finally, consider a classic case of external justification defeat (tracking the phenomenon that has been referred to as *normative defeat* (Lackey 2008)) that has to do with a failure in inquiry, and in which the subject is not faced with the evidence in the way in which Bill is:

**Uninformed Doctor**: Doctor Jones is approaching retirement and hasn’t been reading up on new developments in his field for a while now. Indeed, in spite of the commonly known fact that stomach ulcers are caused by bacteria, he still believes the cause is stress and proceeds accordingly in his practice.

Intuitively, Jones is not justified in believing that ulcers are caused by stress. Just like the case of the sexist scientist, his is also a case of external justification defeat. Note, however, that, as opposed to Bill, Dr. Jones does not believe this in the face of evidence that he is presented with, but rather in the face of evidence that he can easily come to know, and that he should have had in virtue of being a medical doctor.

 These are the cases that I will run with for the rest of this paper. At a minimum, I submit, any reasonably adequate account of defeat should be able to satisfactorily explain the data in these cases. In what follows, I will first argue that internalism about defeat faces in-principle, unsurmountable problems when it comes to extensional adequacy. Further on, I will look at the main varieties of defeat externalism and argue that they struggle to accommodate external defeat. Finally, I will defend my preferred account of defeaters as evidential-probability-decreasing facts that one is in a position to know.

**3. Defeat Internalism**

The first and the classic view on the nature of defeat in epistemology is due to Pollock (1986). According to this view, *D* is a defeater of *E*’s support for *p* for *S* if and only if (i) *E* is a reason to believe *p* for *S*, and (ii) *E*&*D* is not a reason to believe *p* for *S* (henceforth *Pollock’s view*).

 The account has a lot going for it; most crucially, it enjoys a high degree of prior plausibility, since it is, to some extent, neutral between internalism and externalism about defeat, in virtue of remaining silent on the nature of reasons and, most importantly, on what it is for something to be a reason *for S* to believe. Both internalist and externalist accounts of reasons and of the having relation can be conjoined with Pollock’s proposal, to deliver very different accounts of the nature of defeat.

 There are two important standing limitations to Pollock’s view: first, it does not account for partial defeat – nor is it trivial to see how it could be extended to do so without major interventions. Second, its neutrality between externalist and internalist readings also renders it uninformative: since the account remains silent on the nature of reasons and on what it is for something to be a reason for S to believe, it dwells at too high a level of theoretical generality: we need more substance to the view in order to be able to investigate its extensional adequacy.

 Pollock was an internalist evidentialist; thus, one way to spell out Pollock’s view that suggests itself is a traditional, seemings-based recipe: on this account, reasons for S to believe are S’s relevant seemings. Unfortunately, a view like this will get us in trouble with extensional adequacy: on the necessity direction, recall only the sexist Bill, who tunes out whenever a woman speaks to him, or the uninformed doctor Jones. These guys don’t host any relevant seemings – intuitively, however, their beliefs are defeated. Against the sufficiency direction, cases of cognitive penetration will create trouble for a seemings-based defeat account (e.g. Siegel 2012): the fact that it seems to me—due to sexist bias—that women don’t know what they’re talking about, is not enough to defeat my justification to believe their testimony (See also (Kelp 2023) for an argument from defeater defeaters against seemings internalism).

**4. Defeat Reliabilism**

Reliabilist theories of justification have been extremely popular in the last three decades and come in a variety of forms, but the gist of the view is that a belief is justified if and only if formed via a (normally) reliable process, or ability. Reliabilism is a theory of *prima facie* justification. As such, in line with normative theories in general, it needs a theory of defeat.

*Alternative Reliable Processes*

The standard reliabilist account of defeat comes from Alvin Goldman:

**The Alternative Reliable Process Account** (ARP): *S*’s belief is defeated iff there are reliable (or conditionally reliable) belief-forming processes available to *S* such that, if *S* had used those processes in addition to the process actually used, *S* wouldn’t have held the belief in question (Goldman 1979).

One can see how ARP is an elegant reliabilist translation of the Pollockian thought that defeat is the kind of normative entity that, when taken in conjunction with the extant epistemic support for the relevant belief, fails to render it justified.

 Bob Beddor (2015) convincingly argues that ARP is both too weak and too strong. Here is his case against ARP’s sufficiency direction:

**Thinking About Unger:**Harry sees a tree in front of him at *t*. Consequently, he comes to believe the proposition TREE: <There is a tree in front of me> at *t*. Now, Harry happens to be very good at forming beliefs about what Peter Unger's 1975 time-slice would advise one to believe in any situation. Call this cognitive process his Unger Predictor […]. What's more, […] whenever it occurs to Harry that Unger would advise him (Harry) to suspend judgement about *p*, this causes Harry to […] suspend judgement about *p*. So if Harry had used his Unger Predictor, he would have come to […] suspend judgement regarding TREE (Beddor 2015, 151).

What this case~~s~~ shows is that ARP is too weak, normatively: contra ARP, for my belief that *p* to be defeated, it is not enough that I *would* change my mind about *p* due to employing some reliable process. Just because I *would* change my mind, it does not follow that I *should* change my mind: defeat is a normative notion.

 Here is Beddor’s case against ARP’s necessity direction:

**Job Opening:** Masha tells Clarence that her department will have a job opening in the fall. Clarence believes Masha; assuming that Masha is usually reliable, Clarence's belief is prima facie justified. Sometime later, Clarence speaks with the head of Masha's department, Victor, who informs him that the job search was cancelled due to budget constraints. Now suppose that Clarence harbours a deep-seated hatred of Victor that causes him to disbelieve everything that Victor says; what's more, no amount of rational reflection would rid Clarence of this inveterate distrust. Consequently, he continues to believe that there will be a job opening in the fall (Beddor 2015, 155).

Clarence would not believe Victor no matter what he says; thus, even though there is an alternative reliable process available to Clarence, it is not one that would have changed Clarence’s mind. Beddor’s Clarence is a lot like my Ann the climate change denier – i.e., a case of irrational defeat resistance. This case shows that ARP is also too strong: just because, in all words, I *would* irrationally and stubbornly hold on to my belief, it does not follow that I *should* do soOnce again, ARP is not normative enough to do the job it is supposed to do.

More recently, Bob Beddor (2021), Peter Graham and Jack Lyons (2021), John Greco (2010), and Chris Kelp (2023) have proposed novel reliabilist accounts of defeat. I will look at them in turn.

*Reasons Reliabilism*

Bob Beddor’s is a reason-based reliabilism: his account incorporates the notion of a reason into the process reliabilist framework; on his account, reasons are the inputs to available reliable processes. Here is Beddor’s reliabilist theory of reasons (RTR):

**Reliable Reasons (Base Clause)**

If*s*is a non-doxastic state of an agent A, and thereis a reliable process available to A which, when given *s* as input, is disposed to produce a belief in *p*, then *s* is a *prima facie* reason for A to believe *p*.

**Reliable Reasons (Recursive Clause)**

If A has a*prima facie*reason to believe*p*, andthere is some conditionally reliable process available to A which, given a belief in *p* as input, is disposed to produce a belief in *q*, then *p* is a *prima facie* reason for A to believe *q*.

RTR is completed by a closure clause according to which nothing else is a *prima facie* reason to believe. The thought then is that with a reliabilistically respectable account of reasons in hand, process reliabilists can simply adopt the best account of defeat from the evidentialist camp. Beddor opts for Pollock’s treatment in particular.

 Beddor’s account promises, to a large extent, to circumvent the problems faced by ARP: since Beddor’s is a reasons-based account, it is essentially normative~~.~~ Unfortunately, though, the reliabilist unpacking of reasons brings back the not-normative-enough problem through the back door. Here is why: on Beddor’s view, reasons are unpacked in terms of inputs to reliable processes. However, reliable processes are not infallible processes: we should expect them to go wrong on occasion. However, if so, we can easily imagine scenarios in which the failure is an input-level failure: the reliable process in question fails by taking up bad inputs – e.g., wishes, biased beliefs etc. While doing so does not speak against the quality of the process – after all, the process remains reliable – we surely don’t want the result that wishes and biased beliefs are epistemic reasons. *Mutatis mutandis*, we don’t want a view of defeat on which wishes and biased beliefs can defeat justification. RTR does not come with enough normative import.

Here is a further worry for Beddor: suppose that, unbeknownst to me, you implant in me a device that has a trove of trivia stored in it (perhaps from the Encyclopedia Britannica). You hook it up to my nervous system in such a way that it will produce in me a belief in a randomly chosen trivium whenever I have a perceptual experience as of a cow. What I have available to me now is a belief forming process that takes experiences as of cows as inputs and outputs trivia. Note that the vast majority of the claims in the Encyclopedia Britannica are true, which means that the process is highly reliable. Since visual experiences as of cows are the inputs to this belief-forming process, RTR predicts that visual experiences as of cows are prima facie reasons to believe any number of trivia from the Encyclopedia Britannica (Kelp & Simion 2024).

I take it that this is manifestly false. Reasons have contents and their contents matter to what they are reasons for. Any account of reasons that doesn’t capture this fact will not capture the normativity of reasons in a satisfactory manner. The problem for RTR is that it doesn’t do so. Moreover, it is not easy to see that it could do so without ceasing to be a genuine version of process reliabilism. After all, if we modify the account such that relations between contents matter to whether something is a reason for believing, it is hard to see that the resulting view will still be a *bona fide* version of process reliabilism.

 The above considerations spell trouble for RTR in general, as a reliabilist theory of reasons. Since Beddor’s account of defeat rests on RTR this is bad news for the account.

*Warrant Reliabilism*

While kindred in spirit, one important difference between Beddor’s reliabilist account of defeat and Graham and Lyons’s is that Graham and Lyons make do without appeal to reasons. More specifically, while Beddor’s view explains defeat in terms of the inputs to reliable belief forming processes, Graham and Lyons focus on the outputs. Here is the key idea:

S has (*prima facie*) warrant for believing p at t iff a cognitive process that satisfies the general theoretical requirements for *prima facie* (doxastic) justification (a) is available to S, and (b) if used at t, taking as inputs only states that S is already in, does or would likely produce p as output.

With this account of *prima facie* warrant in play, Graham and Lyons also propose to adopt a Pollockian account of defeat, whereby the role that reasons play in Pollock are now played by warrants: on this account, a belief is *prima facie* defeated, roughly, when the agent has available to her a “good” cognitive process (the kind that would yield *prima facie* doxastic justification) that would output the content that that belief is false or unwarranted.

The account faces a serious problem with cases of external defeat. Graham and Lyons hold that *prima facie* warrants are the outputs of justification producing and available cognitive processes that ‘[take] as inputs only states that S is already in’. What we get is a form of (conditional) supervenience: any two agents with exactly the same justification producing and available processes can have different warrants only if they are in different states. Graham and Lyons motivate this in the following passage: “We don’t want the fact that S has an unopened encyclopedia nearby to give S warrant for everything written in the encyclopedia.” If defeat turns on warrant, the thought goes, this would make defeat much too easy to come by. The proposed restriction that inputs to the processes on which warrant turns must be internal states is there to avoid this problem.

Unfortunately, this restriction renders the Graham and Lyons account unable to account for external defeat. To see this, let’s return to our toy case in which Bill, the sexist scientist, regularly dismisses testimony from expert women scientists due to his bias. Indeed, he is so convinced they are not worth listening to that, when women talk about scientific matters, he just zones out. On the Graham and Lyons account we get the result that Bill is justified to hold on to his beliefs in spite of women’s testimony. This is the wrong result. For similar reasons, the account will also struggle to accommodate cases of external defeat due to failure in inquiry, such as the case of the uninformed doctor.

*Virtue Reliabilism*

According to virtue reliabilism, justified belief is a species of a broader normative category, i.e., the category of a competent attempt, where a competent attempt is an attempt that is produced by the exercise of a relevant ability (e.g., Greco 2010, Kelp 2018, Miracchi 2015, Sosa 2021). Greco (2010) and Kelp (2023) have developed virtue reliabilist accounts of defeat.[[2]](#endnote-2)

Greco countenances the following subjective justification condition on knowledge and justified belief, alongside the more familiar reliable ability condition:

*S*’s belief that *p* is [subjectively justified] if and only if *S*’s believing that *p* is properly motivated; if and only if *S*’s believing that *p* results from intellectual dispositions that *S* manifests when *S* is motivated to believe the truth. (Greco 2010, 167)

Greco goes on to offer the following account of a defeater for this subjective justification condition:

*S*’s belief that *q* is a *defeater of S’s subjective justification* for believing that *p* if and only if (i) *S* believes that *q*, (ii) prior to believing that *q*, *S* was subjectively justified in believing that *p*, and (iii) when *S* is properly motivated in present conditions, *S* has a disposition to give up believing that *p* upon believing that *q*. (Greco 2010, 167)

Unfortunately, Greco’s account faces one crucial problem: the prospects of unpacking defeat in terms of dispositions to revise beliefs are dim. To see this, consider an agent so stubborn as to never revise their beliefs (even when properly motivated in Greco’s sense). If defeat is analysed in terms of dispositions to revise beliefs, this stubborn individual will never have any defeaters. After all, they simply don’t have the disposition required to have defeaters. But, clearly, that can’t be right. Whether your justification is defeated does not turn on how stubborn you are.

Another problem with Greco’s account is that it also doesn’t accommodate cases of external defeat. The reason for this is that Greco defines defeat in terms of dispositions/abilities to revise beliefs in light of new beliefs. At the same time, one key feature of cases of external defeat is that external defeaters are not believed.

Kelp’s (2023) knowledge-first virtue reliabilist account improves on this front. Kelp observes that defeat can undermine the competence of attempts more generally. If you are about to take a shot in archery and you know that there is a wind blowing from the right, you need to adjust your aim to shoot competently. If you aim for the bullseye in the same way you would were there to be no wind, your shot will not be competent. Your knowledge that there is a wind blowing from the right constitutes a defeater for any shot that aims straight for the bullseye. In this way, Kelp argues, there is some reason for optimism that a viable virtue reliabilist account of defeat can be found. Accordingly, Kelp conceives of defeaters as shifting the ranges of abilities: In the archery case, your acquiring testimonial knowledge that there is a wind blowing from the right is a defeater for your attempt to hit the target by aiming straight at the bullseye, in that it shifts the range of your ability: if you do aim your shot right at the bullseye, your shot will not be competent. Likewise, when the zookeeper tells you that the pen before you is mostly populated by cleverly disguised mules, you acquire a defeater for your perceptual belief that the animal you are looking at is a zebra. As a result, if you hold on to this belief, your belief will not be competent/justified (Kelp 2023, 41).

 In order to account for cases of external defeat, Kelp supplements the standard ability-based virtue epistemological picture with epistemic abilities that are also proficiencies, that is, ways of forming beliefs that not only have the function of generating beliefs that tend to be knowledge, but also of generating knowledge in agents under certain *triggering conditions*. Proficiencies are not attempt-conditional: for instance, the ways of belief formation involved in forming testimonial beliefs are proficiencies in the relevant sense in that they have the function not only of producing beliefs that tend to be knowledge but also of producing knowledge *when being told something*.

 The problem with Kelp’s account, however, is that, on closer inspection, it remains too agent-based to account for external defeat after all. To see this, note that virtue reliabilisms discuss agential abilities and account for epistemic normative assessment in terms of their employment. Since proficiencies are a species of abilities, they will also be agent-based in the relevant sense. Now, here is the dilemma Kelp’s account faces: on one hand, accounting for justification and defeat in terms of proficiencies the agent has is too weak: by stipulation, the sexist scientist can’t get himself to believe women due to his deep-rooted sexism. If he can’t believe women, he doesn’t have the ability to believe them, nor the corresponding proficiency. As such, an account that predicts that cases of defeat are cases in which a proficiency the agent has is malfunctioning, Bill is off the hook, in virtue of lacking the proficiency to begin with. On the other hand, if, in order to fix this problem, Kelp countenances proficiencies the agent should have had, we are back to square one: the should in question is an epistemic should, and it’s hard to see where the source of this normativity lies within a virtue view.

### 5. The Social and the Moral

### *Moral Normativity and Defeat*

The vast majority of the accounts we have been looking at struggle to accommodate defeat data, most notably cases of external defeat. One might wonder, at this stage, whether some of the external defeat cases we have been worried about aren’t really cases of moral failure rather than cases of genuine epistemic failure (Feldman 2000, 190).

Unfortunately, treating external defeat cases as cases that are morally problematic but epistemically permissible will not work, for several reasons. First, on extensional grounds: intuitively, scientist Bill is not *epistemically* justified in his dogmatic belief: he shouldn’t hold it, assert or act based on it, nor should he reason from it to further scientific conclusions. Second, we know from ethics that there is an epistemic condition on moral responsibility (Rudy-Hiller 2002): one cannot be morally blameworthy if they’ve failed morally through no epistemic fault of their own. Since moral blameworthiness requires some variety of epistemic flaw, however, we cannot diagnose the intuition of impermissibility in the sexist scientist as tracking mere moral impermissibility without epistemic impermissibility: if Bill is morally blameworthy for his dogmatism—which he is—he is epistemically blameworthy as well.

Third, note that we can easily remove the sexist bias from the case, and the intuition of impermissibility survives: even if Bill zones out whenever any colleague tells him he’s made a mistake, the defeat datum remains unchanged.

Finally, we can easily imagine parallel cases of defeat that are clear cases of moral success. This suggests that the source of the intuition is, indeed, epistemic failure. Take, for instance, the case of someone who, in spite of all evidence, cannot get themselves to believe their friend committed a crime (Simion 2024): their suspension is morally laudable, but epistemically problematic.

*Social Normativity and Defeat*

Sandy Goldberg (2016, 2018) develops an account of the normativity of defeat that is thoroughly social. One key thought that motivates Goldberg’s project is that social roles—for instance, being a medical doctor—come with normative expectations. These normative expectations may be, and often enough are, featuring epistemic conditions. For instance, Goldberg argues, there is a social epistemic expectation that medical doctors are up to speed on relevant literature in their field.

Another key thought is that to believe that *p* justifiably one must live up to these legitimate expectations. Doctors who fail to be up to speed with the most recent research in their field are not justified in their corresponding beliefs, in virtue of being in breach of the social expectation associated with their role. A doctor who believes that stomach ulcers are caused by stress, in ignorance of the widely available evidence that suggests that it is caused by bacteria, is not justified to believe that ulcers are caused by stress. As such, Goldberg grounds the normativity of defeat in the social expectations associated with the believer in question’s social role.

It is easy to see that Goldberg’s key thoughts also promise to give us the ideal resources to handle cases like that of Bill the sexist scientist: Bill inhabits a social role—that of scientist—that warrants social expectations that he listen to expert testimony on scientific matters. Since he doesn’t live up to these expectations, he does not believe justifiably.

 Here is the main worry that arises for Goldberg’s view: social expectations can be legitimate, but also illegitimate. Women, for instance, are often illegitimately expected to carry most of the household burden, and to underperform in leadership roles. If so, it would seem as though social expectations cannot play the normative grounding role that Goldberg wants them to play. We seem to need further normative notions to help distinguish between epistemically legitimate and epistemically illegitimate social expectations.

 To answer this question, Goldberg appeals to the fact that we are deeply social creatures who are engaged in practices of information sharing and joint action. These practices are supported by a rationale in that opting out of them would be practically irrational for us. This rationale, in turn, grounds our legitimate expectations of each other. To see why this doesn’t work, however, consider a society that has a practice of not trusting women and black people based on a practical rationale having to do with division of labour in information sharing and joint action. Clearly, even in a society like this, gender and race-based epistemic injustice remains epistemically problematic. In sum: since we can easily imagine cases in which what is beneficial for joint action departs from what is epistemically permissible, the view is bound to get cases like these wrong.

Furthermore, according to error management theory (Haselton and Buss 2003, Haselton 2007), the fallibility of human cognition is the result of natural selection.  Note that it is plausible that evolved epistemically deficient practices are beneficial for both biological and social evolution. Indeed, it seems plausible that e.g. relying on heuristics will be beneficial to the aim of information sharing and joint action—due to limited information and computational power. If so, Goldberg’s view will predict epistemic permissibility in cases of intuitive epistemic failure.

Couldn’t Goldberg appeal to his preferred reliabilist account of justification to restrict the range social expectations that give rise to epistemic normativity? Indeed, Goldberg (2018) himself suggests such a restriction in several places. On this view, it is only social expectations that reliably lead to true beliefs that matter for epistemic justification. Unfortunately, this restriction will not do the trick either. Imagine a case where, unbeknownst to Bill, disregarding the word of women is actually a very reliable practice, in virtue of e.g. the fact that patriarchal structures have marginalised women from knowledge acquisition, thereby decreasing their reliability as testifiers on scientific matters. This is a straightforward gender-infused reverse Norman-the Clairvoyant case: *de facto* unreliability lacks normative import. Bill remains unjustified in his dogmatic beliefs.

### 6. Defeaters as Ignorance Indicators

On the account I favour, the main etiological function[[3]](#endnote-3) of our cognitive capacities is to generate knowledge.[[4]](#endnote-4) In turn, in line with functional traits more generally, our cognitive capacities are properly functioning just in case they work in a way that is, in normal conditions, conducive to generating knowledge. When that happens, the beliefs they generate are justified. I dub my view of justification *Knowledge-First Functionalism* (Simion 2019, 2023, 2024). Here is a more precise formulation of the view:

**Knowledge-First Functionalism (KFF)**: A belief is *prima facie* justified if and only if it is generated by a properly functioning cognitive process that has the etiological function of generating knowledge.

 Defeaters are *pro tanto, prima facie* justification lowerers: they are the proper inputs to our processes of belief formation, and when we have enough defeaters, and the processes in question are otherwise properly functioning, the belief at stake (should it be present) is rendered epistemically unjustified. In turn, when our belief formation processes either fail to take up defeaters that they could have easily taken up, or they take them up but fail to update accordingly, they are malfunctioning. The proper function of belief formation processes, then, on my view, is input dependent: failing to take up the right kind of inputs is an instance of malfunctioning (Simion 2023, 2024).

In my view, defeaters are indicators of ignorance (Simion 2023, 2024):[[5]](#endnote-5) they are facts that one is in a position to know, and that lower one’s evidential probability that p is the case:

**Defeaters as Ignorance Indicators:** a fact *d* is a defeater for S’s evidence *e* for *p* iff S is in a~~n~~ position to know *d* and S’s evidential probability that *p* conditional on *e*&*d* is lower than S’s evidential probability that *p* conditional on *e*.

Defeaters are facts. They can be facts about the world around us, or mere facts about a subject’s psychology. Defeaters are facts that are ignorance indicators, in that they increase distance to knowledge: they decrease one’s evidential probability—i.e., the probability on one’s total body of evidence—of *p* being the case. Not just any psychological facts will constitute defeaters to my justification to believe that there is no furniture in the room. My having a perception as of a table will fit the bill in virtue of having the relevant indicator property. The fact that I wish that there was a table in front of me will not fit the bill, even if, unbeknownst to me, my table wishes are strongly correlated with the presence of tables: wishes don’t affect my evidential probabilities. For the same reason, mere beliefs, as opposed to justified and knowledgeable beliefs, will not be defeat material; they lack the relevant indicator property.

Crucially, in my view, rebutting and undercutting defeaters share one and the same central epistemic normative property: they are evidential probability decreasers. What differs is the mechanism by which they achieve this effect: rebutters lower one’s evidential probability for *p* by raising one’s evidential probability for not-*p*. In contrast, undercutters reduce the degree of confirmation that a particular piece of evidence *e* confers on *p –* i.e. the degree to which that piece of evidence probabilifies p (see also (Kotzen 2019) for a detailed formal treatment along these lines).[[6]](#endnote-6)

Note that the account features a being-in-a-position-to-know condition. What is it for me to be in a position to know *e*? Plausibly, a certain availability relation needs to be instantiated. On my view, availability has little to do with the limits of my skull. Defeaters may consist of facts ‘in the head’ or facts in the world. Some facts—whether they are in the head or in the world, it does not matter—are available to me, they are, as it were, ‘at hand’ in my (internal or external) epistemic environment. Some—whether in the head (think of justified implicit beliefs, for instance) or in the world, it does not matter—are not thus available to me.

 My notion of availability tracks a ‘can’ for a species-typical cognizer of the sort exemplified.. First, there are *qualitative* limitations on availability: we are cognitively limited creatures. There are types information that we just cannot access, or process. There are also *quantitative* limitations on my information accessing and processing: I lack the power to process everything in my visual field, it’s just too much information. Finally, there are *environmental limitations* on my availability, having to do with the laws of nature and the social norms governing my environment.

I take this availability relation to have to do with a fact being within the easy reach of my knowledge generating cognitive capacities. A fact *e* being such that I am in a position to know it has to do with the capacity of my properly functioning knowledge generating processes to take up *e*:

**Being in a Position to Know (BPK)**: S is in a position to know a fact *e* if S has a cognitive capacity with the function of generating knowledge that can (qualitatively, quantitatively, and environmentally) easily uptake *e* in cognizers of S’s type.

BPK is relativised to and tracks the limitations of S’s type of cogniser: limitations pertaining to their cognitive architecture and physical social situation. This will result in a generality problem for the view – but not a very pernicious one: we have a strong pre-theoretic grasp, I believe, on the notion. The fact that you can’t know all arithmetical truths is a limitation to your species-specific cognitive architecture. The fact that you can’t get yourself to believe women is not: rather, it is a problem with your sexist cognitive furniture.

It is also important to note that there is a distinction, on this view, between being in a position to know—where nothing else needs to happen for your cognitive capacities to extend into the world and uptake the relevant facts—and being in a position to come to know. The latter is a much more ubiquitous state: I am in a position to come to know what’s happening next door: I can just walk there and check. I am not, however, in a position to know it: the kind of agent that I am can’t see through walls.

Going back to the cases: the view of defeat defended here nicely predicts that the justification of some occurrent beliefs hosted by the characters in the first two of the toy cases is defeated by the presence of ignorance indicators. Take Ann, the climate change denier: the overwhelming expert testimony she receives puts her in a position to know it, and it lowers her evidential probability that climate change is not happening. Ann’s belief that climate change is not happening is rendered unjustified and should be abandoned. Similarly, Bill, the sexist scientist, is in a position to know that there are problems with his experiment from Mary’s testimony, and the latter lowers his evidential probability for p, the result of the experiment. He should correspondingly lower his confidence in p.

How about case three under consideration, the case of doctor Jones?? Recall doctor Jones, who believes that *p* (=stomach ulcers are caused by stress) but missed a recent development in the field that *q* (=stomach ulcers are caused by bacteria). Jones does not justifiably believe that *p*: *q* is a defeater for this belief which undermines Jones’s justification for believing *p*.

First, I would like to invite you to contrast Jones with a layperson, Gerry, who had been told by their doctor back in the day that *p* and still believes that *p*. Despite the fact that there has been a recent development in the field of medicine, Gerry’s belief that *p* continues to be justified. In particular, *q* does not undermine Gerry’s ’s justification for their belief that *p*. Since the central difference between Jones and Gerry is that Jones occupies a certain social role—i.e., they are a doctor—there is reason to think that social roles can be sources of defeat via giving subjects reasons to inquire (Goldberg 2018).

Note, first, that so far, in developing my account of defeat, I have looked only at epistemic functions in individual agents. And while epistemic functions may arise in individual agents, they also arise in broader social systems. It is precisely this idea that will be of central importance in accounting for cases of inquiry-based external defeat.

 To begin with, I take social systems to be systems that feature multiple agents who are connected to one another in at least some ways. The social roles we are interested in are properties of agents in social systems. Being a doctor, teacher, parent, etc. are properties of agents in social systems.

 One interesting feature of social roles is that many of them have constitutive functional properties in that what it is to be an X (doctor, teacher, baker, fireman etc.) is to have the particular function in question (to treat ill people, to teach people stuff, to make baked goods, to put out fires etc.). The functions constitutive of these social roles generate constitutive epistemic norms of proper functioning. By the same token, an account of how social roles may lead to defeat comes into view.

To get an idea of how this might be, let’s take another look at the case of the doctor. Note that having an up-to-date understanding of their field is part of the proper functioning of doctors in the social system that we occupy: it is part of their proper functioning that doctors engage in inquiries into recent developments in the field, as a result of which they maintain an up-to-date understanding of the field, and thereby know how to treat people. That doctors maintain an up-to-date understanding is a key element in the feedback loop that explains the continued existence of this important social role in the social system we occupy. But since maintaining an up-to-date understanding of the field is part of the proper functioning of doctors in our social system, we get a norm that doctors violate if they fail to maintain an up-to-date understanding of their field and, by the same token, if they fail to engage in the inquiries needed to do so. And, of course, the same holds, *mutatis mutandis*, for many other social roles, including teachers, lawyers, academics, and so on. In turn, since the norms in question are generated by the constitutive functions of these social roles, they will be constitutive norms.

It turns out, then, that we can explain normative defeat as a breach of a constitutive norm, sourced in the constitutive function of these social roles. Doctor X violates an epistemic norm associated with proper functioning for this role when they fail to maintain an up-to-date understanding of their field, e.g., by missing the research that indicates that *q*. Since *q* is a reason against believing that *p*, we get the desired result that X’s justification for believing *p* is defeated.

But can’t there be social roles that are functionally constituted by norms that are bad, epistemically? Consider, for instance, the social role ‘judge’ in a judicial system where discrimination based on race is written into the laws of the land: isn’t this account going to deliver the result that judges shouldn’t update based on the testimony of e.g. black testifiers? It will not. To see this, note that one important advantage my account has over Goldberg’s is that epistemic normativity is not encroached upon by social normativity: the epistemic remains an independent normative domain, with its own independent evaluative structure. On my view some genuine epistemic norms—ones associated with promoting epistemic values, such as knowledge—constitute social roles. Compatibly, norms constituting social roles that are not knowledge-conducive—indeed, that are bad epistemically, in that they conflict with norms sourced in the proper functioning of our cognitive system—such as ‘don’t believe black testifiers!’ – are not epistemic norms: they are mere (bad) social norms with epistemic content. In this, my view of epistemic normativity does not bottom out in social normativity, but in epistemic functions having to do with generating knowledge.

**Conclusion**

This chapter has developed a unified functionalist account of defeat. On this view, defeaters are ignorance indicators: they are facts one is in a position to know, and that decrease one’s evidential probability for the proposition at stake. In turn, whether you are in a position to know an evidential probability decreaser depends on the type of cognizer you are: your cognitive architecture – with its species-specific limitations - and your social role.

**References**

Beddor, B. (2015a). Process reliabilism’s troubles with defeat. *The Philosophical Quarterly,* 65(259): 145–59.

Beddor, B. (2021). Reasons for Reliabilism. In Brown, J. and Simion, M. *Reasons, Justification, and Defeat*, Oxford University Press: 146-177

Brown, J. (2018). *Fallibilism: Evidence and Knowledge*. Oxford: Oxford University

Press.

Brown, J. and Simion, M. (2021). *Reasons, Justification, and Defeat*. Oxford: Oxford University Press.

Dutant, J. and Littlejohn, C. (2021). Defeaters as Indicators of Ignorance. Brown, J. and Simion, M. *Reasons, Justification, and Defeat*, Oxford University Press.

Feldman, R. (2000). The Ethics of Belief, *Philosophy and Phenomenological Research* 60 (3): 667-695.

Goldberg, S. (2016). On the Epistemic Significance of Evidence You Should Have Had. Episteme 13 (4):449-470.

Goldberg, S. (2018). To the Best of Our Knowledge: Social Expectations and Epistemic Normativity. Oxford: Oxford University Press.

Goldberg, S. (2021). The Normativity of Knowledge and the Scope and Sources of Defeat. In Brown, J. and Simion, M. Reasons, Justification, and Defeat, Oxford University Press: 18-38.

Goldman, A. (1979). What is justified belief? In: G. S. Pappas (ed.) Justification and Knowledge. Dordrecht: Reidel: 1–25; reprinted in A. I. Goldman (2012). Reliabilism and Contemporary Epistemology. New York: Oxford University Press: 29–49.

Graham, P. and Lyons, J. (2021). The structure of defeat: Pollock’s evidentialism, Lackey’s distinction, and prospects for reliabilism. In: Simion and Brown (ed.) Reasons, Justification and Defeat. Oxford: Oxford University Press.

Graham, P. (2012). Epistemic Entitlement. *Noûs* 46: 449–82.

Lackey, J. (2008). Learning from Words. Oxford: Oxford University Press.

Kelp, C. (2023). *The Nature and Normativity of Defeat*. Cambridge: Cambridge University Press.

Kelp, C. (2018).  *Good Thinking: A Knowledge-First Virtue Epistemology*. New York: Routledge.

Kelp, C. and Simion, M. (2017). Comodious Knowledge. *Synthese* 194, 1487-1502.

Lasonen-Aarnio, M. ( 2014). Higher-order evidence and the limits of defeat. Philosophy and Phenomenological Research, 88(2): 314–45.

Lasonen-Aarnio, M. (2021). Dispositional Evaluations and Defeat. In Brown, J. and Simion, M. ‘Reasons, Justification, and Defeat, Oxford University Press: 93-116.

Lord, E. and Sylvan, K. (2021) Suspension, Higher-Order Evidence, and Defeat. In Brown, J. and Simion, M. ‘Reasons, Justification, and Defeat, Oxford University Press: 116-146.

Millikan, R. (1984). *Language, Thought, and Other Biological Categories*. Cambridge/MA: MIT Press.

Miracchi, L. (2015). Competence to Know. *Philosophical Studies*, 172 (1): 29-56.

Plantinga, A. (1993). Warrant and Proper Function (Oxford: Oxford UP).

Rudy-Hiller, F. "The Epistemic Condition for Moral Responsibility", The Stanford Encyclopedia of Philosophy(Winter 2022 Edition), Edward N. Zalta & Uri Nodelman (eds.), URL = <https://plato.stanford.edu/archives/win2022/entries/moral-responsibility-epistemic/>.

Siegel, S. (2012). Cognitive penetrability and perceptual justification. *Noûs*, *46*(2), 201–222.

Simion, M. (2023). Resistance to Evidence and the Duty to Believe. Philosophy and Phenomenological Research. Online First.

Simion, M. (2024) *Resistance to Evidence*. Cambridge: Cambridge University Press.

Simion, M. (2019). Knowledge-First Functionalism. *Philosophical Issues*  29/1:  254–267.

Simion, M. (2016). Perception, History and Benefit. *Episteme*,  13:  61–76.

Sosa, E. (2011). *Knowing Full Well*. Princeton/NJ: Princeton University Press.

Sylvan, K. and Sosa, E. (2018).  The Place of Reasons in Epistemology.  *The Oxford Handbook of Reasons and Normativity*,  Star Daniel (ed.). Oxford: Oxford University Press.

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

**Notes**

1. See (Pollock 1986) and (Goldman 1979) for the classic accounts, Brown and Simion (2021) for the first full volume on defeat, and (Kelp 2023) for the first book-length treatment. [↑](#endnote-ref-1)
2. See also (Sylvan and Sosa 2018) for an account of reasons that could be developed into a view of defeat, and (Simion 2023) for discussion. [↑](#endnote-ref-2)
3. See also (Graham 2012), (Millikan 1984), Plantinga (1993), (Simion 2017). [↑](#endnote-ref-3)
4. See (Williamson 2000) for the *locus classicus* for knowledge-first epistemology and (Kelp and Simion 2017) for the distinctive value of knowledge. [↑](#endnote-ref-4)
5. Dutant and Littlejohn (2021) also call defeaters ignorance indicators, but the account is spelled out in very different terms: defeaters are evidence that one is not in a position to know. (Gibbons 2013) and (Kelp 2023) also develop an account along these lines. There are two main problems for accounts like these: for the necessity direction: for an agent to have evidence, they need to be able to process the relevant content; many agents that can undergo defeat are not sophisticated enough to have the relevant contents, however. For the sufficiency direction: matters that are intuitively irrelevant to justification can be evidence of not being in a position to know: not finding the thermometer is evidence that you’re not in a position to know that your grandfather has a fever, but it does not defeat your justification to believe that he has a fever (Jenkins-Ichikawa, p.c.). [↑](#endnote-ref-5)
6. This comes in contrast to literature that gives different treatment to rebutting and undercutting defeat and a variety of scepticism about higher-order defeat for knowledge that has been defended by Maria Lasonen-Aarnio (2014, 2021). See (Brown 2018) for discussion. [↑](#endnote-ref-6)