Undercutting Defeat & Edgington's Burglar[1]

Scott Sturgeon
University of Birmingham

Game Plan

This paper does four things. First it lays out an orthodox position on reasons and defeaters. Then it argues that the position just laid out is mistaken about "undercutting" defeaters. Then the paper explains an unpublished thought experiment by Dorothy Edgington. And then it uses that thought experiment to motivate a new approach to undercutting defeaters.

Reasons and Defeaters

Defeasible reasons are normally thought of as mental states of some kind. In the verbal tradition, at least, reputable philosophers sometimes react to this fact as if the whole idea of a defeasible reason is based on some kind of conceptual confusion or category mistake. Their idea, basically, is that the English word "reason" already has a meaning which rules out mental states as part of its extension. For this reason they see the idea of mental states as reasons as itself utter confusion.

My view is that the meaning of the English word "reason" is irrelevant to debate about defeasible reasons; for the claim that defeasible reasons are mental states--to be made here and found in the literature--should be thought of as a matter of _legislation_: if you like, the phrase "defeasible reason" should be understood as a technical one, something which picks out by fiat, if it picks out anything, whatever plays the role pinned down by it in theory; and the same is true, _mutatis mutandis_, for the word "reason" in what follows.

On the approach discussed here, then, defeasible reasons are mental states. But they are not just any kind of mental state; for the signature function of a defeasible reason is special, something unexecuted by most mental states. After all, the signature function of a defeasible reason is to generate _epistemic pressure_ or _rational bias_. We begin with that working assumption and generalise it when necessary.

More specifically, we use the expression "$x\mathbf{R}B(\Phi)$" to mean that x is a reason to believe Φ, with the basic idea of a reason then being

(R) $x\mathbf{R}B(\Phi)$ =df.   It is possible to become justified in believing Φ on the basis of x.[2]

[2] See, for instance, Pollock (1987a) p.35. Pollock emphasized defeasible reasons from the very beginning of his career. See his (1967), (1970) and his (1974), as well as the classics (1987a&b). I assume here that reasons under discussion are all of equal strength. Nothing turns on the assumption.

Here x is to be a mental state, the kind of thing on the basis of which one can rationally come to believe. We need not delineate, for present purposes, the exact kinds of mental states which play the x-role in (R). We need only assume that beliefs and perceptual states do so. They will be our focus in what follows.

       Next we follow orthodoxy and recognize two kinds of reason: indefeasible and defeasible. The former generate epistemic pressure to believe which cannot be wiped out or undone by the addition of information consistent with the information to hand. The latter generate epistemic pressure to believe which can be wiped out or undone by the addition of such information. Intuitively: indefeasible reasons are the meat and potatoes of rational deduction; and defeasible reasons are the meat and potatoes of rational induction, with reasoning of an inductive sort being the kind of reasoning which exploits the workings of <u>defeaters</u>. In turn these are mental states which wipe out or undo the rational bias put in place by reasons.

       We use the expression "$y\mathbf{D}[x\mathbf{R}\mathrm{B}(\Phi)]$" to mean that y is a defeater for x as a reason to believe Φ, with the basic idea of a defeater then being

(D) $y\mathbf{D}[x\mathbf{R}\mathrm{B}(\Phi)]$ $=_{\mathrm{df.}}$ (i) $x\mathbf{R}\mathrm{B}(\Phi)$,

                                          & (ii) $\neg(x{+}y)\mathbf{R}\mathrm{B}(\Phi)$.

On this way of thinking, y is a defeater for x as a reason to believe Φ exactly when it is possible to become justified in believing Φ on the basis of x but not possible to become justified in believing Φ on the basis of x and y together.

       Next we follow orthodoxy and recognize two kinds of defeater: rebutting and undercutting. Before explaining them, though, we flag an important assumption normally made about defeaters, an assumption which will come under attack later:

The Defeaters Assumption.   Defeaters do their work because they are reasons to believe. They generate their distinctive kind of epistemic pressure—defeating epistemic pressure—in virtue of being reasons to believe.

This assumption plays a crucial role in standard thinking about undercutting and rebutting defeaters.

       To see this, consider the latter. The basic idea of a rebutting defeater for x as a reason to believe Φ is

(RD) $y\mathbf{RD}[x\mathbf{R}\mathrm{B}(\Phi)]$ $=_{\mathrm{df.}}$ (i) $x\mathbf{R}\mathrm{B}(\Phi)$,

                                       & (ii) $y\mathbf{R}\mathrm{B}(\neg\Phi)$.

Or as it is normally put: y is a rebutting defeater for x as a reason to believe Φ exactly when x is a reason to believe Φ and y is a reason to believe $\neg\Phi$. Moreover: these conditions are meant to hold exactly when it is possible to become justified in believing Φ on the basis of x and possible to become justified in believing $\neg\Phi$ on the basis of y. But since rebutting defeaters are meant to be defeaters, of course, standard thinking about defeat—found at (D)—joins with standard thinking about rebutting

defeat—found at (RD)—to entail that it is not possible to become justified in believing Φ on the basis of a reason to believe Φ plus a reason to believe ¬Φ.

It is normal to deploy a similar line for undercutting defeaters. Unfortunately the details of this become delicate straightaway. We shall work up to them with a pair of vignettes: one involving the undercutting defeat of belief, the other involving the undercutting defeat of experience. The orthodox approach to undercutting defeat explains it in slightly different ways depending on whether the reason being undercut is itself a belief or an experience. The need for such a wrinkle in theory will not arise on the new view of undercutting defeat sketched later in this paper.

Here is our first vignette:

The Polling Case

A pollster surveys 1000 voters in Texas at random, asking whether they will vote Republican or Democrat in the next election. Results generate belief in a claim about testimony:

> T = 87% of respondents said they will vote Republican.

Belief in T then generates belief in a generalisation:

> G = Roughly 87% of Texans will vote Republican.

Suppose the pollster then comes to believe

> U = Respondents decided their answer by coin flip.

Intuitively, belief in U is a defeater for belief in T as a reason to believe G. Yet belief in U is not a reason to believe ¬G. U is neutral concerning the truth-value of G. So belief in U is not a rebutting defeater for belief in T as a reason to believe G. Rather, belief in U is an undercutting defeater for that reason. As the discoverer of undercutting defeat, John Pollock, routinely put it in his work: belief in U "attacks the connection" between belief in T and belief in G.[3] But how does belief in U play this role? How does it attack the epistemic link which manifestly exists between belief in T and belief in G?

Pollock answers this question by fleshing out a schema meant to slot into a schema for the undercutting defeat of belief:

$$B(U)\mathbf{UD}[B(T)\mathbf{R}B(G)] \quad =_{df.} \quad (i) \quad B(T)\mathbf{R}B(G),$$
$$\& \ (ii) \quad B(U)\mathbf{R}B(\neg X)].$$

The shape of this schema is familiar: belief in U is an undercutting defeater for belief in T as a reason to believe G exactly when belief in T is a reason to believe G and belief in U is a reason to believe some claim ¬X. This is the same theoretical shape used earlier in our approach to rebutting defeaters. And just as before the conditions put forward are meant to hold exactly when it is possible to become justified in believing G on the basis of belief in T and possible to become justified in believing

---

[3] *ibid.*

¬X on the basis of belief in U. Since undercutting defeaters are meant to be defeaters, moreover, our general approach to defeat—found at (D) and endorsed by Pollock— joins with his schema for the undercutting of belief—found just above—to ensure that it is not possible to become justified in believing G on the basis of a reason to believe G plus a reason to believe ¬X.

This generates a simple question: what plays the X-role in the schema for the undercutting defeat of belief? What is the claim X such that belief in U's being a reason to believe ¬X thereby makes belief in U attack the connection, in the way characteristic of undercutting belief, between belief in T and belief in G?

Throughout his work on undercutting defeat Pollock used surprisingly delicate forms of words to answer this question. And each time he took an initial stab at formulating such a form of words, he re-phrased it immediately. And often those re-phrasings did not look straightforwardly equivalent to what had been rephrased. In my view, this is diagnostic of a weak spot in Pollock's approach to undercutting defeat, with the weakness being exposed in a moment. . For now, though, we shall flesh out the shape of Pollock's approach to undercutting defeat, so we can work effectively to its weaknesses.

In the second edition of _Contemporary Theories of Knowledge_, Pollock and Joe Cruz say that the undercutting defeater U in the Polling Case "is a reason for doubting or denying that you would not have the inductive evidence unless G were true."[4] This loose form of words suggests a plausible thought in the Polling Case—and, as we'll see, a plausible thought in the analogue of the Polling case to do with visual reasons. In turn the plausible thought is this: any information to the effect that political facts on the ground were not responsible, in the Polling case, for the polling data in that case would _thereby_ undercut belief in the polling data as a reason for thinking that G is true. In other words, any reason to deny T-because-G would itself undercut belief in T as a reason to believe G.

This plausible thought will make for an easy-to-understand and initially-plausible theory of undercutting defeat. Unfortunately that theory is subject to clear counter-examples, as we'll see. And while the easy-to-understand view of undercutting defeat is not exactly faithful to Pollock's harder-to-understand approach, once we get clear on the counter-examples to the easy-to-understand theory it will be clear how to construct analogues for the harder-to-understand approach.

To begin, we create an easy-to-understand approach to the undercutting defeat of belief by appeal to the plausible idea sketched two paragraphs back. Let "¬[Φ ⊗ Ψ]" mean "not-(Φ is true because Ψ is true". Then the rebutting-defeat-style schema for the undercutting of belief—the schema of the previous page—can be filled in as follows:

(UDB-easy) $B(\Delta)\mathbf{UD}[B(\Phi)\mathbf{R}B(\Psi)]$ $=_{df.}$ (i) $B(\Phi)\mathbf{R}B(\Psi)$,

& (ii) $B(\Delta)\mathbf{R}B(\neg[\Phi \otimes \Psi])$.

In quasi-English: belief in Δ is an undercutting defeater for belief in Φ as a reason to believe Ψ exactly when belief in Φ is a reason to believe Ψ, and belief in Δ is a reason to believe not-(Φ is true because Ψ is true). These conditions are meant to hold

---

[4] On p.196. I have renamed the claims in the example. Pollock used the same form of words in the singly-authored first edition of the book.

exactly when it is possible to become justified in believing Ψ on the basis of belief in Φ, and possible to become justified—on the basis of belief in Δ—in believing not-(Φ is true because Ψ is true). And since undercutting defeaters are meant to be defeaters, of course, our approach to defeat—found at (D)—joins with the approach to undercutting defeat sketched above—to entail that it is not possible to become justified in believing Ψ on the basis of a reason to believe Ψ together with reason to believe not-(Φ is true because Ψ is true).

This easy-to-understand approach to the undercutting defeat of belief extends gracefully to the undercutting defeat of visual experience, our exemplar type of experiential reason. Consider another vignette:

The Visual Case
You have a visual experience representing a red object before you. On its basis you come to believe R: the claim that there is a red object before you. Then you are informed that

   U =   Local lighting is tricky in that it makes non-red objects look red.

Intuitively, belief in U is a defeater for your visual experience as a reason to believe R. Yet belief in U is not a reason to believe ¬R. U's truth is irrelevant to whether there is a red object before you. Belief in U is not a rebutting defeater for your visual experience as a reason to believe R. It is an undercutting defeater instead: somehow belief in U attacks the connection between your visual experience and your belief in R; and it does so in the way characteristic of undercutting defeat.

But how?

As we have seen, orthodoxy has it that the undercutting of belief turns on the content of one mental state—the belief which is the undercut reason—failing to be related aptly to the content of another belief—the belief which is formed on the basis of the undercut reason. This idea does not generalise to the undercutting of visual experience. The Visual Case (and vignettes like it) involve visual contents identical to belief-contents formed canonically on their basis. Yet no content is a reason for itself, much less an undercut reason; so unlike the undercutting of belief, the undercutting of visual experience cannot be entirely a matter of the content of mental states involved being aptly related to one another.

So how does it work?

The most natural answer—which dovetails with the easy-to-understand approach to undercutting above—turns on the fact that the existence of the defeasible reason itself, in the Visual Case, is explained by the very situation which makes true the belief formed on its basis. The relevant thought is that belief in Δ undercuts your visual experience of Φ as a reason to believe Φ when belief in Δ is reason to doubt or deny that you experience as of Φ because Φ. We can gracefully extend the easy-to-understand approach the undercutting defeat of belief, then, so that it to covers visual reasons as well, by letting "¬[V(Φ) ⊕ Φ]" mean "not-(you visually experience as of Φ because Φ". Then we have an easy-to-understand approach to the undercutting defeat of visual reasons

(UDV-easy) $B(\Delta)\mathbf{UD}[V(\Phi)\mathbf{R}B(\Phi)]$ $=_{\text{df.}}$ (i) $V(\Phi)\mathbf{R}B(\Phi)$,

$$\text{\& (ii)} \quad B(\Delta)\mathbf{R}B(\neg[V(\Phi) \oplus \Phi]).$$

In quasi-English: belief in Δ is an undercutting defeater for visual experience of Φ as a reason to believe Φ exactly when visual experience of Φ is a reason to believe Φ, and belief in Δ is a reason to believe that it is not the case that you visually experience as of Φ because Φ. These conditions are meant to hold exactly when it is possible to become justified in believing Φ on the basis of visual experience of Φ, and possible to become justified—on the basis of belief in Δ—in believing that it's not the case that you experience as of Φ because Φ. And since undercutting defeaters are meant to be defeaters, of course, the approach to defeat found at (D) joins with the approach to the undercutting defeat of visual experience found above to entail that it is not possible to become justified in believing Φ on the basis of a reason to believe Φ together with reason to believe it is not the case that that you experience as of Φ because Φ.

This completes a picture of the traditional approach to reasons and defeaters. It is not strictly Pollock's approach to undercutting defeat, as we'll see. But it is decidedly in the spirit of Pollock's approach; and it has the considerable merit of being easy to understand. According to every approach in the neighbourhood, including the easy-to-understand one we have sketched, reasons are mental states which generate epistemic pressure to believe; and defeaters succeed in destroying that pressure by generating their own epistemic pressure to believe.

This is very puzzling indeed. How can defeat itself spring from epistemic pressure to believe? In the next section, we flesh out this worry and unearth an assumption something like which must lie behind any approach to reasons and defeaters like Pollock's.

<u>Combining Things</u>
When we place our definitions of reason and defeat along side the Defeaters Assumption mentioned earlier—the view that defeaters do their work by being reasons to believe—puzzling explanatory schemata result. Suppose y is a rebutting defeater for x as a reason to believe Φ. The position before us entails the following **e**xplanation of **r**ebutting **d**efeat:

(ERD)   It is not possible to become justified in believing Φ on the basis of $(x+y)$, despite it being possible to become justified in believing Φ on the basis of x alone, ***because*** it is possible to become justified in believing $\neg\Phi$ on the basis of y. In symbols:

$$\neg\blacklozenge B(\Phi)_{[x+y]} \text{ even though } \blacklozenge B(\Phi)_x \ \textbf{\textit{\underline{because}}} \ \blacklozenge B(\neg\Phi)_y.$$

This cannot be bedrock theory. There are stories of its form which do not work at all—e.g. when something intuitively irrelevant is substituted for $\neg\Phi$—and for all we've been told y is a reason for $\neg\Phi$ only when sitting by itself in an agent's psychology, as the only relevant concern, whereas x is a reason for Φ when accompanied by further salient consideration. That sort of asymmetry has not been ruled out.

Or suppose y is an undercutting defeater for a belief x as a reason to believe Ψ. Then x will have a content Φ, and the overall position before us will entail the following **e**xplanation of the **u**ndercutting **d**efeat of **b**elief:

(EUDB)   It is not possible to become justified in believing Ψ on the basis of
         (*x+y*), despite it being possible to become justified in believing Ψ
         on the basis of *x* alone, ***because*** it is possible to become justified in
         believing ¬[Φ ⊗ Ψ] on the basis of *y* [i.e. because it is possible to
         become justified in believing, on the basis of *y*, that not-(Φ is true
         because Ψ is true)].   In symbols:

$$\neg\blacklozenge B(\Psi)_{[x+y]} \text{ even though } \blacklozenge B(\Psi)_x \textbf{\textit{ because }} \blacklozenge B(\neg[\Phi \otimes \Psi])_y.$$

Once again—and for exactly the same kind of reason—the story cannot be bedrock.
There are stories of (EUDB)'s form which do not work at all—e.g. when something
intuitively irrelevant is substituted for ¬[Φ ⊗ Ψ]—and for all we've been told *y* is a
reason for ¬[Φ ⊗ Ψ] only when sitting by itself in an agent's psychology, as the only
relevant concern, whereas *x* is a reason for Φ when accompanied by further salient
consideration. That sort of asymmetry has not been ruled out.
        Or finally: suppose *y* is an undercutting defeater for visual experience *v* as a
reason to believe Φ. Then *v* will have a content Φ—or so we are supposing—and the
overall position before us will entail the following **e**xplanation of the **u**ndercutting
defeat of **v**isual experience:

(EUDV)   It is not possible to become justified in believing Φ on the basis of
         (*v+y*), despite it being possible to become justified in believing Φ
         on the basis of *v* alone, ***because*** it is possible to become justified in
         believing ¬[*v* ⊕ Ψ] on the basis of *y* [i.e. because it is possible to
         become justified in believing, on the basis of *y*, that not-(you
         experience as of Φ because Φ).   In symbols:

$$\neg\blacklozenge B(\Phi)_{[v+y]} \text{ even though } \blacklozenge B(\Phi)_v \textbf{\textit{ because }} \blacklozenge B(\neg[v \oplus \Phi])_y.$$

Once again the story does not look to be bedrock. Stories of its form do not work;
and we have not been told how *v* and *y* work epistemically within an agent's
psychology.
        Having said all that, *something* seems right about the stories about defeat just
canvassed. The explanation of rebutting defeat—found at (ERD)—obviously has
something going for it; and so do the explanations of the undercutting defeat of
belief—found at (EUDB)—and the explanation of the undercutting defeat of visual
experience—found at (EUDV). Why is that? More specifically, why do these stories
seem to have something going for them despite the fact that they are obviously not
bedrock theory?
        Consider two answers.
        Answer 1. When dealing a composite state (*x+y*), we presuppose that one can
rationally believe a claim Φ on its basis, which can be rationally believed on the basis
of *x* or *y* alone, ***only if*** it is rationally possible to believe all such Φ conjointly. When
x is a reason to believe Φ and *y* is a reason to believe Ψ, therefore, the presupposition
of Answer 1 is that composite state (*x+y*) will be a reason to believe Φ, or reason to
believe Ψ, only if it is rationally possible conjointly to believe Φ and believe Ψ.
        This story renders cogent the explanation of rebutting defeat found at (ERD).
After all, it is not rationally possible conjointly to believe Φ and ¬Φ. When *x* is a

reason to believe Φ and y is a reason to believe ¬Φ, therefore, the composite state (x+y) will be composed of elements which are each reasons for claims it is not rationally possible conjointly to believe. Answer 1 thus makes sense of the explanation of rebutting defeat found at (ERD).

   Unfortunately, the line does not handle the stories about undercutting defeat canvassed earlier [at (EUDB) and (EUDV)]. The key to Answer 1, after all, is this idea:

(⊷A1) defeaters do their work because they are reasons to believe
    something which cannot be rationally believed while also believing
    the claim for which the defeated reason is a reason.

Suppose y is a rebutting defeater for x as a reason to believe Φ. Then y is a reason to believe ¬Φ. Since it is not rationally possible to believe Φ and ¬Φ, y is a reason to believe something which cannot be rationally believed while also believing the claim for which x is a reason. Hence y is a defeater for x (and *vice versa*).

   Nothing like this is occurs when a reason is undercut. To see this, recall the details of undercutting defeat. In the Polling Case, for instance, T is the claim that 87% of Texas respondents to a poll said they will vote Republican, G is the claim that roughly 87% of Texans will vote Republican, and U is the claim that respondents decided their answer by coin flip. When x is a belief in T and y is a belief in U, x is reason to believe G and y is an undercutting defeater for x as a reason to believe G. According to the story canvassed earlier, y undercuts x as a reason to believe G because y is itself reason to believe a complex negation: ¬(T ⊗ G). This is the claim that it is not the case that testimonial evidence is due to the fact that roughly 87% of Texans will vote Republican. The key thought behind Answer 1 will attempt to underwrite this story by saying that it is rationally impossible conjointly to believe G and ¬(T ⊗ G), i.e. it is not rationally possible conjointly to believe that roughly 87% of Texans will vote Republican and it is not the case that the testimonial evidence is due to the fact that roughly 87% of Texans will vote Republican.

   That is simply not true. In the original Polling Case, after all, such a combination of beliefs is precisely what you should have if you come across further sound polling which indicates that G is true after all, that you were in a polling analogue of a Gettier case. This means that Answer 1 does not render cogent the overall story about the undercutting of belief that we have been considering.

   The same holds true for the overall story about the undercutting defeat of visual experience we've been considering. To see this, recall the details of such undercutting. In the Visual Case R is the claim that there is a red object before you, and U is the claim that tricky lighting makes non-red objects look red. When v is a visual experience of R and y is a belief in U, v is a reason to believe R and y is an undercutting defeater for v as a reason to believe R. According to the line before us, y undercuts v as a reason to believe R because y is itself reason to believe a complex negation: ¬(v ⊕ R). This is the claim that it is not the case that you experience as of a red object before you because there is a red object before you. The key thought behind Answer 1 will attempt to underwrite this overall story by saying that it is rationally impossible conjointly to believe both R and ¬(v ⊕ R), i.e. it is rationally impossible conjointly to believe that there is a red object before you but that it is not the case that you experience as of a red object before you because there is a red object before you.

That is also not true. In the original Visual Case, after all, that is precisely what you should do if you come across further evidence that you suffer veridical hallucination. In that event you should believe precisely that there is a red object before you while denying, because of the tricky lighting, that it looks to you as if there is a red object before you because there is such an object before you. Answer 1 does not render cogent the story about undercutting of visual experience found at (EUDV).

Answer 2. When dealing with a composite state ($x+y$), we presuppose that one can rationally believe a claim Φ on its basis, which can be rationally believed on the basis of x or y alone, **_only if_** it is rationally possible to believe all such Φ conjointly on the basis of (x+y). When x is a reason to believe Φ and y is a reason to believe Ψ, therefore, the presupposition of Answer 2 is that composite state ($x+y$) will be a reason to believe Φ, or reason to believe Ψ, only if it is rationally possible to believe Φ and to believe Ψ conjointly on the basis of ($x+y$).

This story also renders cogent the explanation of rebutting defeat found at (ERD); and it does so, of course, for the same reason that Answer 1 did so. It is not rationally possible conjointly to believe Φ and $\neg\Phi$. Hence it is not rationally possible to do so on the basis of ($x+y$). When x is a reason to believe Φ, and y is a reason to believe $\neg\Phi$, therefore, the composite state ($x+y$) will be composed of elements which are each reasons for claims it is not rationally possible conjointly to believe. It follows that ($x+y$) is composed of elements which are reasons for claims it is not rationally possible conjointly to believe on the basis of ($x+y$). Answer 2 thus renders cogent the explanation of rebutting defeat found at (ERD).

Moreover, that Answer helps underwrite the explanations of undercutting defeat canvassed earlier. It has positive purchase not only on the story about undercutting defeat of belief—found at (EUDB)—but also on the story about undercutting defeat of visual experience—found at (EUDV). When x is belief in claim T—that 87% of Texas respondents said they will vote Republican, and y is belief in the claim U—that respondents decided their answer by coin flip—then, intuitively, x is a reason to believe G—the claim that roughly 87% of Texans will vote Republican—and y is an undercutting defeater for x. But as we have seen: it is rationally possible to believe that roughly 87% of Texans will vote Republican while also denying that the testimonial evidence turned out as it did because of the political facts on the ground. It is _not_ rationally possible, however, to believe both these things _on the basis of_ ($x+y$). It is not rationally possible to believe—on the basis of a composite state consisting of a belief that 87% of respondents said they'd vote Republican and a belief that they did so by appeal to coin flips—that roughly 87% of Texans will vote Republican while also denying that 87% of respondents said that they would vote Republican because roughly 87% of Texans intended to vote Republican. Answer 2 renders cogent the explanation of undercutting of belief found at (EUDB).

It also renders cogent the explanation of undercutting of visual experience. When v is such an experience of R—the claim that there is a red object before you— and y is belief in the claim U—that tricky lighting makes non-red objects look red— then, intuitively, v is a reason to believe R and y is an undercutting defeater for v as a reason to believe R. But as we have seen: it is rationally possible to believe that there is a red object before you while denying that it looks to you as if there is because there is. It is _not_ rationally possible, however, to believe both these things _on the basis of_ ($v+y$). It is not rationally possible to believe—on the basis of a composite state consisting of a visual experience as of a red object before you, and a belief that tricky lighting makes non-red objects look red—that a red object is before you while

9

denying that it looks to you as if a red object is before you because a red object is before you. Answer 2 renders cogent the explanation of undercutting of visual experience found at (EUDV).

The hope, then, is that the basic approach to reasons and defeaters before us turns out to be cogent if we assume that whenever one can rationally believe a claim Φ, on the basis of composite state (*x+y*), which can also be rationally believed on the basis of a component of (*x+y*), it is rationally possible to believe every such Φ conjointly on the basis (x+y). This assumption would at least render sensible an aspect of the approach which looks decidedly puzzling at first, namely, the fact that on the approach defeaters function _as such_ by virtue of being reasons to believe. One could be forgiven for not understanding this at first, for puzzling at the thought that being a reason to believe could itself invest a state with the power to wipe out another state's effectiveness as a reason to believe. Assumption 2 may be wrong, of course. But it helps to make sense of a very puzzling aspect of a Pollock-style approach to reasons and defeaters; and it does so for rebutting and undercutting defeaters alike.

Having said that, even if we accept Answer 2 we can make trouble for the general approach to undercutting defeat before us. Indeed with a bit of work we can see that undercutting defeaters turn out to function in an entirely different way than their rebutting cousins. Explaining why the Pollock-style approach to undercutting is no good will be our next task in the next section. Explaining the ways in which undercutting defeaters function differently than rebutting defeaters will be done in the section after that.

Counter-examples to Orthodoxy about Undercutting Defeaters.
I begin with counter-examples to both directions of the easy-to-understand story about undercutting defeat. This will be done for the story's *spiel* about the undercutting of visual experience. It will be obvious, though, how to alter the examples so that they cover other forms of experiential reason and the undercutting defeat of belief. Once the easy-to-understand story has been found wanting, we'll turn our attention to Pollock's officially story about undercutting defeat. By then it will be clear how to construct counter-examples to that story too.

Consider the following vignette:

The Milk Taster

Subject S tastes a bit of milk to see if it's gone off. Being a normal milk taster, S is unaware that her view of the milk is based on smell more than taste. Indeed we may suppose that S believes her view of the milk is not based on smell, not even in part, even though she knows, of course that she smells the milk as well as tastes it. When the milk taster imbibes the milk, however, she has a complex gustatory and olfactory experience; and like other normal milk tasters she comes to believe on its basis that the milk is o.k. Like those milk-tasters, though, S is unaware that she bases her view of the milk on smell.

Suppose the milk taster is then told, by someone she trusts, that her nose is bunged up, that she is subject to random olfactory hallucination. This leads her, after a bit of reflection, to deny that her overall gustatory and olfactory experience of the milk was due to the milk being o.k. After all, she realises that her overall gustatory and olfactory experience of the milk includes the olfactory part of that experience; and she believes herself to

10

be subject to random olfactory hallucination. Nevertheless, her new information does not, and should not, lead her to change her view of the milk. She continues rationally to believe that the milk is o.k.; and she continues to do so on the basis of her complex gustatory and olfactory experience.

Let e be the milk taster's complex experience of the milk, OK be the claim that the milk is o.k., and H be the claim that the milk taster is subject to random olfactory hallucination. Then we have the following in the Milk Taster case:

(i) e is a reason to believe OK: it is possible to become justified in believing OK on the basis of e.

(ii) Belief in H is a reason to believe $\neg(e \oplus OK)$: the milk taster can become justified in believing, on the basis of her belief in H, that it is not the case that her overall gustatory and olfactory experience of the milk is produced by the milk.

(iii) Even after coming to believe H, the milk taster is rational in continuing to believe, on the basis of her overall gustatory and olfactory experience of milk, that the milk is o.k.

From these facts it follows that satisfying the easy-to-understand approach to the undercutting of experience is not actually sufficient for the undercutting of experience. With arrows marking the basing relation in thought, we can diagram the moral here as follows: when someone manifests this pattern of thought

$$B(\Phi) \qquad\qquad B[\exp(\Phi) \oplus \Phi]$$
$$\uparrow \qquad\qquad\qquad \uparrow$$
$$\exp(\Phi) \qquad\qquad B(U),$$
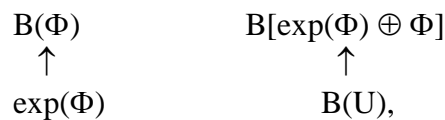
the easy-to-understand approach to undercutting entails that U is an undercutting defeater for Φ-experience as a reason to believe Φ. That is simply not so. Milk tasters demonstrate that manifesting the cognitive situation diagrammed is insufficient for the undercutting defeat of experience. In turn that cuts against the easy-to-understand approach to such defeat.

Now consider another vignette

The Not-Because-Presupposer

Subject S is a normal person who begins with belief in the complex negation said by the easy-to-understand approach to be evidentially supported by undercutting defeaters. Let her start out with closed eyes, a desire to know if a red thing is before her, and a steadfast *presupposition* that it is not the case that the world will look as it does, *in situ*, because the world is in fact the way it looks. S opens her eyes, it looks as if a red thing is before her, and S comes to believe on that basis that a red thing is before her.

She should not have formed the belief. S has undercutting defeat for her visual experience as a reason for her belief. But she does not have an

undercutting defeater in the easy-to-understand sense; after all, she does not accept the negation of the complex because-claim in play here on the basis of a reason which supports that complex negation. She merely presupposes the negation is true, i.e. she merely presupposes that it is not the case that the world looks as it does, *in situ*, because the world is the way it looks. Contrary to the easy-to-understand approach, such a presupposition alone is sufficient for undercutting defeat.

Let *v* be the presupposer's visual experience as of a red object before her, R be the claim that a red object is before her, and ¬[*v* ⊕ R] be the complex negation in play. By stipulation S presupposes this complex negation. So we have the following in the Not-Because Presupposer case:

(i)   *v* is a reason to believe R: it is possible for the presupposer to become justified in believing R on the basis of *v*.

(ii)   The presupposer believes ¬(*v* ⊕ R) but not on the basis of a reason (or anything else). She merely presupposes that it is not the case that it looks as if R because R.

(iii)   The presupposer should not come to believe R on the basis of *v*: her belief that a red object is before her is irrationally formed on the basis of its looking as if a red object is before her.

From these facts it follows that satisfying the easy-to-understand approach to undercutting defeat is not necessary for such defeat. When someone manifests this pattern of thought

$$B(R)$$
$$\uparrow$$
$$v(R) \qquad B[v(R) \oplus R],$$

the easy-to-understand approach entails that *v*(R) is an undefeated reason to believe R. That is not so. The Not-Because-Presupposer makes clear that manifesting the cognitive situation above is sufficient for undercutting defeat. In turn that cuts against the easy-to-understand approach to such defeat.

We may conclude that the easy-to-understand approach is no good. Its framing conditions are neither necessary nor sufficient for undercutting defeat. Milk Tasters show that those conditions can happen when undercutting defeat fails to occur. Not-Because-Presupposers show that undercutting defeat can happen when conditions put forth by the easy-to-understand view fail to occur. But that very view is not Pollock's, of course; it is a close cousin of Pollock's approach which is simply easier to understand. Perhaps the view's difficulties go away once complications are introduced, complications which make exactly for a harder-to-understand Pollock-style approach to undercutting defeat.

Unfortunately, difficulties faced by the easy-to-understand approach are close cousins of difficulties faced by Pollock's harder-to-understand view(s). To see this, recall how we arrived at the easy-to-understand theory. In the Polling Case: T is the claim that 87% of respondents said that they will vote Republican, G is the claim that roughly 87% of Texans will vote Republican, and U is the undercutting claim that

12

respondents decided their answer by coin flip. Pollock and Cruz described U as "a reason for doubting or denying that you would not have the inductive evidence unless G were true". This led to the easy-to-understand idea that undercutting defeaters for belief are reasons for denying that one truth happens because of another, and the equally easy-to-understand idea that undercutting defeaters for experience are reasons for denying that experience which portrays the world a certain way occurs because the world is that way.

Just after making the remark which sets up these easy-to-understand ideas, Pollock and Cruz re-describe U as reason to believe that "it is false that T would not be true unless G were true",[5] and then they rephrase this new idea with the following gloss:

> "This can be read more simply as 'T does not guarantee G'. Undercutting defeaters [of belief] are reasons for the claim that T does not guarantee G."[6]

Further still, in his classic paper "Defeasible Reasoning" Pollock describes undercutting defeaters as

> "reason for thinking it false that the premises of the inference [they attack] would be true unless the conclusion were true. More simply, we can think of [an undercutting defeater of belief] as giving us a reason for believing that (under present circumstances) the truth of the premises do not guarantee the truth of the conclusion."[7]

These extra remarks on undercutting defeater complicate our discussion. They lead to a pair of hard-to-understand takes on the complex negation being reason for which is definitional, on Pollock's approach to undercutting defeat, of being an undercutting defeater. In turn this is true because the extra remarks lead to a pair of hard-to-understand takes on the claim negation of which plays the key role in Pollock's approach. As we are about to see neither of these hard-to-understand claims is equivalent to the other; and neither is equivalent to the easy-to-understand claim used earlier.

In the Polling Case, as we have seen, one claim the negation of which plausibly plays the key role is the claim that T is true because G is true (i.e. the claim that 87% of polled Texans said that they will vote Republican because roughly 87% of Texans will vote Republican). Earlier we symbolized the claim this way:

$$[\Phi \otimes \Psi] \qquad\qquad \Phi \text{ is true because } \Psi \text{ is true.}$$

The further remarks on the Polling Case quoted above generate two more claims which might play the key role in a Pollock-style theory of undercutting defeat. One is a subjunctive conditional we'll symbolize this way

[5] *Ibid.* P197.
[6] *Ibid*
[7] (1987): p.486.

[$\Phi \dashrightarrow \Psi$] $\qquad$ Φ wouldn't be true unless Ψ were true.[8]

And the other is a claim about some kind of situational guarantee we'll symbolize this way

[$\Phi \,_{pc}\!>>\Psi$] $\qquad$ Φ's truth guarantees Ψ's truth, in present circumstances.

But notice: instances of any of these schemata do not entail analogue instances of the others. It might be the case that Φ is true because Ψ is true, for example, but not the case that Φ wouldn't be true unless Ψ were true, and also not the case that Φ's truth guarantees Ψ's truth *in situ*. Just think of a situation in which Φ is true because Ψ is true, but where Ψ concerns a weak back-up system *in situ*, one which happened quite improbably to have overridden the situation's dominant system Δ. Or it might be the case that Φ's truth guarantees Ψ's truth *in situ*, but neither the case that Φ is true because Ψ is true, nor the case that Φ wouldn't be true unless Ψ were true. Just think of a situation in which a weak back-up system Σ makes both Φ and Ψ true, while the dominant system Δ, the one overwhelmingly likely to have activated, only makes Φ true when activated. And so on. Instances of each of the indented schemata fail to entail analogue instances of the others.

$\qquad$ Yet each the schemata can be used in a Pollock-style approach to undercutting defeat. We have seen that the view which results when the easy-to-understand schema is used is itself no good. We are about to see that its problems have direct analogues for approaches based on the less-easy-to-understand schemata. Since the subjunctive-based one is by far the most common in Pollock's work on undercutting defeat, that will be our focus. Difficulties faced by the view of such defeat resulting from its use have direct analogues for the approach to undercutting defeat built from the other hard-to-understand schema.

$\qquad$ Now, Pollock's work on undercutting defeat makes heavy use of subjunctive conditionals. His view of the undercutting of belief is normally put this way:

(UDB-Pollock) $\quad$ B(Δ)**UD**[B(Φ)**R**B(Ψ)] $\qquad$ =$_{df.}$ $\quad$ (i) $\quad$ B(Φ)**R**B(Ψ),

$\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ & (ii) $\quad$ B(Δ)**R**B($\neg$[$\Phi \dashrightarrow \Psi$]).

Belief in Δ is said to be an undercutting defeater for belief in Φ as a reason to believe Ψ exactly when belief in Φ is a reason to believe Ψ, and belief in Δ is a reason to believe that it is not the case that Φ wouldn't be true unless Ψ were true. And each time Pollock offers a subjunctive-based approach to the undercutting defeat of belief he extends it directly to the undercutting of experience. With visual experience as our exemplar:

(UDV-Pollock) $\quad$ B(Δ)**UD**[B(Φ)**R**B(Φ)] $\qquad$ =$_{df.}$ $\quad$ (i) $\quad$ V(Φ)**R**B(Φ),

$\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ & (ii) $\quad$ B(Δ)**R**B($\neg$[V(Φ)$\dashrightarrow \Phi$]).

---

[8] I choose the dotted arrow here because it is quite clear from Pollock's career-long work on undercutting defeat that he thought some kind of binary connective was appropriate; but it is equally clear that he didn't not have a stable conception of how to vocalize that connective in English.

Belief in Δ is said to be an undercutting defeater for visual experience as of Φ as a reason to believe Φ exactly when visual experience as of Φ is a reason to believe Ψ, and belief in Δ is a reason to believe that it is not the case that you wouldn't visually experience as of Φ unless Φ were true.

   This approach to undercutting defeat will not work. Like its cousin the easy-to-understand theory, conditions it lays down for undercutting defeat are neither necessary nor sufficient for such defeat. It is straightaway clear, for instance, that a presupposer counter-example plagues the view. Consider the following vignette:

<p align="center">The Not-Subjunctive-Presupposer</p>

 Subject S is a normal person who begins with belief in the complex negation said by (UDV-Pollock) to be evidentially supported by undercutting defeaters. Let her start out with closed eyes, a desire to know if a red thing is before her, and a steadfast presupposition that it is not the case that the world wouldn't look as it does if the world weren't the way that it looks. S opens her eyes, it looks to her as if a red thing is before her, and, on that basis, S comes to believe that a red thing is before her.

 S should not have formed the belief. She has undercutting defeat for her visual experience as a reason for her belief. But S does not have an undercutting defeater in the style of (UDV-Pollock). After all, she does not accept the negation of the complex subjunctive on the basis of a reason supporting that negation. She presupposes the negation to be true. She presupposes that it is not the case that the world wouldn't look as it does if the world weren't the way that it looks. Contrary to (UDV-Pollock), this presupposition alone is sufficient for undercutting defeat.

Let v be S's visual experience as of a red object before her, R be the claim that a red object is before her, and $\neg[v \dashrightarrow R]$ be the complex negation in play. By stipulation S presupposes this negation. So in the Not-Subjunctive Presupposer case we have:

(i) v is a reason to believe R: it is possible for the presupposer to become justified in believing R on the basis of v.

(ii) S believes $\neg(v \dashrightarrow R)$ but not on the basis of a reason (or anything else). She presupposes that it is not the case that it wouldn't look as if R if it weren't the case that R.

(iii) S should not come to believe R on the basis of v: her belief that a red object is before her is irrationally formed on the basis of its looking as if a red object is before her.

From these facts it follows that satisfying (UDV-Pollock) is not necessary for undercutting defeat. When someone manifests this pattern of thought

$$B(R)$$
$$\uparrow$$
$$v(R) \qquad B[v(R) \dashrightarrow R],$$

Pollock's theory entails that $v(R)$ is an undefeated reason to believe R. That is not so. The Not-Subjunctive-Presupposer makes clear that manifesting the cognitive situation above is sufficient for undercutting defeat. In turn this means that satisfaction of Pollock's conditions for such defeat is not actually required for it to occur.

Nor is the satisfaction of Pollock's conditions sufficient for undercutting defeat. Consider our final vignette:

The Mixed-up Assessor

Subject S is paid to watch a parade of Fs go by on a conveyor belt. The Fs come in two varieties: V1 and V2. S is paid to discriminate between them observationally. Like everyone, though, S is mixed-up about her capacity to do so. She thinks it is a purely visual capacity when in fact it is an olfactory one. S believes that when she is confronted with an F, somehow, she exploits subtle visual cues to register whether the F before her is V1 or V2. In fact S exploits subtle olfactory cues to make the discrimination. Sight plays no role at all.

S is aware, of course, that she smells Fs as well as sees them as they parade by on the belt. But S mistakenly believes—like everyone else—that Fs all smell the same to her while subtly looking different from one another. In fact the reverse is true: S discriminates V1s from V2s by exploiting subtle olfactory ways in which Fs are portrayed to be different from one another.

Suppose S is faced with an F. It looks and smells a certain way to her. In deploying her capacity to discriminate, S comes rationally to believe, on the basis of how the F smells, that it is V1 rather than V2.

Then S is told, by someone she trusts, that her nose is bunged up, that she is subject to random olfactory hallucination. This leads her to deny, after a bit of reflection, that the F wouldn't have smelled as it does if it weren't as it is. After all, she believes herself to be subject to random olfactory hallucination. Yet the information to that effect does not, and should not, lead her to change her view of the F. She continues rationally to believe that the F is V1 rather than V2; and she continues rationally to do so on the basis of how the F smells to her.

Let e be the mixed-up assessor's olfactory experience of the F, V1-F be the claim that the F before her is V1 rather than V2, and H be the claim that S is subject to random olfactory hallucination. Then we have the following in the Mixed-up Assessor case:

(i) e is a reason to believe V1-F: it is possible to become justified in believing V1-F on the basis of e.

(ii)   Belief in H is a reason to believe $\neg(e \dashrightarrow$ V1-F): the assessor can become justified in believing, on the basis of her belief in H, that it is not the case that the F before her wouldn't have smelled as it did if it weren't that way.

(iii)   Even after coming to believe H, the assessor is rational in continuing to believe, on the basis of her olfactory experience of the F before her, that the F in question is V1 rather than V2.

From these facts it follows that satisfying Pollock's subjunctive-based approach to the undercutting of experience is not sufficient for such undercutting. With arrows marking the basing relation in thought, we can diagram the moral as follows: when someone manifests this pattern of thought

$$B(\Phi) \qquad\qquad B[\exp(\Phi) \dashrightarrow \Phi]$$
$$\uparrow \qquad\qquad\qquad \uparrow$$
$$\exp(\Phi) \qquad\qquad B(U),$$

Pollock's subjunctive-based approach entails that U is an undercutting defeater for Φ-experience as a reason to believe Φ. That is simply not so. Mixed-up assessors demonstrate that manifesting the cognitive situation diagrammed is insufficient for the undercutting defeat of experience.

      Something is fundamentally wrong with Pollock-style approaches to undercutting defeat. Next we begin to make a case that such defeat works in an entirely different way than rebutting defeat. The key thought will be located at the heart of an unpublished thought experience of Dorothy Edgington.

Edgington's Burglar: Some Notation.
The original thought experiment appears in an unpublished paper called "Tale of a Bayesian Burglar". That paper was sent privately by Edgington to David Lewis. It prompted Lewis to write an unpublished reply called "Advice to a Bayesian Burglar". As the titles make clear, the original discussion was framed within a Bayesian setting: epistemic agents were supposed to have point-valued degrees of belief, with the crucial issue at hand being how such degrees of belief should be updated.

      In our discussion of undercutting defeat, however, we have proceeded under the assumption that belief is undercut, not degree of belief. That is standard practice in the literature, of course; but it is entirely inessential to the topic of undercutting defeat. After all, everyone (these days) agrees that experiential input can be undercut, no matter whether its initial impact is on belief or its degrees. This means that undercutting defeat requires theoretical treatment within the epistemology of degrees of belief just as it does within the epistemology of belief.

      In what follows we prescind, therefore, from any take on the grain of states undercut by new information. We make no assumption about whether those states are coarse-grained like belief, disbelief or suspended judgement, or whether they are fine-grained like point-valued subjective probabilities. We distinguish simply between strong positive attitudes taken to a claim, neutral attitudes taken to a claim, and strong negative attitudes taken to a claim. And we represent them respectively as

      ☺(Φ)   =   strong pro attitude taken to Φ
      ☺(Φ)   =   neutral attitude taken to Φ
      ☹(Φ)   =   strong con attitude taken to Φ.

Now suppose you are a burglar. We set up our Burglar Case by appeal to three claims to which you take attitudes:

A   =   There is an **_a_**larm on a given house.
B   =   You **_b_**reak into the house.
C   =   You get **_c_**aught.

In the burglar case you begin with an initial take on things, then look at the house before you, and then update your view of the world. Hence you manifest two overall takes in the relevant thought experiment. Your initial take on the world is **old**(-); and your post-experience take is **new**(-).

Unsurprisingly, you want to break into the house but you do not want to get caught. You start with a strong pro attitude taken to the claim that you will get caught, given you break into the house and it has is an alarm. You start with a strong con attitude taken to the claim that you will get caught, given you break in and the house does not have an alarm. You know that you will look to see whether the house has an alarm, that you are next to certain to detect whether it does so, and that you will make up your mind whether to break in on that basis. You begin with a neutral take on whether the house has an alarm.

In the notation we shall be using the set-up is this:

Desire: (B+¬C)
old-☺(C/BA)
old-☹(C/B¬A)
☺(A).

But you are a sensible burglar. So the initial set-up makes for two further aspects of your initial take on things:

       old-☹(B/A)

and

       old-☺(B/¬A).

You begin with a strong con attitude to your breaking into the house given it has an alarm; and you begin with a strong pro attitude to your breaking into the house given that it has no alarm. Thus we have a key consequence of the initial burglar set-up:

       old-☹(B/A) ≠ **old**-☺(B/¬A).

Next you look to see whether there is an alarm on the house. It visually appears as if there is an alarm on the house. This prompts a shift in your view. Your old neutral take on whether there is an alarm on the house--**old**-☺(A)--is replaced with to a new pro take on the claim that there is--**new**-☺(A).

Then something interesting should happen. You should go from a state in which your take on breaking in given there is an alarm is distinct from your take on breaking on given no alarm--as we have just seen--to a state in which the conditional commitments are identical. In other words: after updating on visual experience you should go from an old take on which

       old-(B/A) ≠ **old**-(B/¬A)

to a new take on which

$$\textbf{new}(B/A) \; = \; \textbf{new}(B/\neg A).$$

This means updating on your visual input makes your take on your breaking in go _from_ being dependent on whether there is an alarm on the house _to_ being independent of whether there is such an alarm.

Why is that? Well, as David Lewis put it:

> "[There is] a good reason why **old-**☹(B/A) and **old-**☺(B/¬A) should differ: before you looked, you thought that if there was an alarm you would most likely spot it and be deterred, whereas if there wasn't you would probably be undeterred and go ahead. But after you've already done your looking, and revised your [view], this reason no longer applies. There is no good reason why **New**(B/A) and **New**(B/¬A) should differ. Rather, they should be equal…for once you have finished looking the influence of the burglar alarm on whether you break in or not is over and done with."[9]

Updating on visual input shifts your take on breaking in from one on which it is dependent on whether there is an alarm to one on which it is not. This is so because you knew that facts about the alarm would control for your break-in behavior via your visual input, that mediating input had happened, and that your new take on there being an alarm was based on that input.

The lesson is both stark and general. Updating on visual episodes does not solely involve shifting commitment about the seen world. It also involves shifting commitment about the way in which first-order commitment about the seen world is itself founded. Visual updating routinely involves higher-order commitment about the basing of first-order commitment. When we observe the world and change our mind about it on the basis of that observation, we also routine mark the fact that our first-order commitment is itself based on observation; and we do so with a higher-order commitment, another belief or investment of credence or some such. We register that our new view of the world is got by appeal to observation.

This happens so routinely that it has gone largely unnoticed in the literature. But it is crucial to the way that undercutting defeat works. There is a tight connection between when a commitment is undercut by new information and when there is a higher-order commitment about the way in which the undercut commitment is itself based. Explaining this is my next task.

Higher-order Mental States and Undercutting Defeat
Suppose you adopt a strong pro take on a claim Φ. Let U be the claim that source of information S is untrustworthy about Φ. Let Φ-BOS be any claim which forges a strong link between your take on Φ and source of information S.[10] Then a bi-conditional of note will be true:

(↔)   ☺(U) undercuts ☺(Φ) ↔ (∃x)x = ☺(Φ-BOS).

---

[9] "Advice to a Bayesian Burglar". I have brought notation into line with this paper.

[10] Φ-BOS might be the claim that your take on Φ is based on S, or the claim that your take on Φ is likely to be based on S, or the claim that your take on Φ has a high objective chance of being based on S, etc. Φ-BOS can be any claim commitment to which amounts to commitment to there being a strong link between source of information S and your take on Φ.

In these circumstances a strong pro take on U undercuts such a take on Φ exactly when there is commitment to a strong link between one's take on Φ and source of information S.

Consider the right-to-left direction of (↔):

(RTL-↔): Having ☺(Φ-BOS) → ☺(U) undercuts ☺(Φ).

This looks plausible straightaway. If you are committed to the view that your strong pro take on Φ is itself based on source of information S, for instance, and then come to adopt a strong pro take on the view that S is untrustworthy about whether Φ is true, that strong pro take on S's untrustworthiness undercuts your strong pro take on Φ. And this looks to be so whether or not your old take on Φ was in fact based on source of information S.[11]

Similarly, consider the left-to-right direction of (↔):

(LTR-↔) ☺(U) undercuts ☺(Φ) → One has ☺(Φ-BOS).

This also looks plausible. Suppose you are rationally committed to Φ but are not committed to there being a link between your commitment to Φ and source of information S. Then you come to adopt a strong pro attitude to the view that S is untrustworthy about whether Φ is true. Is your take on Φ undercut by your new commitment? The milk-taster and mixed-up assessor scenarios make it clear that the answer is No. If you do not have a strong pro attitude to the view that your take on Φ is linked to S, therefore, coming to adopt such an attitude to the view that S is untrustworthy about Φ will not undercut your take on Φ. The left-to-right direction of (↔) is just the contraposition of this thought.

The plausibility of each direction of (↔) suggests that we distinguish two fundamentally different types of defeat. The idea would be that rebutting defeaters generate a distinctive kind of pressure on their own—all by themselves, as it were—whereas undercutting defeaters generate a distinctive kind of pressure only in concert with higher-order commitment about the basing of lower-order commitment. That is a main suggestion of this paper.

In the event, orthodoxy about rebutting defeat is fine. Rebutting defeaters do their work in splendid isolation, by being reasons to adopt a strong pro commitment of a certain sort. On the approach to undercutting defeaters suggested here, however, such defeaters work in an entirely different way; and that is why orthodoxy about them breaks down. Undercutting defeaters do their work in tandem with other mental states. They join forces with higher-order commitment about the basing of lower-order commitment.

Think of how dogmatists see the epistemic role of perceptual experience.[12] They say that visual states, for instance, generate epistemic pressure on their own to adopt a strong pro attitude to their contents, without appeal to background commitment or achievement. Non-dogmatists deny this, of course, insisting that visual states generate epistemic pressure only in concert with background commitment or achievement. My suggestion is that rebutting defeaters are the defeat-

---

[11] The same would be true for any other higher-order content commitment to which entails commitment to a strong link between S and your commitment to Φ.

[12] The term "dogmatism" comes from Pryor (2000). Pollock was an early and vigorous defender of the view: see his (1974) or (1987), for instance.

theoretic analogue of dogmatic experience, and undercutting defeaters are the defeat-theoretic analogue of non-dogmatic experience. Rebutting defeaters function the way dogmatists say perceptual experience works—in line with orthodoxy about rebutting defeat, without reliance on background epistemic factors; and undercutting defeaters function as non-dogmatists insist that perceptual experience works—contrary to orthodoxy about such defeat, in tandem with background commitment about the basing of lower-order commitment.

        If this suggestion is right, the Defeater's Assumption is only partly correct. In our current terminology that assumption is this:

The Defeaters Assumption.   Defeaters do their work because they are reasons to adopt a strong pro epistemic stance; they generate their distinctive kind of epistemic pressure—defeating epistemic pressure—in virtue of being reasons to do so.

My suggestion is that the Defeaters Assumption is true only for rebutting defeaters. Only they generate defeating pressure in virtue of being reasons to adopt a strong pro attitude. Undercutting defeaters do not work that way. They join forces with higher-order commitment about the basing of such attitude, with the resulting complex of mental states being no reason to believe at all. Rather, it is a reason to avoid basing commitment on untrustworthy sources of information. When an agent is committed to a strong link between her pro take on Φ and source of information S, then, and only then, does commitment to the Φ-untrustworthiness of S undercut her take on Φ. Rebutting and undercutting defeaters work in fundamentally different ways.

References

Edgington, Dorothy. Unpublished: "Tale of a Bayesian Burglar".

Lewis, David. Unpublished: "Advice to a Bayesian Burglar".

Pollock, John. 1967: "Criteria and Our Knowledge of the Natural World", *Philosophical Review*.

---- 1970: "The Structure of Epistemic Justification", *American Philosophical Quarterly*.

---- 1974: *Knowledge and Justification* Princeton University Press.

---- 1987a: *Contemporary Theories of Knowledge* Rowman & Littlefield. Second edition co-authored with Joe Cruz.

---- 1987b: "Defeasible Reasons", *Cognitive Science*.

Pryor, Jim. "The Skeptic and the Dogmatist", *Nous*.