

Understanding Undermining Defeat

This is the accepted version of an article whose final and definitive form is published in Philosophical Studies. Philosophical Studies is available online at: link.springer.com. The article is available at the following permanent link: <https://doi.org/10.1007/s11098-013-0238-z>. Please cite the published version only.

ABSTRACT

I articulate a view according to which an important difference between undermining and overriding defeaters is that the former require the subject to engage in some higher-order epistemic thinking, while the latter don't. With the help of some examples, I argue that underminers push the cognizer to reflect on the way she formed a belief by challenging the epistemic worthiness of either the source of justification or the specific justificatory process. By contrast, overrides needn't pose any such challenge. I also provide some details on how undermining defeat works in different cases.

1. INTRODUCTION

The notion of undermining (aka undercutting) defeat is acquiring increasing relevance in epistemology in virtue of its connection with many issues, such as the relationship between higher-order evidence and ordinary evidence, the debate on peer disagreement, and the debate between dogmatism and conservatism in the philosophy of perception.¹ This paper is concerned with the understanding of the nature of undermining defeat, regardless of its bearing on any other debate.

In a recent paper critical of Pollock's account of the way epistemic defeaters work, Scott Sturgeon (2012) argued that while overrides (aka rebutters) generate their distinctive kind of pressure on their own, underminers "generate it only in concert with higher-order commitments about the

¹ For two examples of how undermining defeat plays a role in those debates, see Christensen (2010) and Pryor (ms).

basing of lower-order belief'. Some years before, Albert Casullo (2003: 45-6) observed that "[t]ypical undermining defeaters show either that the source [of justification] is defective in some way [...], or that the source is operating in an environment for which it was not well adapted". I take the two passages to inspire the view according to which underminers pertain to what we might call 'the higher-level of belief-management', while overrides can be limited to what we might call 'the lower-level of belief reception and rejection'. This paper explores this view and sketches some of its details.

2. THE VIEW AND A PROBLEM

The distinction between undermining and overriding defeaters has been introduced by John Pollock (1974: 42-3). Let p be a previously justified proposition (for a subject S at a time t): while overrides overtly suggest that $\text{not-}p$, underminers don't. Instead, they are reasons to believe that the justification is not strong enough to justify belief in p .² Consider the following examples, where e is the evidence, p the proposition the subject is justified to believe, and d the defeater³ (and e and d are the only pieces of evidence relevant for p that are available to the subject):⁴

- (1) $e = \langle \text{Andreas says that puffins are a species of mosquitoes} \rangle$
 $p = \langle \text{Puffins fly} \rangle$
 $d = \langle \text{Filippo tells me that Andreas's knowledge in naturalistic matters is poor} \rangle$
- (2) $e = \langle \text{I remember having left the book on the desk} \rangle$

² As it will be evident by my constant reference in cases where the belief is *based on* some evidence, I'm working with a notion of doxastic justification.

³ I will appeal to justificational triads e, p, d throughout and, for ease of exposition, I will take e, p and d to be propositions. However, I don't wish to commit myself to the claim that all evidence is propositional, but only to the much weaker claim that for every piece of evidence available to a subject, there is a proposition that can be used to represent it.

⁴ The parenthetic remark is meant to rule out cases of epistemic overdetermination, where the subject has access to other pieces of information for or against p , besides the mentioned e and p .

$p = \langle \text{The book is on the desk} \rangle$

$d = \langle \text{I now see that the book is not on the desk} \rangle$

It can quickly be seen that while d in (2) entails that not- p ,⁵ d in (1) suggests that the source of justification is not trustworthy, but it is compatible with p . In (2) we have an overrider; in (1) we have an underminer.

Let's recall the two remarks that inspire the view I wish to explore. Sturgeon claimed that while overrides generate their distinctive kind of pressure on their own, underminers "generate it only in concert with higher-order commitments about the basing of lower-order belief". If we look at example (1) again, we can see that d appeals to the commitment that the belief that puffins fly was based on Andreas's testimony—in other words: d couldn't do its defeating job, unless that commitment (or some other very similar to it) were in place. No similar commitment is appealed to in example (2): d would have been an equally effective defeater if, say, the belief that the book is on the desk were warranted by testimony rather than memory.⁶

The second inspiring remark comes from Casullo: "[t]ypical undermining defeaters show either that the source [of justification] is defective in some way [...], or that the source is operating in an environment for which it was not well adapted". Again, look at example (1): the suggestion is precisely that the source is not reliable with respect to the subject matter. In example (2), on the other hand, no suggestion is made with respect to the defectiveness of the source, or its working in a problematic environment.

Putting the two suggestions together, we can thus formulate the following view:

(View #1) Underminers suggest that something was wrong with the source of justification, and they operate their defeat by appealing to the higher-order commitment that the belief in question was based on that source.

⁵ Assuming that 'see' is factive.

⁶ In general, underminers appear to be source-sensitive (they defeat only the justification provided by a specific source), while overrides appear to be source-neutral (they defeat regardless of the source of justification). This difference, noted by Casullo (2003:45-6), might very well be a symptom of the feature Sturgeon describes.

The question from which I shall begin my investigation is: what exactly are the suggestions made by the underminers with respect to the source?

In the already mentioned passage, Casullo provides a twofold reply: underminers typically show that the source is defective, or that it is working in an environment for which it is not well adapted.

However, on a common understanding of the notion of source, we can easily come across cases in which neither of the two options offered by Casullo seem to apply. If by ‘source of justification’ we mean something on a quite large scale, like the five senses, memory, introspection, and whatever mental ability might deliver a priori justification,⁷ the following example appears to pose a problem for (*View #1*):

- (3) $e = \langle [S\text{'s apparent proof of } p] \rangle$
 $p = \langle [\text{a seemingly logical theorem}] \rangle$ ⁸
 $d = \langle \text{A logician tells } S \text{ that there's a mistake in the proof} \rangle$

In this case it seems natural to individuate the relevant source of justification in the subject’s proving abilities, as it were. Suppose that *S* is generally good at proving theorems, that the proof wasn’t especially difficult, and that there were no disturbing elements involved, such as having had too much wine, or having executed the calculation in a noisy place. If so, the underminer does not suggest either that the source was defective (we are supposing that the agent is generally good at carrying out proofs, and he is surely allowed an occasional mistake), or that it was not operating in the proper environment (we are supposing that there were no disturbing elements). What does the underminer suggest with respect to the defectiveness of the source—or, more generally, with respect to the defectiveness of the subject’s basis of belief—in case (3)?

⁷ This is the notion of source endorsed by Casullo himself (2003: 35-38).

⁸ Square brackets are meant to suggest that *e* and *p* are propositions, rather than the mere stand-in for propositions that are inside the brackets.

3. THE DIFFERENCE IN THE EFFECT OF UNDERMINING AND OVERRIDING DEFEATERS

I believe that, in order to make sense of case (3), we need to focus the attention on something of a rather smaller scale than the sources of justification. A notion that might be helpful here is that of ‘justificatory process’. Let us agree that a justificatory process begins with the gathering of the evidence and ends with the formation of the belief.⁹ The reason why I think that this notion is helpful is that it’s perfectly fine for a non-defective source to, occasionally, bring about a defective justificatory process (unless one is bold enough to impose the rather demanding requirement that a source of justification has to be infallible).

In case (3) the justificatory process consists mainly in the actual flawed proof produced by the subject, and the natural thought concerning the suggestion made by *d* is that the process, rather than the source, was defective in some way. This is a suggestion concerning the higher-order commitment that the belief that *p* was based on that particular alleged proof, and it is an open challenge to the opinion that it was a good idea to base the belief that *p* on that specific alleged proof. Therefore, it is perfectly in line with the view under exploration. We might now try to reformulate the proposal:

(View #2) Underminers suggest that something was wrong with the source of justification or with the justificatory process, and they operate their defeat by appealing to the higher-order commitment that the belief in question was based on that source or that process.

However, the appeal to justificatory processes generates another worry. Whereas it’s plausible that overrides say nothing about the defectiveness of the source (think about case (2): *d* does not suggest that memory is a bad source), things are not so when it comes to justificatory processes. Indeed, it seems that overrides suggest that the justificatory process was defective in some way too. Recall that

⁹ The choice to focus on evidence-supported justified beliefs is not due to a disregard for other options, but only to a desire to keep the exposition as simple as possible.

overrides say or entail that not- p , where p is the proposition previously justified for subject S . Now, since surely one of the epistemic goals is to believe what is true, if S formed the belief that p , and p is false, something must have gone wrong in the epistemic path that led S to believe p . If so, the way in which underminers and overrides work doesn't look that different after all.

Here is a reply: the way in which underminers and overrides suggest that the justificatory process was defective is different. Overrides merely suggest that the justificatory process failed to lead to truth, underminers suggest that a disturbing event (like a mistake in the proof) has caused the process to fail. The former suggestion is perfectly compatible with the justificatory process having been executed impeccably, and having been delivered by a reliable source working in good circumstances; the latter is not. I wish to defend this line of reply by considering the examples.¹⁰

Example (1) is not problematic for the view, since d clearly says that the source (Andreas's testimony) wasn't reliable. The problematic case was (3), and we should now compare it with the overriding in (2) to see if we still have a difference. As we have seen, d in (3) it suggests that there was a mistake in the justificatory process (the alleged proof). By contrast, in (2), the overrider (obtained by looking at the desk, and seeing that there was no book) does not suggest anything with respect to either the defectiveness of the source, its working in adverse circumstances, or the occurrence of some mistake, or of any other disturbing event, in the justificatory process.¹¹ All that (2) seems to suggest with respect to the justificatory process is that it failed to lead to truth, and it's in the nature of justificatory processes originated by fallible sources of justification to occasionally fail to produce a true belief. It might be that, under close examination, the failure to produce a true belief in (2) turns out to be due to some glitch in the justificatory process (maybe I didn't inspect my memories

¹⁰ There might be another reply available here: in a nutshell, one could agree that both overrides and underminers suggest that something went wrong in the justificatory process, but still contend that while the consideration of that thought plays a role in the process of belief revision generated by an underminer, it needn't play any role in the belief-revision process generated by an overrider.

¹¹ It's worth saying something on the notion of disturbing event at this stage. By 'disturbing event' I mean something that interfered with the justificatory process and caused its failure. I take disturbing events to belong to two main categories: those for which the agent can be blamed (e.g. a mistake in a proof due to carelessness), and those for which the agent cannot be blamed (e.g. the acquisition of a false information from a reliable source).

carefully, or I was given some memory-distorting drug), but it needn't be so. And, more importantly for our purposes, the overrider does not suggest that things are so.¹²

Another example should help to make the point more vivid:

- (4) $e = \langle \text{All swans observed at } t_1 \text{ are white} \rangle$
 $p = \langle \text{All swans are white} \rangle$
 $d = \langle \text{At } t_2 \text{ a black swan is observed in Australia} \rangle$

Again, the overrider here does not challenge either the reliability of the source (inductive observation), its working in the proper environment, or the good standing of the specific inductive process. All that the overrider suggests is that that particular induction failed to lead to truth, and that is perfectly compatible with the induction being executed impeccably (say, a very high number of observations were made, they were all singularly carried out with care, etc.). Just like it seems to be in the normal course of events that sometimes memory delivers a false piece of information, so it's in the normal course of events that inductions are sometimes proven wrong.

Let me reformulate the view one more time:

(View #3) Underminers suggest that something was wrong with the source of justification or with the justificatory process, and they operate their defeat by appealing to the higher-order commitment that the belief in question was based on that source or that process. If the suggestion is that the process, rather than the source, was defective, the defectiveness is to be understood as the occurrence of a mistake or some other disturbing event.

¹² One could argue that faculties like memory and perception, if they are working at the best of their possibilities in optimal circumstances cannot deliver a false belief. On this view, the acknowledgment that memory and perception are not infallible is, at bottom, an acknowledgement that they hardly ever work at their best in optimal circumstances. If this view is correct, there is a sense in which overriders entail that something went wrong in the process in the sense I attach to underminers. However, the point stands that overriders do not suggest that things are so, and don't need things to be so to defeat the subject's original belief.

4. MORE DETAILS

I will now try to be a little bit more specific about what happens when a cognizer accepts an underminer. In line with the examples already presented, I use propositional triads to represent in a schematic way the pieces of information available to the subject when she encounters a defeater:

e = <evidence>

p = <supported proposition>

d = <defeater>

My contention is that when d is an underminer, it suggests something with respect to e or to the relationship between e and p . By contrast, when d is an overrider, there needn't be any such suggestion. More precisely, when d is an underminer, the suggestion might be that:

- (a) e —the evidence for p —is not true¹³
- (b) e does not support p
- (c) there is no justification for e
- (d) there is no justification for the claim that e supports p

Let me go back to the examples already considered to show how this connects with what I have already said. In (1) d does not impinge on the truth of e ; instead, it appears to suggest that e does not support p , on the grounds that the source which delivered p is defective (unreliable with respect to the subject matter). On the other hand, in (3) d does not challenge the claim that e , if true, would support p —the execution of logical calculations by a competent subject is a good way of knowing about logical theorems—but it does challenge the truth of e : at least if e is composed by the conjunction of

¹³ The specification that e is evidence for p is meant to make it clear that the suggestion made by the underminer is a higher-order one.

the statements which make up the alleged proof, then the logician's claim that there is a mistake in the proof amounts to saying that at least one of the statements involved is false, and thus the conjunction which includes them all is false as well. By contrast, consider the cases of overriding defeat in (2) and (4): neither the truth of e , nor the support relation holding between e and p are challenged by the overrider d .

I will now present two more examples, just to illustrate how features (c) and (d) might be instantiated. Here's the first, instantiating (c). Let's assume that when there is a railway workers' strike, the company affected usually advises its customers to check in advance whether the train they are booked on will travel or not.

- (5) $e = \langle \text{A train to Edinburgh will depart tomorrow at noon} \rangle$ [as learned by checking the station's timetable]
- $p = \langle \text{The train will get to Edinburgh tomorrow afternoon} \rangle$
- $d = \langle \text{The newspaper says that a railway workers' strike has been called for tomorrow} \rangle$

Does d here suggest that e is false? It wouldn't seem so. If the advice of the company is to be taken at face value, upon learning that a strike is announced, the subject is not able to assign a truth-value to a proposition like $\langle \text{Train } T \text{ will travel} \rangle$.¹⁴

Does d in (5) suggest that e is not relevant for p ? Not really. Learning that a railway workers' strike has been called does not appear to question that checking the timetable provides evidence for forming beliefs about train travel times—the evidence in the case at hand is not as strong as it would be on a strike-free day, but it is nevertheless a decent piece of evidence: it is enough to make it worthwhile continuing the investigation and getting in touch with the railway company.

¹⁴ One might think that the suggestion here is that e is false, but that the suggestion isn't strong enough to warrant belief in the falsity of e . I don't disagree with this way of putting it; it's just that I understand the suggestion made by the underminer to be the higher-order thought—about the basing of the first-order belief—that the thinker is warranted to believe. And what the thinker is warranted to believe in this case is that there is no justification to believe that e is true, not that that e —the evidence for p —is false.

What d does in (5) is to provide the cognizer with a good reason to doubt that e is the case, thereby pushing him to suspend his judgment over e . After having acquired d , he is no longer in the position to commit himself to the truth of e , but neither he is in the position to commit himself to the falsity of e . The suggestion then would be that there is no justification for e , and that would account for the defectiveness of the justificatory process (the source does not seem to be problematic here)¹⁵: its early stages, namely the evidence-gathering stages, are hindered by circumstances such that it's not clear whether the proposition representing the evidence can be accepted.

One more example, to illustrate feature (d):

(6) $e = \langle \text{My experience is in all respects as if I saw Stephan entering the building} \rangle$

$p = \langle \text{Stephan is in the building} \rangle$

$d = \langle \text{Bob has just told me that Paula was determined to put some hallucinatory pills in my tea} \rangle$

What does d suggest here? Not that e isn't true: the content of the relevant experience is not called into question by d . For the same reason the suggestion made by d can't be that there is no justification for e . What about the claim that e supports p ? Just like in the previous case, I think that, upon learning that d , the subject is still in the position to accept the claim that e supports p : the degree of support that e provides for p is not as strong as it would be in a situation devoid of d , but it would seem that e still supports p to some degree. To make the case more plausible, let's suppose that, thanks to some background information, I figure that Paula had about 50% chances of succeeding in adding some hallucinatory pill to my drink without me noticing. If so, it appears reasonable to withhold the belief in the claim that e supports p : what d seems to suggest is that there is no justification for the claim that e supports p . A way, maybe familiar by now, to summarize the story just told is: the underminer here suggests that the source of justification might not have been working in good enough circumstances.¹⁶

¹⁵ It might be a little tricky to specify the source here. One option might be: vision plus the required theoretical luggage needed to read timetables.

¹⁶ This example pushes for a further refinement on (*View #3*): the suggestion made by underminers with respect to the source or the process can be either that they were defective, or that they *might have been* defective.

5. MIXED CASES

Let's backtrack a bit: I have endorsed—and tried to provide some details of—a view according to which underminers do their work by interacting with some higher-order commitment concerning the basing of the belief, and that they do so by challenging the epistemic worthiness of the relevant commitments. On the other hand, overrides don't need to interact with higher-order commitments concerning the basing of the target belief, and don't have to challenge them. The examples discussed lend support to the view I have defended.

However, nothing seems to forbid that, in principle, there are mixed cases: overrides that do interact with some higher-order commitment about the basing of the target belief, and that do challenge the way in which the belief was formed. Here's an example:

- (7) $e = \langle \text{The sun moves around the earth} \rangle$
 $p = \langle \text{The sun orbits the earth} \rangle$
 $d = \langle \text{The best astronomical theory that we have today says that the earth orbits the sun} \rangle$

Suppose the cognizer is a relatively grown-up school kid, and that she gets e by observation, while she learns d in school. There seems to be a sense in which she is mandated to reason along lines similar to the following:

Astronomy says that the earth orbits the sun, yet I believe I saw the sun moving around in the sky and on that basis I formed the belief that the sun orbited the earth. Thus, science has contradicted what I've seen. Is there anything wrong in my vision? It doesn't seem so, I can rely very well on my vision when I play football or when I cross the streets. Eureka! There are some things for which vision is good for and others for which it is not good for: forming beliefs about the movements of the stars is one of the things that vision is not good for; forming beliefs about the position of the ball in the football pitch is a thing that vision is good for.

We might suppose that the reasoning just presented relied on a higher-order commitment like “school teachers are always right (in school related matters)”, which makes it the case that *d* could call into question the epistemic quality of the other higher-order commitment “my belief that the sun orbits the earth is based on vision”.¹⁷ In this reading of case (7), the acceptance of the overrider pushed the young cognizer to reflect on the way in which the belief was formed, the final suggestion being that *e* was false (because the source was not dealing with a subject matter for which it was well adapted). I call this a case of ‘impure overriding’ because the overrider behaved in an underminer-like way.¹⁸

Note, however, that the effectiveness of the overrider, or even its acceptance, is not dependent on the higher-order reasoning of the sort just presented. The school kid could have accepted the overrider, given up her belief in *p*, and at the same time remained unreflective about the way in which the belief in *p* was formed. Such behaviour is probably not ideal, but it doesn’t seem that epistemic rationality forbids it. In the case of undermining defeaters, however, things appear to be different: if the view I’ve tried to sketch is on the right track, a cognizer cannot accept an underminer, and at the same time remain unreflective about the way in which her belief was formed.¹⁹

To conclude, let me briefly draw the general moral of this story. Calling into question the way the belief was formed is a crucial part of the way in which underminers work, and accepting that one’s justification has been undermined requires some reflection on how the belief was formed. It is in this sense that underminers pertain to the level of belief-management: they require the subject to worry about the way her beliefs are formed, retained, and abandoned.

¹⁷ Or, maybe more precisely, “vision plus the theoretical luggage needed to handle notions like orbit, astronomical theory, sun, earth”.

¹⁸ Jim Pryor (ms: § 2), by using a framework different than mine, comes to the conclusion that there are cases of mixed overriding/undermining defeat.

¹⁹ In other words, it is a consequence of the view presented that unreflective agents cannot suffer undermining defeat.

ACKNOWLEDGEMENTS

I wish to thank everyone at the Northern Institute of Philosophy for creating such a supportive environment, and especially Crispin Wright, Carrie Jenkins, and Federico Luzzi for discussions and feedback that have been crucial in shaping the ideas here presented. A predecessor of the present article was presented at the SIFA Graduate Conference, at the University of Cagliari in September 2013: I am grateful to the organizers and the participants for constructive and helpful comments.

REFERENCES

Casullo, A. (2003). *A Priori Justification*. Oxford: Oxford University Press.

Christensen, D. (2010). Higher-Order Evidence. *Philosophy and Phenomenological Research*, 81 (1), 185-215.

Pollock, J. L. (1974). *Knowledge and Justification*. Princeton: Princeton University Press.

Pryor, J. (ms). Problems for Credulism. To appear in Tucker, C. *Seemings and Justification: New Essays on Dogmatism and Phenomenal Conservatism*. Oxford: Oxford University Press. <http://www.jimpryor.net/research/papers/Credulism.pdf>. Accessed 15 August 2013.

Sturgeon, S. (2012). 'Pollock on Defeasible Reasons'. *Philosophical Studies*, published online: 17 April 2012. <http://link.springer.com/content/pdf/10.1007%2Fs11098-012-9891-x.pdf>