**Defeat**

Mona Simion, Cogito Epistemology Research Centre, University of Glasgow

mona.simion@glasgow.ac.uk

### 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 put forth on the market.[[1]](#footnote-1) This chapter does two things: first, it surveys classical as well as recent accounts of defeat in both the internalist and the externalist camps; 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 adequacy. On this account, defeaters are ignorance indicators: 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[[2]](#footnote-2) 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 with regard to what it is supposed to map on to, between the target of defeat on the one hand, and its nature on the other.

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 XXX, Lackey XXX). 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. One main 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.

Finally, 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 congnizer’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, but she isn’t).

Given all this, in this chapter, 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 zooms out. One day his colleague Mary, who’s 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 traditionally been referred to as normative defeat) that has to do with a failure in inquiry, and in which the subject is not even 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 a 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 on the market, 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 nicely promises to cut across normative domains, in virtue of being framed in terms of reasons; after all, epistemologists hardly enjoy exclusivity on reasons. Second, Pollock’s view makes good on the intuitive thought that defeaters are actualizers of the possibility of a positive normative status to be overridden or undercut. Third, crucially, the account is, in principle, neutral between internalism and externalism about defeat, since it remains 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.

Pollock’s view has been discussed at length in the literature: I will not run through all objections and refinements of the account here. In my view, 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 himself was an internalist evidentialist. One way to spell out Pollock’s view that suggests itself, given his evidentialist leanings, 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 really fast: 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, notably, cases of cognitive penetration will create trouble for a seemings-based defeat account (see e.g. Lyons 2009): 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.

Along similar lines, Chris Kelp (Forthcoming) offers an argument from defeater defeaters against seemings internalism about defeat. Kelp invites us to consider two cases, in which the protagonists - call them X and Y- believe that they are a doctor at the local hospital and that they have just received notice to perform an emergency surgery in OR1. While X and Y are preparing for surgery, someone else – R - tells them that they are not a doctor and asks them to step away from the patient. However, the protagonists believe R to be a notorious drug abuser and disruptive element in the hospital, and thereby dismiss their testimony and hold on to their beliefs that they are a doctor. Here is how the three cases differ, however: X knows that they are a doctor, and they know that R is a notorious drug abuser and disruptive element. In contrast, Y falsely believes R is not trustworthy because they have been brainwashed into thinking this. Note, also, that we may assume that X and Y are psychological duplicates. Internalism in epistemology is in trouble: X and Y are not exactly alike justificationally. After all, X justifiably dismisses the defeater at issue here, whereas Y and Z don’t. Note, also, that in the case of X, in which the defeater is justifiably dismissed, the defeater X has ends up being defeated itself. In other words, the case of X is not only a case of defeat, but also a case of defeat defeat. In contrast, in the case of Y, in which the defeater isn’t justifiably dismissed, the defeater Y has ends up not being defeated itself. In other words, the cases of Y is a case of defeat, but not a case of defeat defeat. But, of course, if this is correct, internalism about defeat is in trouble once more:. After all, since X and Y are exactly alike psychologically, they must be exactly alike in terms of defeat. As a result, internalism about defeat cannot allow that the case of X is a case of defeat defeat, while the cases of Y is a case of defeat without defeat defeat (Kelp Forthcomin, XXX).

**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) is the *locus classicus* for criticism of ARP; if Beddor is right, ARP is both too weak and too strong. Against ARP’s sufficiency direction, Beddor offers the following case:

**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 XXX).

What this cases 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* in a counterfactual world due to employing some reliable process. Just because I would change my mind in world *W*, it does not follow that I *should* change my mind in world *W*: 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 XXX).

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 counterfactual words, I would irrationally and stubbornly hold on to my belief, it does not follow that I *should* do so. Once 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 (XXX), and Chris Kelp (Forthcoming) have proposed novel reliabilist accounts of defeat. I will look at them in turn.

*Reasons-Based 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.

It is easy to see that Beddor’s account promises, to a large extent, to circumvent the problems faced by APR: after all, the main issue with APR was that it was not normative enough. Since Beddor’s is a reasons-based account, it is essentially normative from ground up.

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. Indeed, there is nothing wrong with a reliable processes that fails on occasion: that’s what merely reliable entities do. 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. Consider the following case: Suppose, unbeknownst to me, that you have tampered with my colour vision to produce a slight malfunctioning: I now can no longer distinguish between two shades of yellow. At the same time, you have fitted me with a device that takes as input my ostensible memory that my name is Mona and outputs a belief that my colour vision is currently malfunctioning. Since my colour vision is indeed malfunctioning, this belief forming process is highly reliable. One awkward result of RTR is that my ostensible memory that my name is Mona is a defeater for beliefs formed via colour vision. Another is that any belief I form via colour vision will come out as ultima facie unjustified. Perhaps there is a case to be made that when the process is actually run and produces a belief that my colour vision is currently malfunctioning, I do have a (misleading) defeater for all my beliefs formed via colour vision. Crucially, however, on Beddor’s view, whether the process is actually run is immaterial to whether I have a defeater. All that matters is that the process is available to me. And available it is, all the time. The result that we get, then, is that I have a standing defeater for my beliefs about colour vision. None of them are justified. And that is implausible: Defeat doesn’t come that easily.

*Warrant-Based 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 of 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 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.

Here is one first potential drawback of the view (Kelp and Simion 2024): It looks as though not only beliefs can be defeaters. For instance, if I have a visual experience as of a red dot on a white background and I believe that I am looking at a completely white surface, my belief will not be justified. In particular, my visual experience as of a red dot provides a defeater for my belief that the surface is completely white. Graham and Lyons allow that the outputs of the relevant reliable cognitive processes in question don’t have to be beliefs. Rather, they hold that the outputs are propositions. Now, one might worry that it is a little unclear what it would take for a cognitive process to output a proposition. After all, propositions are abstract entities, and it is not entirely clear what it would take for a cognitive process to output an abstract entity. That said, one can fix this problem by simply holding that the outputs are states with propositional contents that don’t have to be beliefs. Given that belief forming processes may output visual experiences, we can see how Graham and Lyons may be able to handle the above case after all.

At the same time, one question that arises at this stage is exactly what sorts of states can be the outputs of belief-forming processes. For instance, we may wonder whether belief forming processes can output desires. Note that if the answer here is yes, the Graham and Lyons account is in trouble: the fact that I might desire that p, where this desire is produced by a reliable belief-forming process, does not mean that p can now be a defeater for some other belief I hold. More generally, the point here is that once Graham and Lyons allow that reliable belief-forming processes can output states that aren’t beliefs, and once we recognise the fact that not all states can serve as defeaters, they face the question as to why it should be that belief forming processes can output only the kind of states that can serve as defeaters.

Graham and Lyons briefly consider this worry in a footnote and credit it to Bob Beddor. Here is what they say:

Could a belief-forming process produce as outputs states that aren’t at all belief-like? If an alternate reliable process produced the desire that p, it doesn’t seem like this would be a defeater for the belief that ~p. *Maybe we need to restrict the outputs to states with assertoric force or which are in some other way sufficiently belief-like.* We won’t try to sort out the details here (Graham and Lyons XXX).

While the kind of restriction strategy that Graham and Lyons envisage may look promising at first glance, I have some serious reservations. To get them into clear view, note that the difference between states that can serve as defeaters and states that can’t is a normative difference: the ones that can have the normative power to defeat, the ones that can’t do not have this power.

On a different note: The account also faces a 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 incapable 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 zooms out. One day his colleague Mary, who’s 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 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. On the Graham and Lyons account we get the result that Bill is perfectly justified to hold on to his belief in spite of Mary’s testimony. This is overtly the wrong result here.

Maybe Graham and Lyons can dig in their heels and claim that there is some internal state with the relevant properties that is present in the case of Bill after all – maybe Bill hosts some experiences as a result of Mary’s testimony, and they can appeal to those to explain the defeat intuition. Two things about this: first, note that it is up to the objector the construct the case as they wish: the way I envisage it, Bill is such a terrible sexist that he completely zooms out when Mary speaks. Now, here is one result Graham and Lyons don’t want: the more we build up the sexism in the case, the more plausible it is no relevant experience is present, and thereby their account will predict no defeat and presence of justification. Clearly however, increased sexism should not make you a better epistemic agent; this is overtly the wrong result.

Finally, and for similar reasons, the Graham and Lyons account will also struggle to accommodate cases of external defeat due to failure in inquiry, such as the case of the uninformed doctor.

*Virtue-Based 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 2017, Sosa 2021). Recently, Greco and Kelp have developed virtue reliabilist accounts of defeat.[[3]](#footnote-3)

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 (Kelp & Simion 2024).

It might be suggested, on behalf of Greco, that while the stubborn agent does have a disposition to revise their beliefs, they ought to have it. Accordingly, perhaps Greco can analyse defeat in terms of dispositions agents ought to have rather than in terms of dispositions agents do have. To see the trouble with this proposal, however, note first that the ‘ought’ here will have to be an epistemic ought. Crucially, the key claim of ability reliabilism is to explain epistemic normativity in terms of reliable abilities, where abilities in turn are taken to be dispositions. The problem is that it is hard to see how the epistemic ought at issue in dispositions one ought to have could plausibly be analysed in terms of reliable abilities/dispositions.

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.

Again, one might think that this problem can be overcome by tweaking the account. In particular, one proposal that may appear promising is to abandon the idea that defeat requires the acquisition of new beliefs and instead hold that all that is needed are new beliefs one ought to have formed. But, again, the ‘ought’ at issue here will have to be an epistemic ought. At the same time, according to ability reliabilism, epistemic oughts are to be explained in terms of reliable dispositions to succeed (get to truth) upon attempting (believing). If beliefs are attempts, then a belief that one ought to have but doesn’t is an attempt that one ought to have made and didn’t. The trouble is that it is hard to see how virtue epistemology, construed as the normativity of attempts one makes, can account for attempts that *should* have been made.

Kelp’s (Forthcoming) 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 Forthcoming 41).

In order to account for cases of external defeat – in which no belief is formed - , 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*. In this, 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 mere abilities+, they will also be agent-based in the relevant sense. Now, here is the dilemma Kelp’s account is faced by: 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 to begin with. Indeed, many of the champions of the views above will appeal to exactly this reply.

Here is, for instance, Richard Feldman, putting the point succinctly:

It's surely true that there are times when one would be best off finding new evidence. But this always turns on what options one has, what one cares about, and other non-epistemic factors. As I see it, these are prudential or moral matters, not strictly epistemic matters (2004, 190).

Unfortunately, treating external defeat cases as cases that are morally problematic but epistemically permissible will not fly, for several reasons. First, on mere 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, in terms of theoretical adequacy: we know from the extensive literature on blameworthiness that there is an epistemic condition on moral blame: one cannot be morally blameworthy if they’ve failed morally through no epistemic fault of their own: should, for instance, Bill have done all proper inquiry in the issue and not discovered any evidence that his experiment was flawed, and should e.g. his belief that p have resulted in actions that led to extensive moral harm, Bill would be properly diagnosed as morally blameless. 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 zooms 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 plausibly cases of moral success. This suggest that the source of the intuition is, indeed, epistemic failure (absent other normative constraints at the context). Take, for instance, the case of Lucy, the optimist spouse (Simion 2023): When her partner Dan spends more and more evening hours at the office, she’s happy that his career is going so well. When he comes home smelling like floral perfume, she compliments his taste in fragrance. Finally, when she repeatedly sees him having coffee in town with his colleague Alice, she is glad he’s making new friends. Is Lucy justified to believe as she does that Dan is a faithful, loving husband? Clearly not. Note, however, that it’s hard to find moral flaws with Lucy’s epistemic ways: after all, many moral philosophers (and a good number of epistemologists) agree that we owe more trust to our friends and family than to people we have never met:[[4]](#footnote-4) if so, Lucy’s suspension is morally laudable, but epistemically problematic.

*Social Normativity and Defeat*

In extensive work, Sandy Goldberg has taken up the task of developing 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 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: this guy 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 social expectations, but also, illegitimate social expectations. 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 get the problem with Goldberg’s account into view, notice first that, since on this normative picture epistemic norms are grounded in social expectations, which, in turn, are grounded in cooperation-generated rationales, the scope of epistemic normativity only reaches as far as our rationale-supported practices of information-sharing and joint action. This is a theoretically heavy burden to carry: it amounts to a claim that epistemic normativity strongly co-varies with a particular subset of practical normativity.

To see why this is a problem, consider a society that has practices of sharing information and acting jointly on a wide range of issues. Suppose furthermore, these practices are supported by a rationale in the way envisaged by Goldberg. But now suppose that this society also has a practice of not trusting women and black people. Clearly, this practice not only fails to give rise to epistemic expectations, it gives rise to bad epistemic expectations. Note, also, that one does not even have to come up with very far-fetched examples to illustrate this point. We do, as a matter of fact, live in a world where many societies have a practice of disbelieving women and people of colour. We can imagine that one might even come up with a practical rationale for these practices – having to do, e.g., with division of labour. Nevertheless, gender and race-based epistemic injustice remains epistemically problematic.

Couldn’t Goldberg say that this practice is not, as a matter of fact, supported by a cooperative rationale? Unfortunately, the answer here is no. The reason for this is that normativity is modal: even if this practice is not, as a matter of fact, supported by a rationale, insofar as the practice *may* be supported by a rationale - in that it may be practically irrational to opt out of it - , it may, on Goldberg’s view, generate legitimate social expectations. 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 wrong.

Furthermore, research in cognitive psychology (e.g. Kahneman et al. 1982, Nisbett and Ross 1980, Gilovich et al. 2002) notably indicates that human beings tend to rely on heuristics when engaged in probabilistic reasoning, heuristics making people prone to commit elementary probabilistic fallacies. Also, according to error management theory (Haselton and Buss 2003, [Haselton 2007](https://www-oxfordscholarship-com.ezproxy.lib.gla.ac.uk/view/10.1093/acprof:oso/9780199674800.001.0001/acprof-9780199674800-bibliography-1" \l "acprof-9780199674800-bibItem-200)), the fallibility of human cognition, at least in many cases, is the result of natural selection. Evolutionary psychologists argue that, given the limited information and computational power with which organisms must contend, an inference mechanism can be advantageous if it often enough draws accurate conclusions about real-world environments, and does so quickly and with little computational effort (2002: 152). The heuristics humans rely on in probabilistic reasoning, some of these psychologists maintain, are mechanisms of just that sort.

Note that it is plausible that these evolved epistemically deficient practices are beneficial for both biological and social evolution – otherwise, it seems implausible they would have been selected to begin with. Indeed, it seems plausible that relying on heuristics like those above 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 all of these cases of intuitive epistemic failure.

To put the worry in more theoretical terms, here is the problem: if our model predicts that epistemic norms are grounded in the rationality of our practices of information sharing and joint action, and the latter are (very plausibly) aimed at the survival of our species, then our model predicts that epistemic norms will track survival norms: our belief producing processes, for instance, need only be as reliable as needed for survival. However, there is nothing to ensure that the socially and biologically-set reliability threshold will coincide with the epistemically-needed reliability threshold. That is, the threshold of reliability required for epistemic purposes may well be higher than what is needed for our practices of information sharing and joint action, and in turn for biological benefit. Socially and biologically reliable enough need not coincide with epistemically reliable enough. Similarly, epistemic norms for sensitivity to evidence and for evidence gathering may set the threshold for epistemic permissibility differently than social and practical norms for joint action and survival.

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 himself alludes to such a restriction in several places (e.g. XXXX). 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: reliability is not sufficient for justification. 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. This is a straightforward gender-infused Norman-the Clairvoyant case: de facto unreliability lacks normative import in isolation. Still, clearly, Bill the sexist is not justified to disregard the expert testimony by Mary.

Goldberg is aware of these difficulties. In more recent work (Goldberg In Progress), in response to these problems, he develops a novel, moral-normativity-infused account of epistemic justification.[[5]](#footnote-5) Here is how the view goes: according to Goldberg, in cases like the above, in which disbelieving a testifier has a high potential for moral harm, hearers shoulder a moral obligation to inquire further before settling into a particular doxastic attitude. This moral obligation, when breached, affects epistemic justification.

Recall, however, that we have already seen, in the previous section, that appealing to moral justification to explain these cases will not work, for a variety of reasons. Maybe most importantly indeed because some of the problem cases can be spelled out as cases of moral success (remember Lucy, the overly trustworthy spouse). If so, switching from practical to moral normativity will not help isolate cases in which epistemically legitimate social expectations deliver external defeat.

### 7. Defeaters as Ignorance Indicators

On the account I favour, the main etiological function of our cognitive capacities is to generate knowledge. In turn, in line with functional traits more generally, they 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): the account is functionalist in that it follows people like Burge (2010), Graham (2012) and Millikan (1984) in taking the epistemic normativity of belief to drop out of the epistemic function of our cognitive capacities. It is knowledge-first epistemological (Williamson 2000) in that, unlike traditional, truth-first functionalism, it unpacks the function at stake in terms of knowledge. 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.

On this knowledge-centric picture, good belief is knowledgeable belief, while justified belief is belief generated by a properly functioning cognitive process that has the etiological function of generating knowledge. The standards for proper functioning are thus natural normative standards, and they are constitutively associated with promoting knowledgeable beliefs.

Now, I want to start by noting that defeaters constitute epistemic oughts: when our justification is defeated, by definition, it is impermissible to ignore defeat and hold on to the corresponding doxastic attitude. Since it is impermissible to ignore defeaters, it follows that they constitute epistemic oughts (since it is always permissible to ignore mere permissions).

What grounds the epistemic ought carried by defeat, in my view, is proper epistemic functioning. 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. One illuminating analogy here is the proper functioning of the lungs: what it is for our lungs to function properly is, partly, for them to take up the right amount of the right stuff, i.e. oxygen, from the environment. Lungs that fail to do so, in environments where oxygen is easily available, are improperly functioning. Similarly, our cognitive capacities are malfunctioning if failing to take up easily available inputs, i.e. easily available evidence and defeaters (Simion 2023).

In my view, defeaters are indicators of ignorance:[[6]](#footnote-6) 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 an 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*.

Or, slightly more formally:

**Defeaters as Ignorance Indicators:** a fact *d* is a defeater for S’s evidence *e* for *p* iff S is in a position to know *d*, and P(*p*/*e*&*d*) < P(*p*/*e*).

Defeaters are facts. They can be facts about the world around us, or mere facts about a subject’s psychology. My having a perception as of a table in front of me is a psychological fact; it (*pro tanto, prima facie*) defeats my justification to believe that there is no furniture in the room. So does the fact that there is a table in plain view in front of me.

Defeaters are facts that are ignorance indicators, in that they enhance 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 have a perception as of there being a table in front of me is a fact that I am in a position to know and that decreases my evidential probability that there is no furniture in the room. 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* (see also (Kotzen 2019) for a detailed formal treatment along these lines). This comes in stark contrast to literature that gives different treatment to first- and higher- order evidence, or rebutting and undercutting defeat. Most centrally, my unified account outright opposes a variety of scepticism about higher order defeat for knowledge that has been introduced and made popular by Maria Lasonen-Aarnio’s work on the topic (e.g. XXX).

I think this is the right result: undercutters and rebutters afford a unified treatment, and we should, all else equal, prefer this unified treatment on grounds having to do with theoretical adequacy. Here are some quick reasons why scepticism about undercutting defeat doesn’t work: One’s evidence comes with a having relation and an evidential support relation (aka degree of confirmation). Plausibly, one’s confidence in p should match the degree of confirmation that one’s evidence offers to p. Higher order evidence/undercutting defeat works via raising/lowering the degree of confirmation that first order evidence provides to p.

Now, here is a case of higher order evidence that increases the degree of confirmation of the first-order evidence: I believe p based on my neighbor’s George’s testimony (alternatively, I have e.g. .8 credence that p). Mary tells me that George is the top expert in the world on the matter. Her testimony is evidence that q: ‘George’s testimony gives very high support to p’. In probabilistic terms: if George’s original testimony probabilifies p to x, Mary’s testimony translates roughly as ’George’s testimony probabilifies p to y & y>x (how high y is will depend on Mary’s epistemic credentials). So now that Mary has spoken, I am in a position to know that r: I have a .8 credence that p based on George’s testimony and the probability that p conditional on George’s testimony is .9. Intuitively: I should revise to .9

Moving on to a corresponding case of undercutting defeat: I believe p (indeed, I know p) based on my neighbor’s George’s testimony (alternatively I host a e.g. .8 credence). Mary tells me that George is a well-known liar on p-issues: Her testimony is evidence that q: ‘George’s testimony gives lower support to p’. George’s original testimony probabilifies p to x; Mary’s testimony, just like before, amounts to: ’George’s testimony probabilifies p to y & y<x (it will depend on Mary’s epistemic credentials how low y is). Just like before, I am in a position to know that r: I have a .8 credence that p based on George’s testimony and the probability that p conditional on George’s testimony is .4. Intuitively, I should revise to .4. Undercutting defeat scepticism is wrong. Defeat affords unified treatment.

Moving on to further unpacking the view: 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 psychological ‘can’ for an average cognizer of the sort exemplified – i.e. with the relevant kind of cognitive architecture. First, there are *qualitative* limitations on availability: we are cognitively limited creatures. There are of *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 (with S’s cognitive architecture).

It is important here to note that there is an important 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 net 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.

How does this view explain that there is such a thing as an ought to take up defeat governing our belief formation? After all, on the account proposed, defeaters are ignorance indicators – i.e. facts that lower evidential probability for one. Why should a system with the function of generating knowledge be under an obligation to take up ignorance indicators, which, by stipulation, enhance distance to knowledge?

Two things about this: first, note that defeaters are not merely ignorance indicators for *p*, but also either knowledge indicators for not-*p* (rebutting defeaters), or knowledge indicators for ‘Piece of evidence e does not confirm/offers weaker confirmation for p’ (undercutters). This, in turn, suggests that defeaters exercise normative pressure on our cognitive systems in virtue of them also being knowledge indicators.

Second, and furthermore, I think that there is more to the normative pressure of defeat, than this merely general evidential normative pressure; in particular, I think that the normative pressure of defeaters for my belief that *p* can also be explained in a *p*-centric fashion. Here it goes: consider, for starters, a case in which I know that *p* at t1, and (misleading) evidence that not-*p* shows up in my environment at t2. In a situation like this, crucially, whether I take up the defeating evidence or not, my knowledge that *p* is defeated: I am now left, at best (depending on the relevant evidential weights), with a somewhat - but not fully - justified true belief that *p*. My full belief has thereby been rendered impermissible: I now hold a doxastic attitude that is stronger than what the evidence affords. Note, also, that this is so even if what I started with is not knowledge, but rather, e.g., a justified credence of .6. Now that defeat is available in my environment, whether I take it up or not, my .6 credence is no longer permissible, since my level of justification has been lowered by defeat.. What is it that my cognitive system ought to do now that it’s hosting an impermissible doxastic attitude? The answer that suggests itself is: abandon it, either altogether (if defeat is full defeat), or in favour of the weaker attitude that remains supported by the evidence.

Note, though, that abandoning the (now) impermissible doxastic attitude in conjunction with not taking up the relevant defeaters seems irrational: if I, at the same time, (mistakenly) continue to take my evidence to support a .6 credence, and adjust to a .5. credence, something has gone amiss, rationality-wise: I cannot be justified to believe that *p* based on a total body of evidence E if E includes evidence I don’t take up. In sum: it would seem as though, if I don’t take up the relevant defeat from the environment, I am faced with a normative dilemma: either I hold on to an impermissible doxastic attitude that is no longer supported by my evidence, or I hold a novel doxastic attitude that enjoys propositional justification but not doxastic justification, for lack of proper basing. Grasping neither of these horns is knowledge conducive. If so, it would seem, the only path left for knowledge conduciveness is taking up the relevant defeaters, and adjusting my doxastic attitude in light of them. That’s how the view explains why our knowledge-generating cognitive capacities ought to take up defeat.

Going back to the cases: It is easy to see that 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 we started with 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? Recall doctor Jones, who believes that *p* (=somach ulcers are caused by stress) but missed a recent development in the field that *q* (=stomach ulcers are caused by a 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 (Kelp & Simion 2023).

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.[[7]](#footnote-7)

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.). To see this, note that we cannot even fully understand the roles in question without understanding their functions: To fully understand what a doctor is you have to understand what the function of a doctor is, i.e. to treat ill people. It is easy enough to see that the same is true of a whole host of other social roles, including teachers, bakers, firemen, among many others.

Given that many social roles have constitutive functional properties, the prospects for an analysis of epistemic norms constituting these social roles already look bright. The question that we need to consider is: are the functions constitutive of these social roles generating any constitutive epistemic norms of proper functioning? If the answer to this question is yes, the route to an account of epistemic norms constitutive of social roles is a going to be a short one. By the same token, an account of how social roles may lead to defeat may come 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 normal functioning of doctors in the social system that we occupy. More specifically, it is part of normal 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. In fact, 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:[[8]](#footnote-8) doctors’ understanding informs their treatment practices, the fact that it is kept up to date enhances their success rate of these treatments, which in turn explains why the social role of doctor continues to exist in our social system. But since maintaining an up-to-date understanding of the field is part of normal functioning of doctors in our social system, we get the by now familiar normative import. It is thereby part of proper functioning. This, in turn, means that 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, it is easy enough to see that 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. Since our doctor X is a practicing doctor, they occupy the social role of doctor. As a result, they violate 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*. In this way, it is epistemically proper for them to believe that *q*. And since *q* is a reason against believing that *p*, we get the desired result that X’s justification for believing *p* is defeated (Kelp and Simion 2023).

But won’t my account suffer from a problem parallel to Goldberg’s view? Recall that Goldberg wanted to explain evidence one should have had in terms of social expectations. Recall, also, that we said that epistemically illegitimate social expectations make problems for Goldberg’s account: if epistemic normativity reduces to social normativity, bad social expectations cannot be further explained in epistemic normative terms.

Won’t my account have the same problem, in virtue of appealing to social roles? 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 our 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 it does not ground epistemic norms in social norms, and, in 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- 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, our view of epistemic normativity does not bottom out in social normativity, but in epistemic functions having to do with generating knowledge.

**Conclusion**

This chapter developed a functionalist, unified 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. What differs between rebutters and undercutters 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.

**References**

1. See (Pollock XXX) and (Goldman XXX) for the classic accounts, Brown and Simion 2021) for the first full volume on defeat, and (Kelp Forthcoming) for the first book-length treatment. [↑](#footnote-ref-1)
2. See (Dutant and Littlejohn 2021) for a more detailed taxonomy. [↑](#footnote-ref-2)
3. See also (Sosa and Sylvan xxx) for an account of reasons that could be developed into a view of defeat, and (Sosa 2021) for an account of epistemic negligence that could serve a similar purpose. See also (Simion 2023 and 2022 for respective criticisms of these proposals). [↑](#footnote-ref-3)
4. REFERENCES [↑](#footnote-ref-4)
5. I have argued often in p.c. that the view is a variety of moral encroachment, but since it is not a threshold view (like e.g. the Basu & Schroeder classic), in order to avoid confusion, I avoid this label here. [↑](#footnote-ref-5)
6. Dutant and Littlejohn (2021) put forth an account of defeaters as ignorance indicators, but spell it out in very different terms. I think their account doesn’t work. On their account, defeaters are evidence that one is not in a position to know. (Gibbons 2013) also develops an account along these lines; see also (Kelp Forthcoming) for a similar knowledge-based view. The account is further spelled out in probabilistic terms – embracing the thesis that it is rational for one to believe that p just in case the probability that one is in a position to know p is sufficiently high. The are two main problems for accounts like these: the first is structural: these accounts are epistemically second order, in that defeaters are evidence that some epistemic status is missing. But for an agent to have this 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. The second problem parallels a wrong-kind-of-reasons problem: matters that are intuitively irrelevant to justification can raise and lower the probability of one being in a position to know. Here is a case from Jonathan Jenkins-Ichikawa (p.c) that convinced me that the probability of knowing and being in a position to know is normatively irrelevant to justification (and, m*utatis mutandis*, defeat): Say that your grandfather is not feeling well and you’re searching for the thermometer to check whether he has a fever. Now, finding the thermometer increases the probability that you are in a position to know that your grandfather has a fever, and, indeed, the probability that you will come to know that he does. It is also plausible, indeed, that it constituted evidence that you are in a position to know that your grandfather has a fever. However, clearly, it does not affect the rationality of your corresponding belief. Conversely, not finding the thermometer decreases the likelihood of your being in a position to know/coming to know that your grandfather has a fever, and it is evidence that you’re not in a position to know that he does, but it surely does not defeat whatever justification you might have had to believe that he does have a fever. [↑](#footnote-ref-6)
7. Note that the existence of agents who occupy social roles need not imply the existence of other agents. We might want to allow that a doctor continues to be a doctor even if they are the only human being left alive. However, it does imply the existence of other agents at least at some point in the past. If there had only ever been one human being around, this person could not have been a doctor. At best, they may have engaged in healing practices, but they couldn’t have been a doctor. We take this to be independently plausible. After all, social roles are just that: *social* roles. [↑](#footnote-ref-7)
8. These are, of course, not essential features of doctors, but merely constitutive features: doctors remain doctors even when they are in breach of some of the norms constitutive of their social role. This is a general feature of functionalist normativity: hearts remain hearts even when they malfunction. It is also a general feature of constitutive normativity more generally: one can break several constitutive norms of a game or language and still count as playing the game/speaking the language. Note, though, too widespread a breach of constitutive norms with maximal systematicity will lead to discontinuing the constituted activity: if I only utter kakakakaka I don’ count as speaking English. Similarly, someone who: doesn’t have a medical degree, lacks any understanding of medicine, etc. will not count as being a doctor due to being in too widespread a breach of the constitutive norms of the role. [↑](#footnote-ref-8)