1 Introduction

Suppose that you are preparing to go and see a show by your favourite band. You have bought tickets months ago and, finally, the big day has arrived. You have double- and triple-checked the dates and times. You believe that the show is happening today. Clearly, your belief is justified. Just as you are about to leave, you are notified that the show is cancelled because one of the band members has fallen ill. (In what follows, I will refer to this case as 'the cancellation case'.)

Suppose, next, that you are a detective. You have excellent evidence that a certain suspect committed the murder you are investigating: the suspect's fingerprints are on the murder weapon, their DNA is found at the crime scene, and they have an excellent motive for the deed. Based on the evidence, you come to believe that the suspect did indeed commit the crime. Again, it is hard to deny that your belief is justified. A little later, you find out that the suspect has an ironclad alibi for the time of the deed. There are eyewitnesses and video surveillance placing them at a different location ('the alibi case').

These are paradigm instances of what epistemologists call *defeat*, that is, the loss of justification for believing something in light of new information. Defeat will take centre stage in this Element. Its central aim is to work towards developing a novel account of defeat.

It is easy enough to see that achieving this aim is of considerable importance. We live in a world in which an abundance of information is just clicks away. As a result, we now have unprecedented opportunities to acquire new information. But since defeat is the loss of justification in light of new information, by the same token, we have unprecedented opportunities to acquire defeat for our justified beliefs. The expansion of our informational horizon that we have witnessed is by the same token an expansion of potential defeat for the justification of our beliefs. Many of our most deeply held beliefs about the likes of politics, religion, and ethics may be most affected. This is because the amount of information supporting claims that run contrary to our beliefs is particularly large. If these beliefs are indeed defeated, the range of our justified beliefs is much smaller than we may have thought (cf. Frances 2005). What's more, it is widely agreed that in order to permissibly assert our beliefs, we must at least have justification for them (Kelp and Simion 2021; Lackey 2008; Williamson 2000). As a result, the range of beliefs we may assert threatens to turn out to be much smaller than we may have assumed, and our practices of conducting debates about them may need to be revised dramatically. Since a viable account of defeat will offer precise conditions under which our justified beliefs suffer from defeat, such an account will be key to ascertaining how great the threat to

our justified beliefs really is. What's more, it will provide an invaluable tool for understanding the extent to which we may need to revise the practices for debating our beliefs.

I said that defeat is the loss of justification for believing something in light of new information. I hope this characterisation of defeat is not only useful but also theoretically lightweight in that it is acceptable to most contributors to the debate on defeat. In fact, I think that the lightweight characterisation of defeat can be fleshed out a little further. When defeat occurs, what happens is that the new information leads to the loss of justification *by providing reason against holding certain beliefs.* In the cancellation case, the information that the show has been cancelled provides a reason against believing that the show is happening tonight. And in the alibi case, the information provided by the eyewitness testimony and surveillance videos is a reason against believing that the suspect committed the crime.

Note that in both cases, the information provides reason against holding the beliefs in question by providing reason for thinking that the beliefs are false. In the cancellation case, you acquire reason for thinking that the show is *not* happening tonight, and, in the alibi case, you acquire reason for thinking that the suspect is *not* guilty. Can we refine our lightweight characterisation of defeat by saying that what happens when defeat occurs is that new information leads to the loss of justification by providing reason for thinking that the target belief is false? No. While acquiring reason for thinking that some of our beliefs are false is one important way of acquiring defeat, it is not the only one. To see this, consider the following case.

You are looking at a surface in front of you, which looks red. Based on this, you form a belief that the surface is indeed red. This is yet another case in which your belief is justified. Shortly after, I tell you that the surface is illuminated by red light and, as a result, would look red to you even if it were white ('the illumination case').

In this case, you do not acquire a reason for thinking that your belief is false. After all, even if the surface is illuminated by red light, the question as to whether it is red remains entirely open. The table may be red illuminated by red light. Rather, what you learn is that your source, that is, visual perception, is inadequate to give you justification for believing that the surface is red. After all, visual perception doesn't allow you to tell whether it is red or white.

This is why we will do well to stick to the more inclusive characterisation of defeat in terms of information that provides a reason against holding certain beliefs. Note that this characterisation can accommodate both kinds of defeat. After all, both provide us with reason against holding some of our beliefs, one

by providing reason for thinking that the beliefs are false, the other by providing reason for thinking that their sources are inadequate.[1]

Now, there is lots of new information out there that provides reason against holding various of our beliefs. Suppose someone you have never met or even heard of and who is currently in a faraway place has just asserted that your employer will be unable to pay their employees next month ('the employment case'). That someone made this assertion is information that provides reason against holding the belief that you will receive your pay as per usual. But it does not defeat the justification you have for your belief that your employer will continue to pay your salary. Why not? The answer is that for new information to defeat some justification you have, you must be related to it in the right way.

The question about this relation is one of the key divides in the epistemological debate on defeat. First, one might think that the relation is psychological (henceforth also 'the psychological view'). On this view, for new information that provides reason against holding some belief to defeat some justification one

---

[1] Consider the process reliabilist 'alternative reliable process' account of defeat (Goldman 1979; Lyons 2009). Very roughly, according to process reliabilism, whether you believe justifiably turns on whether your belief is produced via reliable cognitive processes, that is, processes that tend to produce true beliefs. And whether your justification for believing something is defeated turns on whether you have an alternative reliable process available such that were it to be used, you would not hold your belief. Crucially, defeat is explained purely in terms of processes, reasons don't feature at all here. Note also that this is no accident. Process reliabilists take pride in the fact that their account of justification is naturalistically respectable. Part of what makes it so is that it features no normative properties such as reasons. In light of this, one may wonder just how lightweight the characterisation really is.

  Three comments on this. First, even though process reliabilists don't state their account of defeat in terms of reasons, it doesn't follow that their account of defeat isn't compatible with the lightweight characterisation of defeat. After all, it may be that reasons can be analysed in terms of available reliable processes. Note that, if so, the process reliabilist account achieves a nice fit with the lightweight characterisation. After all, the lightweight characterisation isn't meant as a substantive account of defeat and leaves open the possibility that the key property of reasons against believing admits of further analysis, including along process reliabilist lines. Second, defeat is a general normative phenomenon: it doesn't only occur in the epistemic domain. At the same time, it is far from clear that, in all normative domains in which defeat occurs, justification and defeat can be unpacked along process reliabilist lines. (For instance, process reliabilism is structurally a kind of rule consequentialism. While this may be plausible for the epistemic domain, it is not clear that it is equally plausible, for example, for the practical domain, which may require an act consequentialist treatment.) If it cannot, the prospects for a fully general account of defeat in terms of alternative reliable processes are dim. What's more, we may just have to revert to a general account in terms of reasons, perhaps with a process reliabilist account of what it takes to have reasons in the epistemic domain. Third, to the best of my knowledge, the alternative process account of defeat is the only account that doesn't naturally fit with the lightweight characterisation of defeat. At the same time, it has come under heavy criticism (some of which I will review in Section 3). And in order to develop a viable account of defeat, process reliabilists have started to try and make room for reasons (or something in the vicinity) in their epistemology. In light of this, no matter whether literally everyone is on board with the lightweight characterisation, I take it that it is lightweight enough to provide a useful starting point for theorising about defeat.

has for this belief is for one to stand in a to-be-specified psychological relation to the information. For instance, one could say that the relation is belief. In that case, some new information that provides reason against holding some belief defeats some justification one has only if one believes it. The reason why the assertion about your employer doesn't defeat your justification for your belief about continued pay is that you do not believe this piece of information.

The main rival to the psychological view takes the relation to be normative (henceforth also 'the normative view' (e.g. Alston 2002; Graham and Lyons 2021; Simion 2021)). On this view, for new information that provides reason against holding some belief to defeat some justification one has for this belief is for one to stand in a to-be-specified normative relation to the information. For instance, one could unpack this relation in terms of epistemic propriety to believe. In that case, some new information that provides reason against holding some belief defeats some justification one has only if it is epistemically proper for one to believe this piece of information. The reason why the assertion about your employer doesn't defeat your belief about continued pay is that it is not the case that it is epistemically proper for you to believe this piece of information.[2]

We have looked at some cases of defeat and at a lightweight characterisation of the phenomenon in terms of loss of justification in light of new information, where the new information provides reason against holding certain beliefs. With these points in play, the phenomenon that takes centre stage in this Element should now be in clear view. Before getting down to more substantive theorising about defeat, I'd like to look at some kinds of defeat that are familiar from the literature. Surveying the myriad ways in which defeat may occur will allow us to get an even better handle on the phenomenon. It will also give me the opportunity to highlight some important allegiances and flag some important assumptions that I will make in the remainder of this Element.

We have already encountered the distinction between rebutting and undercutting defeat (e.g. Pollock 1986). Very roughly, rebutting defeat involves a reason for thinking that some belief we hold is false (the cancellation and alibi cases), and undercutting defeat involves a reason for thinking that the source of some beliefs is inadequate (the illumination case). The distinction between undercutting and rebutting defeat is widely endorsed in the literature and I will take it on board for present purposes.

---

[2] Note that it is possible to have a conjunctive view on which the relation must be both psychological and normative (e.g. what's needed is belief that it is epistemically proper for one to hold), or a disjunctive view on which a psychological relation is sufficient, and a normative relation is also sufficient.

Another important distinction is between misleading and non-misleading defeat. Very roughly, misleading defeat involves information that provides reason against holding a certain belief by providing evidence for a falsehood. To see what I have in mind, consider a version of the illumination case in which my statement that the surface is illuminated by red light is a lie. In fact, the light shining on the surface is standard white light. Or consider a version of the cancellation case in which the notification that the show is cancelled is sent by mistake. In fact, the show is going ahead as planned. In both cases, the defeaters you acquire are misleading. In the tweaked illumination case, my testimony provides reason against believing that the surface is red by providing evidence for a falsehood, to wit, that the surface is illuminated by red light. Likewise, in the tweaked cancellation case, the notification you receive provides reason against believing that the show is happening tonight by providing evidence for a falsehood, to wit, that the show was cancelled.

Most contributors to the debate register an intuition that there is such a thing as misleading defeat. For instance, they maintain that, in the tweaked illumination and cancellation cases, intuitively, you lose justification for your beliefs in light of misleading defeaters. In fact, most contributors to the debate also register an intuition that you can even lose knowledge via misleading defeat. For instance, in the tweaked illumination and cancellation cases, it may well be that initially your beliefs qualified as knowledge. Even so, once the new information comes in, you lose your justification and, indeed, your knowledge as a result of misleading defeaters. While I will look at views that take at least knowledge to be indefeasible in Section 1, for now I'd like to flag that I side with the majority here in that I take it that justification and knowledge admit of defeat even via misleading defeaters.

The next distinction I'd like to highlight is between justification defeaters and (pure) knowledge defeaters. The most important idea here is that, alongside the class of justification defeaters we have already familiarised ourselves with, there is a class of defeaters that only target knowledge, leaving the justification of our beliefs intact. (Pure) knowledge defeaters have been of particular interest for champions of defeasibility accounts of knowledge, which aim to solve the Gettier (1963) problem by invoking defeat (e.g. Lehrer and Paxson 1969; Swain 1974). To get an idea of how this is meant to work, consider a standard Gettier case. You acquire a justified true belief that it's 8:22 by taking a reading from your watch. Crucially, unbeknownst to you, your wristwatch stopped exactly twelve hours ago ('the stopped clock case'). In this case, your justified true belief that it's 8:22 falls short of knowledge. According to defeasibility accounts of knowledge, the reason for this has to do with defeat. More specifically, the thought is that the fact that your clock is stopped defeats your knowledge that it's 8:22.

Of course, if this project can be made to work, my characterisation of defeat
requires reworking. To see why, note that I characterise defeat in terms of loss of
justification. If defeasibility theorists are right, then there is defeat of knowledge
without defeat of justification. And, of course, if that's correct, any characterisation
of defeat in terms of loss of justification cannot be adequate. Now, I do not mean to
foreclose the possibility of (pure) knowledge defeat. That said, for present pur-
poses, I'd like to set it aside and restrict my focus to justification defeat. By the
same token, any problem that the existence of knowledge defeat may cause for my
characterisation of defeat in terms of loss of justification can safely be ignored.

The last distinction I'd like to look at is perhaps the most controversial. This is the
distinction between internal and external defeat.[3] While, in cases of internal defeat,
the defeater is psychologically registered by the relevant believer, in cases of external
defeat it isn't.[4] One may wonder why I say that this distinction is controversial. After
all, if there is such a thing as (pure) knowledge defeat as defeasibility theorists would
have us think, then there must also be such a thing as external defeat. After all, the
kinds of (pure) knowledge defeaters that would explain the absence of knowledge in
Gettier cases are standardly not psychologically registered. For instance, in the
stopped clock case, the fact that your wristwatch is stopped, which constitutes the
(pure) knowledge defeater here, is not psychologically registered.

I agree that once we allow for (pure) knowledge defeaters, then we will also have
to countenance external defeat. What I mean when I say that the distinction between
internal and external defeat is controversial is that it is controversial that there is
such a thing as external defeat of justification (henceforth just 'external defeat').
That is to say, what is controversial is that we can lose justification for believing
something in light of information that we do not psychologically register. Note that
all cases of (justification) defeat we have encountered so far are cases of internal
defeat. In all cases you register the relevant information that provides reasons
against holding the relevant beliefs. For instance, in the cancellation case, you
learn about the cancellation, and, in the alibi case, you find out about the eyewitness
testimony and the video surveillance. What's more, it may be hard to see how
information that we don't register could lead to the defeat of justification. For

---

[3] Note that the phenomenon sometimes goes by the labels 'propositional defeat' (Bergmann 2006)
or 'normative defeat' (e.g. Lackey 2008). I am not happy with either term. The first suggests an
analogy with propositional justification. Propositional justification is something that internalists
can and indeed happily do acknowledge the existence of. The existence of what I call external
defeat is incompatible with internalism. This is why I prefer not to go with 'propositional defeat'
as a label. Normative defeat suggests that there is also non-normative defeat. Again, I take it that
this is false, which is why I am not satisfied with this label either. That said, nothing hinges on this
choice of terminology.

[4] What does it take for a defeater, d, to be psychologically registered? For present purposes, I follow
Lackey (e.g. 1999, 2003) in that one needs to have some doxastic attitude towards d (like beliefs
or doubt).

instance, how could your justification for believing that the show is happening tonight be defeated if you never received the notification that it was cancelled? And how could your justification for believing that the suspect is guilty be defeated if the information about the alibi and the video surveillance never got through to you?

What's more, things don't look much more promising when we look at the classical cases of external defeat in the literature. Here is one. Suppose I want to play a prank on you and make you believe that I am in San Francisco for the summer, while, really, I am in Rome. To this end, I mail a bunch of letters with descriptions of my stay in San Francisco to a friend in San Francisco who mails them from San Francisco to you. At present they are in a pile of unopened letters on your desk that has been accumulating for days. You happen to call my office and are told that I am spending the summer in Rome ('the unopened letter case', Harman 1980, 164).

The thought here is that the fact that the letters on your desk say that I am in San Francisco constitutes a defeater. While originally the case was meant as a case of defeat of knowledge, it is sometimes considered a case of defeat of justification also (Nottelmann 2021, 1185). I agree that it is not clear that this is a genuine case of defeat. Even the claim that it is a case in which knowledge is defeated strikes me as tenuous. The claim that it is a case of justification defeat is more controversial.

In light of this, it becomes more and more understandable why the distinction between internal and external defeat has remained controversial. Even so, one key ambition of this Element is to mount a case that there is such a thing as external defeat and that, as a result, any adequate epistemology of defeat must make room for external defeat. Accordingly, one of the central tasks for the account of defeat I am to develop is to show how this can be done.

Note that the case for external defeat has implications for the debate between the psychological view and the normative view. To see this, recall that cases of external defeat involve defeaters that aren't psychologically registered. This means that, in cases of external defeat, the kind of relation that the psychological view posits doesn't obtain. As a result, if there is such a thing as external defeat, the psychological view cannot be correct.[5] Unsurprisingly, I favour a version of the normative view. Accordingly, another central task for this Element, which I take on in Section 5, is to develop a version of the normative view that can allow

---

[5] That said, for present purposes, I don't need to take a stance on whether some psychological relation is sufficient for a defeater to undermine justification. What matters is that no psychological relation is required for this. After all, that's what is needed to make room in our epistemology for external defeat. Accordingly, for present purposes, the correct view may well be a hybrid one. That said, I worry that the psychological view overgenerates defeat. Sexists may have defeaters for the word of women simply in virtue of the fact that they believe that women are not trustworthy. In my view, this can't be right. Defeat cannot be that easy to come by. In light of this, I will take it that the normative view is correct here.

for external defeat without overgenerating defeat. For instance, in the employ-
ment case, we will want the view to predict that your belief that you will continue
to receive your pay continues to be justified (i.e. not defeated) even after some
stranger in a faraway place has asserted that your employer will be unable to pay
their employees next month.

Here is a game plan for this Element. I distinguish among three broad views
in the epistemology of defeat: scepticism, internalism, and externalism. I argue
that there is excellent reason to think that sceptical and internalist accounts of
defeat remain unsatisfactory. More specifically, I argue, *pace* sceptics, that we
do need to countenance defeat in epistemology (Section 1). *Pace* internalists,
I provide reason to believe that the correct epistemology of defeat must be
externalist. The perhaps most important reason for this, albeit not the only one,
is that the correct epistemology of defeat must countenance the existence of
external defeat (Section 2). As a result, any viable account of defeat must be
externalist. While externalists have the resources to accommodate external
defeat in principle, I argue that extant externalist accounts have failed to
successfully accommodate external defeat in practice. The reason for this is
that externalist accounts of defeat are either too restrictive or too inclusive
(Section 3). Finally, I begin to develop a better alternative. More specifically,
I introduce my preferred account of justification and develop an account of
defeat for it, which, I argue, can handle not only familiar cases of internal defeat
(Section 4) but also the particularly difficult cases of external defeat (Section 5).

2 Defeat Scepticism

2.1 Introduction

Sceptics about defeat think that there is no such thing as defeat. To be fair, sceptics
about defeat typically focus on knowledge and hold that knowledge does not
admit of defeat. In this way, defeat sceptics may not be all that radical after all.
Even so, if it turns out that a central epistemic property is indefeasible, that's an
interesting result in its own right. What's more, it does raise the question whether
defeat scepticism generalises beyond knowledge. After all, as we have seen in the
Introduction, intuitively knowledge does admit of defeat. If defeat sceptics are
right and this intuition is mistaken, we have to take seriously the possibility that
we make a similar mistake in cases in which, intuitively, justification is defeated.

The central aim of this section is to argue against defeat scepticism. More
specifically, I will provide reason to believe that the perhaps most prominent
general argument for defeat scepticism fails (Section 1.2).[6] More importantly

---

[6] Note that there are also more modest approaches which aim to defend scepticism about specific
kinds of defeat (e.g. defeat via higher-order evidence, see e.g. Coates 2012; Lasonen-Aarnio 2014;

yet, if defeat scepticism is to be viable at all, its advocates will need an alternative explanation of what's going on in intuitive cases of defeat; that is, they need an error theory. Section 1.3 takes a closer look at the two main error theories in the literature. I argue that one ultimately doesn't succeed in making do without defeat. While the other one is more promising, I argue that if it is successful, it can be used, and indeed is better used, as a non-sceptical account of defeat. Finally, Section 1.4 argues that there is reason to think that the more promising account doesn't work in any case, no matter whether it is used as an error theory by defeat sceptics or as a non-sceptical account of defeat.

2.2 The Case for Defeat Scepticism

Let's begin by taking a look at why one might think that knowledge does not admit of defeat. The most popular answer is that our best accounts of knowledge do not appear to allow for the possibility of defeat. Maria Lasonen-Aarnio (2010) argues this point forcefully for safety accounts of knowledge, which are among the leading accounts of knowledge on the market.[7]

The core idea of safety accounts of knowledge is that knowledge is belief that is safe from error, where safety from error is a matter of avoiding error across nearby possible worlds. It is widely agreed that this core idea needs refinement. In particular, the safety condition on knowledge needs to be relativised to ways in which the relevant belief is held. While there are several ways of making this idea more precise, I'll be working with the following rough view: One's belief that p is safe if and only if there are no nearby worlds at which one holds a false belief on whether p in the same way.[8]

To get a feeling for how the view works, let's consider a couple of cases. I know that there is a laptop in front of me. Here my belief that there is a laptop in front of me is safe. After all, it is held via my ability to recognise laptops and there are no nearby worlds at which this ability delivers a false belief on whether there is a laptop in front of me. Contrast this case with a standard Gettier case such as the stopped clock case. Recall that, in this case, you acquire a justified

---

Wedgwood 2012). Due to limitations of space, I will not look at these approaches here. Instead, I will focus entirely on approaches that aim to establish a more general form of defeat scepticism.

[7] For a similar line of argument see Baker-Hytch and Benton (2015). An alternative route to defeat scepticism is to go for an account of knowledge on which knowledge is very easy to get. One example here is the view that knowledge is just true belief (e.g. Sartwell 1992), or at least that minimal knowledge is (e.g. Hetherington 2001). It's hard to see how knowledge could be defeated on this kind of view (though of course it can be lost via loss of belief). I will not take the time to discuss this kind of view here. The main reason for this is that I think this kind of view fails for reasons unrelated to defeat (see e.g. Turri 2012).

[8] Prominent defences of safety accounts of knowledge include Pritchard (2005); Sosa (1999); Williamson (2000), for criticisms see, for example, Comesaña (2005); Kelp (2009); Neta and Rohrbaugh (2004).

true belief that it is 8:22 by taking a reading from your stopped wristwatch.
Here, too, your belief that it is 8:22 is held via your ability to read the clock.
Crucially, here your belief is not safe. There are nearby worlds at which you
look at your wristwatch a minute earlier or later. Since at those worlds, your
watch is still stopped, it does not display the right time and you end up with
a false belief that it is 8:22.

Now, in a nutshell, the trouble is that the question as to whether a belief is safe
does not turn on whether one has a defeater in the way we'd need it to. To see
this, consider the following case. My nephew is accused of a crime. I have
ample evidence that he is innocent, and I believe that he is innocent ('the
nephew case'). Now consider two ways of developing this case. In the first,
my belief that my nephew is innocent is held based on careful consideration of
the evidence I have. Here my belief is safe (or so we may assume). After all,
there are no nearby worlds at which I hold my belief on whether my nephew is
innocent via careful consideration of the evidence, and yet I end up with a false
belief. Safety accounts deliver the desired result that my belief qualifies as
knowledge. In the second, despite all the evidence that I have, my belief that my
nephew is innocent is held via a coin toss. Here my belief is unsafe. After all, at
many nearby worlds at which I form a belief on whether my nephew is innocent
based on a coin toss, I don't get lucky and end up with a false belief. Again,
safety accounts deliver the desired result.

For present purposes, the important take home lesson concerns the key
relativisation to ways of holding beliefs at issue in the safety condition. How
a given belief is held is independent from what evidence one has. Rather, it is
a contingent matter of psychological fact. In particular, even if one has excellent
evidence for believing a certain proposition, the question of whether one's
belief qualifies as knowledge remains open. If one holds one's belief via careful
consideration of the evidence (as in the first development of the nephew case), it
may count as knowledge. However, there just isn't any guarantee that one will
hold one's belief in this way. In fact, it may well be that one holds one's belief in
a different way, one that prevents it from qualifying as knowledge (as the second
development of the nephew case clearly indicates).[9]

With these points in play, let's return to the question as to why safety accounts
will struggle to accommodate standard cases of defeat. To see the answer, recall
one of our toy examples from the Introduction, the illumination case. More
specifically, consider a version of the case in which I know that the surface in
front of me is red. According to safety accounts, this is because I have a safe

---

[9] These points are familiar from the literature on propositional and doxastic justification. The
distinction dates back to Firth (1978). For more on propositional and doxastic justification see, for
example, Silva and Oliveira (2022).