Chapter 14
A Formal Account of Epistemic Defeat

Matthew Kotzen

Abstract The goal of this chapter is to disentangle several related—but importantly distinct—notions of evidential defeat. The broadest distinction in the literature on epistemic defeat is that between rebutting and undercutting defeat; very roughly, the idea is that rebutting defeaters provide a "positive" reason to disbelieve the conclusion, whereas an undercutting defeater merely "blocks" existing reasons to believe the conclusion. In this chapter, I formalize these two notions and explore some related (and under-discussed) phenomena such as "hybrid" defeat (where a single defeater can both rebut and undercut), "bidirectional" defeat (where some information that serves as evidence for a conclusion can become a defeater in the presence of another piece of evidence for the conclusion), and "redundant" defeat (where an undercutting effect is generated by the non-independence of two pieces of information, rather than by the "blocking" phenomenon that occurs in more typical cases of undercutting).

Keywords Formal epistemology · Degrees of justification · Evidential defeat · Belief revision · Justification defeat · Bayesian epistemology · Types of defeaters · Bayesian approaches to defeaters · Epistemology

14.1 Introduction

It is standard in epistemology to distinguish between two different kinds of defeaters of *S*'s evidence for her belief that *q*—"undercutting" defeaters and "opposing"[1] defeaters. Very roughly, an undercutting defeater undermines the effect that the

The original version of this chapter was revised: typographical errors were corrected throughout the chapter. The correction to this chapter is available at https://doi.org/10.1007/978-3-030-04522-7_16

[1] This latter class is more commonly referred to as "rebutting" defeaters; for reasons that will become clear later, I prefer the term "opposing." The term is due to Jim Pryor.

M. Kotzen (✉)
UNC Department of Philosophy, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA
e-mail: kotzen@email.unc.edu

B. Fitelson et al. (eds.), *Themes from Klein*, Synthese Library 404,
https://doi.org/10.1007/978-3-030-04522-7_14

evidence has on q, without directly providing evidence against q.[2] An opposing defeater, on the other hand, serves as "direct" evidence against q.

Suppose, for instance, that Julio tells me that it is raining outside and that I believe on that basis that it is raining outside. If I were to learn that Julio in fact has no idea what the weather is like right now but likes to make up claims about the weather and confidently share them with others, that would plausibly serve as an undercutting defeater of my evidence that it's raining. This new information about Julio isn't "directly" evidence that it's not raining outside; rather, it merely undermines the evidential force that I originally took his testimony to have. If, however, I were to go outside and see for myself that it's sunny, that experience would serve as an opposing defeater of my evidence that it's raining; here, my new experience does serve as "direct" evidence *against* the proposition that it's raining.

One thing we might wonder about is whether we can draw this distinction in a more principled and precise way—and, if so, how we should go about doing it. This is particularly important because some philosophers—for instance, Jim Pryor (see Manuscript and 2013)—have expressed doubts about whether the distinction can be captured using the standard Bayesian probabilistic model of partial belief. If this doubt is justified, then it provides some reason to pursue a richer formal model that can capture the distinction; after all, we want a formal model of belief to capture those distinctions that play a crucial role in epistemology.

In this chapter, I will give a formal account of epistemic defeat generally, and also of the distinction between undercutting and opposing defeaters, using the standard Bayesian apparatus. I will also introduce a third kind of defeat that seems to have gone unnoticed in the epistemological literature, which I call "bidirectional" defeat. My account will also handle cases of "partial" defeat, where the evidential effect under consideration is only partly undermined, as well as cases of "hybrid" defeat, where the same defeater plays both an undercutting and an opposing role (or an undercutting and a supporting role). I will close by addressing a concern that my account cannot properly distinguish undercutting defeat from evidential redundancy.

14.2 A First Stab

Consider again the case where I take Julio's testimony that it's raining to be some evidence that it's raining.

Let E be Julio's testimony that it's raining out and let H be the proposition that it is raining out. Let's assume that, given the background information that I have at the

[2] There has been a lot of discussion in the literature recently on so-called "higher-order" defeaters, which work by inducing doubts about the reliability of the cognitive process(es) that produced a belief. See, e.g., Christensen (2007a, b, c, 2009, 2010), Elga (unpublished), Kelly (2010), Lasonen-Aarnio (2014), and Schechter (2011). I consider these sorts of higher-order defeaters to be undercutting defeaters, though I will for the most part focus on lower-order kinds of undercutting defeat in this chapter.

time that I hear Julio's testimony, E is evidence (for me) for H. In Bayesian terms, this amounts to the assumption that (my) $p(H|E) > p(H)$.

Call my credence in H before collecting E, $p(H)$, my **prior credence in *H***. Call my credence in H after collecting E but before acquiring any defeaters, $p(H|E)$, my **evidential credence in *H***.[3] Call my credence in H after collecting E and after acquiring defeater D, $p(H|\ E\ \&\ D)$, my **defeated credence in *H***.

In my rough characterization above of the difference between undercutting and opposing defeaters, I suggested that while opposing defeaters constitute direct evidence against the relevant proposition, undercutting defeaters do not. Undercutting defeaters merely undermine the evidential force of some other evidence for the proposition, and therefore do not constitute direct evidence *against* the proposition. One might think, then, that we can characterize opposing defeaters as those propositions which, if learned, lower our credence in H (on the assumption that E is known):

> Opposing IFF Credence-Lowering: D is an opposing defeater for the evidence that E provides for H just in case $p(H|\ E\ \&\ D) < p(H|E)$.

But this clearly will not work. However we should understand the distinction between the "directness" of opposing defeaters and the "indirectness" of undercutting defeaters, both kinds of defeaters are defeaters, and both can therefore lower our credence in H. Take, for instance, the information that Julio is a highly unreliable testifier about the weather—a paradigm case of an undercutting defeater (of Julio's testimonial evidence that it's raining). Once I learn this, it seems clear, I ought to become less confident that it's raining than I was before finding out how unreliable Julio is. Thus, in this case, it's true that $p(H|\ E\ \&\ D) < p(H|E)$ even though the D in question is an undercutting defeater. Therefore, Opposing IFF Credence-Lowering can't be used to uniquely characterize opposing defeaters.

Still, it's somewhat plausible that the condition that $p(H|\ E\ \&\ D) < p(H|\ E)$ characterizes defeaters in general, even if it doesn't distinguish between opposing and undercutting defeaters. It's plausible that, when E's evidential effect on H is defeated, the agent's defeated credence in H is lower than her evidential credence. So let's provisionally accept:

> Defeater IFF Credence-Lowering: D is a defeater for the evidence that E provides for H just in case $p(H|\ E\ \&\ D) < p(H|E)$.

One advantage of Defeater IFF Credence-Lowering over several extant theories is that it accounts for cases of *partial* defeat which do not cross the "threshold of binary belief"—i.e., cases where the defeater lowers the agent's credence in H from its evidential value, but does not lower it from a point above the threshold to a point below the threshold. Since several other theories of defeat focus only on binary beliefs, they are insensitive to this kind of partial defeat.

[3] Here and throughout, I will assume Conditionalization—i.e., I will assume that the new rational credence for me to have in H, after collecting exactly evidence E, is my old $p(H|E)$.

For example, in the course of developing a "No True Defeaters"-style solution to the Gettier Problem, Klein (1971, 1976) develop an account of defeat according to which a (true) proposition *D* defeats[4] *S*'s justification to believe *p* just in case, "if S were to add d to whatever justified p for S, p would no longer be justified for S" (1976, 802). Similarly, in *Contemporary Theories of Knowledge* (Pollock and Cruz 1999, 37), Pollock and Cruz give the following characterization of defeaters[5]:

> If *E* is a reason for *S* to believe *H*, *D* is a defeater for this reason if and only if *D* is logically consistent with *E*, and *E & D* is not a reason for *S* to believe *H*.

Chisholm provides an account that is similar in spirit to both Klein's and Pollock and Cruz's, though Chisholm sets things up in terms of defeat of some evidence's "tendency to make a hypothesis evident":

> *D* defeats *E*'s tendency to make *H* evident $=_{df}$ *E* tends to make *H* evident; and *D& E* does not tend to make *H* evident. (Chisholm 1989, 53)

Since these accounts focus on binary belief, they do not capture cases where the agent's defeated credence in *H* is lower than her evidential credence in *H*, and yet where either both or neither of those credences is above the threshold for belief. Of course, this is not an *objection* to any of these accounts; it is a perfectly legitimate epistemological project to give an account of the defeat of binary beliefs, and I have no particular claim on the term "defeat" to characterize the more general phenomenon that I'm interested in. But it is also a legitimate epistemological project to characterize this more general phenomenon, and an advantage of DEFEATER IFF CREDENCE-LOWERING is that it is able to do so.

Even if DEFEATER IFF CREDENCE-LOWERING is correct, we still have the problem of distinguishing opposing defeaters from undercutting defeaters. One natural idea is that, since undercutting defeaters merely (perhaps only partially) undermine the evidential effect that *E* has on *H*, the net effect of acquiring some undercutting defeater should be to push the agent's credence in *H* back toward the value it had before the agent acquired *E*, so that her defeated credence in *H* is lower than her evidential credence in *H* but no lower than her prior credence in *H*. By contrast, since an opposing defeater actually gives us independent evidence against *H*, it has the capacity to push the agent's credence in *H* below the value it had before the agent acquired *E*, leading to a defeated credence that is lower than the prior credence. Thus:

> UNDERCUTTING IFF DEFEATED CREDENCE IS NO LOWER THAN PRIOR CREDENCE: *D* is an undercutting defeater for the evidence that *E* provides for *H* just in case $p(H) \leq p(H| E \,\&\, D) < p(H|E)$.

> OPPOSING IFF DEFEATED CREDENCE IS LOWER THAN PRIOR CREDENCE: *D* is an opposing defeater for the evidence that *E* provides for *H* just in case $p(H| E \,\&\, D) < p(H) < p(H|E)$.

[4] Klein uses the term "disqualifying proposition" in Klein (1971), but he uses the term "defeater" in Klein (1976).

[5] Here and throughout, I have changed the notation of the theories I'm discussing for the sake of consistency.

Notice that each of Undercutting IFF Defeated Credence Is No Lower Than Prior Credence's condition and Opposing IFF Defeated Credence is Lower Than Prior Credence's condition individually entail Defeater IFF Credence-Lowering's condition; since both undercutting defeaters and opposing defeaters are defeaters, this is desirable. Notice too that Defeater IFF Credence-Lowering's condition entails the disjunction of Undercutting IFF Defeated Credence Is No Lower Than Prior Credence's condition and Opposing IFF Defeated Credence is Lower Than Prior Credence's condition; if you think that all defeaters are either undercutting defeaters or opposing defeaters, then this is desirable as well.

However, just because some opposing defeaters *can* push the agent's credence below what it was before acquiring E, it clearly doesn't follow that *all* opposing defeaters have that effect. The opposing defeater in the rain example above—my going outside and seeing for myself that it's sunny out—is a particularly strong opposing defeater; barring skeptical scenarios and elaborate tricks, the experience I have when I walk outside and seem to see sunny and cloudless skies makes me virtually certain that it's not raining out, regardless of whose testimony about the weather I've previously heard. Thus, assuming that I found it at least moderately credible that it was raining out before hearing Julio's testimony, the combined effect of Julio's testimony and the opposing defeater of that testimony (i.e., my sunny experience) will be to make me less confident that it's raining out than I originally was, satisfying Opposing IFF Defeated Credence is Lower Than Prior Credence's condition. But there surely are weaker opposing defeaters possible. For example, suppose that instead of going outside and observing the weather for myself, I had instead come across Jill, someone in whom I have the same confidence that I do in Julio. If Jill were to tell me that it's not raining out, it's fairly plausible that this would serve as a (weaker) opposing defeater of Julio's testimony. But since I have equal confidence in Julio and Jill, it's plausible that my defeated credence in H would (approximately) equal my prior credence; after all, there's a natural sense in which these two pieces of testimonial evidence "cancel each other out." Similarly, if I had slightly less confidence in Jill's reliability than I do in Julio's, her testimony that it's not raining would still serve to lower my credence somewhat that it's raining, though not all the way down to my prior credence. These are both counterexamples to the necessity of Opposing IFF Defeated Credence is Lower Than Prior Credence, and also to the sufficiency of Undercutting IFF Defeated Credence Is No Lower Than Prior Credence.

So far, nothing that I have said rules out the sufficiency of Opposing IFF Defeated Credence is Lower Than Prior Credence. Even if we grant that not every opposing defeater generates a defeated credence in H lower than the subject's prior credence in H, still it might be true that only opposing defeaters can do so. And indeed, it's at least somewhat plausible that $p(H| E \& D) < p(H) < p(H|E)$ does specify a sufficient condition for opposing defeat. After all, if D merely undermines the effect of E, how could D leave us with a defeated credence in H that is lower than our prior credence in H? It seems that in order to end up with a defeated credence lower than our prior credence, we'd need some independent reason to disbelieve H; a reason merely to doubt the evidential force of E (i.e., an undercutting defeater) seems like it couldn't be such an independent reason to disbelieve H. Thus, only opposing defeaters look to be capable of generating a defeated credence which is lower than the prior credence. So we can fix the problem presented by the counterexamples to the necessity of Opposing IFF Defeated Credence is Lower Than Prior Credence by simply dropping the "only if":

> Opposing If Defeated Credence is Lower Than Prior Credence: D is an opposing defeater for the evidence that E provides for H if $p(H| E \& D) < p(H) < p(H|E)$.

And we can fix the problem presented by the counterexamples to the sufficiency of Undercutting IFF Defeated Credence Is No Lower Than Prior Credence by dropping the "if":

> Undercutting Only If Defeated Credence Is No Lower Than Prior Credence: D is an undercutting defeater for the evidence that E provides for H only if $p(\mathrm{H}) \leq p(H| E \& D) < p(H|E)$.

However, we'd like to be able to say more. We'd like well-motivated necessary *and* sufficient conditions both for undercutting and for opposing defeat. And we'd like a way to distinguish between those cases where $p(H) < p(H| E \& D) < p(H|E)$ and where D is an undercutting defeater from those cases where $p(H) < p(H| E \& D) < p(H|E)$ and where D is an opposing defeater. Neither Opposing If Defeated Credence is Lower Than Prior Credence nor Undercutting Only If Defeated Credence Is No Lower Than Prior Credence gives us any guidance here.

14.3 A Different Approach

In order to pursue a different tack, let's look at how Pollock and Cruz characterize opposing defeaters:

> If E is a defeasible reason for S to believe H, D is a[n *opposing*] defeater for this reason if and only if D is a defeater (for E as a reason for S to believe H) and D is a reason for S to believe ~H. (Pollock and Cruz 1999, 37)

The idea behind Pollock and Cruz's account of opposing defeaters can't be that opposing defeaters are defeaters that motivate us to lower our credence in H (i.e., to have a defeated credence that is lower than our evidential credence); as already discussed, undercutting defeaters do that too. Nor can it be that opposing defeaters leave us with a low credence in H; if your prior credence in H was low, then a fairly strong undercutting defeater can do that too. Rather, a more plausible reading of their account is that opposing defeaters are reasons to lower our credence in H *even when we ignore E.*

In other words, a natural thought here (regardless of whether it is Pollock and Cruz's thought or not) is that the core difference between an opposing defeater and an undercutting defeater is that, since an undercutting defeater merely undermines (perhaps only partially) the evidential effect of E on H, an undercutting defeater would have no effect on H once E is removed from consideration. By contrast, an opposing defeater provides some general reason to think that H is false; thus, its effect on H isn't mediated by E. In other words, an undercutting defeater reduces our credence in H only on the assumption of E, while an opposing defeater reduces our credence in H regardless of whether we assume E or not. To distinguish the two types of defeaters, then, we need to take a look at the effect of the defeater on H when we haven't already taken E into account. This motivates the following:

> Undercutting IFF Defeater Has No Independent Effect: D is an undercutting defeater for the evidence that the E provides for H just in case $[p(H| E \& D) < p(H|E)]$ & $[p(H|D) = p(H)]$.

> Opposing IFF Defeater Has Independent Effect: D is an opposing defeater for the evidence that E provides for H if and only if $[p(H| E \& D) < p(H|E)]$ & $[p(H|D) < p(H)]$.

Obviously, each of Undercutting IFF Defeater Has No Independent Effect's condition and Opposing IFF Defeater Has Independent Effect's condition individually entails Defeater IFF Credence-Lowering's condition. Moreover, notice that Undercutting IFF Defeater Has No Independent Effect's condition and Opposing IFF Defeater Has Independent Effect's condition are mutually exclusive; again, this is desirable if we think that undercutting defeat and opposing defeat are mutually exclusive. I'll say more about this in Sect. 14.4.

It would be noteworthy if the condition specified in Opposing IF Defeated Credence is Lower Than Prior Credence entailed the condition specified in Opposing IFF Defeater Has Independent Effect, given our background assumptions.[6] As already argued, it's plausible that Opposing IF Defeated Credence is Lower Than Prior Credence specifies a sufficient condition for opposing defeat. And if Opposing IFF Defeater Has Independent Effect is true, and the condition specified in Opposing IF Defeated Credence is Lower Than Prior Credence entails the condition specified in Opposing IFF Defeater Has Independent Effect, then it follows that Opposing IF Defeated Credence is Lower Than Prior Credence is true too. This would be a prima facie mark in favor of Opposing IFF Defeater Has Independent Effect.

The condition specified in Opposing IF Defeated Credence is Lower Than Prior Credence might *look* to entail the condition specified in Opposing IFF Defeater Has Independent Effect. Opposing IF Defeated Credence is Lower Than Prior Credence's condition entails Defeater IFF Credence-Lowering's condition, which is the first conjunct of Opposing IFF Defeater Has Independent Effect's condition. After all, Opposing IF Defeated Credence is Lower Than Prior Credence's condition says that $p(H| E \& D) < p(H)$, and we're assuming throughout this discussion that $p(\mathrm{H}) < p(H|E)$. By transitivity, $p(H| E \& D) < p(H|E)$, which is the first conjunct of Opposing IFF Defeater Has Independent Effect's condition. What about the second conjunct?

It might seem as though Opposing IF Defeated Credence is Lower Than Prior Credence's condition does entail the second conjunct of Opposing IFF Defeater Has Independent Effect's condition. After all, Opposing IF Defeated Credence is

[6] Namely, that $p(H|E) > p(H)$.

Lower Than Prior Credence's condition says that $p(H| E \& D) < p(H)$—i.e., that $E \& D$ lowers the agent's credence in H to below her prior credence in H. But we're assuming that $p(H|E) > p(H)$—i.e., that E alone raises the agent's credence in H to above her prior credence in H. So if E alone raises the agent's credence in H, and the conjunction $E \& D$ lowers the agent's credence in H, isn't the only explanation for this that D alone must lower the agent's credence in H (and by more than E raises it)? After all, if D raised (or was neutral to) the agent's credence in H, how could the combined effect of E and D be to lower the agent's credence in H, given that E alone raises it?

The above argument, however, is invalid. The somewhat plausible-sounding principle

> Conjunctions of Confirmers are Confirmers (CCC): If A is evidence for H, and B is evidence for H, then $A \& B$ is evidence for H.

is false. Here's a counterexample: Suppose that someone is applying for a job in your department, and that you don't know which graduate program she's coming from. You know that Graduate Program X produces above-average percentages of Metaphysics students (compared to other graduate programs), and also above-average percentages of Logic students. However, suppose that you also know that Program X never produces students who do both Metaphysics and Logic; students are forced to choose between these areas in their first year, and aren't permitted to work in both. However, this policy is unique to X; though this happens somewhat infrequently, other programs do produce students who work in both Metaphysics and Logic. Suppose first that you find out only that the applicant does Metaphysics (it's left open when she works in other areas too). Since you know that X produces an above-average number of Metaphysics students, this is some evidence that the applicant comes from X. If, instead of finding out that she does Metaphysics, you had instead found out that she does Logic, that too would have been some evidence that she comes from X. But if you learn both that she does Metaphysics and that she does Logic, this is clearly conclusive evidence that she is not from X, since X doesn't produce any students who do both Metaphysics and Logic. Thus, the fact that the applicant does Metaphysics is a confirmer that she's from X, and the fact that she does Logic is a confirmer, but the joint effect of these confirmers is to disconfirm. Hence CCC is false.

Suppose that, in this case, you learn first that the applicant does Metaphysics. As argued above, this is some evidence that she's from X. When you learn that she also does Logic, you become certain that she's not from X. It's quite natural, then, to characterize the information that she does Logic as a defeater of your evidence that she's from X. But there's something a bit strange going on here. If you hadn't already learned that she does Metaphysics, then the information that she does Logic would have been evidence *for* the hypothesis that she comes from X. Thus, whether the information that she does Logic is a confirmer or a defeater of the hypothesis that she's from X seems to turn on the order in which you learn her specialties.

I don't know how to answer the question whether, in the case above, the information that the applicant does Logic is an undercutting defeater or an opposing defeater. I'm hesitant to classify it as an undercutting defeater, since (given that we also know that she does Metaphysics) the defeated credence is lower than the prior credence. And I'm hesitant to classify it as an opposing defeater, since it actually confirms H if we assume that we don't know that she does Metaphysics. I don't think that our pre-theoretic notions or intuitions are clear enough to deliver an unambiguous verdict here. I propose that we classify this case as involving a new type of defeater, which I call a bidirectional defeater, characterized as follows:

> Bidirectional IFF E Flips D From a Confirmer to a Disconfirmer: D is a bidirectional defeater for the evidence that E provides for H if and only if $[p(H|\ E\ \&\ D) < p(H|E)]\ \&\ [p(H|D) > p(H)]$.

So where are we now? We have the following tripartite distinction:

> Undercutting IFF Defeater Has No Independent Effect: D is an undercutting defeater for the evidence that E provides for H just in case $[p(H|\ E\ \&\ D) < p(H|E)]\ \&\ [p(H|D) = p(H)]$.

> Opposing IFF Defeater Has Independent Effect: D is an opposing defeater for the evidence that E provides for H if and only if $[p(H|\ E\ \&\ D) < p(H|E)]\ \&\ [p(H|D) < p(H)]$.

> Bidirectional IFF E Flips D From a Confirmer to a Disconfirmer: D is a bidirectional defeater for the evidence that E provides for H if and only if $[p(H|\ E\ \&\ D) < p(H|E)]\ \&\ [p(H|D) > p(H)]$.

We retain

> Defeater IFF Credence-Lowering: D is a defeater for the evidence that E provides for H just in case $p(H|\ E\ \&\ D) < p(H|\mathrm{E})$.

as a general characterization of defeaters, since the three conditions above individually entail it (and it clearly entails the disjunction of the three conditions). Now let's see whether this gets the cases that we started with right.

Take the case where Julio's testimony that it's raining out is undercut by the information that Julio is an unreliable testifier about the weather. Does this case satisfy Undercutting IFF Defeater Has No Independent Effect's condition? Clearly, $p(H|\ E\ \&\ D) < p(H|E)$ here. When I first heard Julio's testimony, I became more confident that it was raining out. When I learned that he was unreliable, I became less confident that it was raining out. Notice that this inequality holds regardless of whether the defeater is that Julio is highly unreliable or that he is only somewhat unreliable. Either way, it's plausible that the defeater induces a defeated credence lower than my evidential credence. What about the second conjunct? Before hearing Julio's testimony, I had my prior credence that it's raining out, $p(H)$. How would this credence be affected by learning only that Julio is unreliable? Intuitively, not at all. After all, I haven't heard Julio's testimony yet, so I don't even know what his testimony is, and I certainly haven't taken it into account yet. I might not have even heard of Julio, so the information that he tends to get weather reports wrong shouldn't seem particularly relevant to me. Thus, my $p(H) = p(H|D)$, so both conjuncts of Undercutting IFF Defeater Has No Independent Effect's condition are satisfied.

Now take our paradigm cases of an opposing defeater, where Julio's testimony that it's raining is defeated either by my own sunny experience outside or by Jill's testimony that it's not raining. Is Opposing IFF Defeater Has Independent Effect's condition satisfied? Again, it's fairly clear that $p(H| E \& D) < p(H|E)$; both of these defeaters give me reason to lower my credence in H on the assumption that I've already taken E into account. Moreover, it's fairly clear that the second conjunct of Opposing IFF Defeater Has Independent Effect's condition is also satisfied. Since each of these defeaters intuitively constitutes direct evidence against H, it should follow that acquiring the defeaters would force me to lower my credence in H, even if E hasn't already been taken into account. Even if I haven't heard Julio's testimony or don't know who Julio is, still a sunny experience or Jill's testimony that it's not raining constitute evidence that it's not raining. Accordingly, $p(H|D) < p(H)$, so Opposing IFF Defeater Has Independent Effect's condition is satisfied.

14.4 Hybrids

I think that Defeater IFF Credence-Lowering, Undercutting IFF Defeater Has No Independent Effect, Opposing IFF Defeater Has Independent Effect, and Bidirectional IFF E Flips D From a Confirmer to a Disconfirmer do a fairly good job characterizing ordinary types of epistemic defeat. However, these principles do less well when we consider *hybrid* cases of defeat.

Consider the following case. I have a visual experience as of Ada writing a large check to a worthy charity, and on that basis my credence increases that she is a morally upstanding person.[7] Then, I hear testimony from a reliable friend that Ada surreptitiously put a visual hallucinogen in my coffee this morning.[8] Intuitively, two reactions seem appropriate here. First, I ought to (at least partially) discount the confirming effect that I took my visual experience as of Ada's donation to have on the proposition that Ada is morally upstanding. Since I've acquired some evidence (in the form of the auditory testimony about Ada's behavior this morning) that the visual experience I had as of Ada's donation was a hallucination, I ought to regard that visual experience as a significantly less reliable indicator of Ada's moral upstandingness than I previously thought. Second, it seems, I ought to "directly" decrease my credence that Ada is morally upstanding. After all, surreptitiously putting drugs in people's coffee is a morally bad thing to do; since I have some

[7] Some positions in metaethics—notably, some versions of non-cognitivism—entail that my attitudes about Ada's moral upstandingness are non-cognitive in nature, and hence do not involve my being related to a proposition about Ada's moral upstandingness. For the purposes of simplicity, I simply ignore these positions here; I assume in the text that being confident that Ada is morally upstanding is simply a matter of having a high credence in the proposition that Ada is morally upstanding. However, nothing essential turns on this choice, and the example could be modified (at the cost of simplicity) to avoid this complication.

[8] For the purposes of this example, suppose that the hallucinogen at issue causes visual, but never auditory, hallucinations; thus, there is no reason for concern about whether your experience as of your friend speaking to you is veridical.

reliable evidence that Ada performed this morally bad act, I have "direct" evidence against the proposition that Ada is morally upstanding.

The trouble here is that Undercutting IFF Defeater Has No Independent Effect's condition and Opposing IFF Defeater Has Independent Effect's condition are incompatible, since Undercutting IFF Defeater Has No Independent Effect's condition entails that $p(H|D) = p(H)$ and Opposing IFF Defeater Has Independent Effect's condition entails that $p(H|D) < p(H)$. Thus, if these principles are correct, then hybrid cases of undercutting and opposing defeat are impossible.

In the case under consideration, it's plausible that $p(H|D) < p(H)$; even if I haven't had the visual experience as of Ada writing the check, my friend's testimony that Ada drugged my coffee is some evidence against the proposition that she is morally upstanding. So, Opposing IFF Defeater Has Independent Effect counts this as a case of opposing defeat. But that's only part of the story; intuitively, this case is also a case of undercutting defeat, and Undercutting IFF Defeater Has No Independent Effect's condition isn't satisfied (indeed, necessarily, it can't be satisfied if Opposing IFF Defeater Has Independent Effect's condition is). If this is right, then Undercutting IFF Defeater Has No Independent Effect's condition can't be a necessary for undercutting defeat (though the hybrid cases don't put any pressure on the claim that it is a sufficient condition for undercutting defeat).

I think that this sort of hybrid case motivates a reconsideration of our strategy for characterizing undercutting defeat. Though I think that it is false, Pollock and Cruz's account is again instructive here. Here is how they characterize undercutting defeaters:

> If believing E is a defeasible reason for S to believe H, D is an *undercutting* defeater for this reason if and only if D is a defeater (for believing E as a reason for S to believe H) and D is a reason for S to doubt or deny that E would not be true unless H were true. (Pollock and Cruz 1999, 37)

This characterization is a bit odd. First, there are notorious problems with counterfactual analyses, and there are the usual wrinkles related to overdetermination, pre-emption, etc. in this case. But more importantly, suppose again that Julio tells me that it's raining out (E), and that I believe on that basis that it is raining out (H). According to Pollock and Cruz, D is an undercutting defeater of Julio's testimony (for me) if only if (D is a defeater and) D gives me a reason to doubt or deny that: *Julio wouldn't say it's raining unless it were raining*. In some cases, this account works just fine: a paradigm undercutting defeater like information that there is no correlation whatsoever between Julio's weather reports and the actual weather would indeed give me a reason to deny that *Julio wouldn't say it's raining unless it were raining.*

However, suppose that I knew in advance that Julio is very unlikely to say that it's raining when it's not in fact raining. Now, suppose I learn that Julio is also very unlikely to say that it's raining when it *is* raining; in fact, we can even suppose that he's equally as unlikely to say that it's raining regardless of whether it's raining or not. This information should clearly be a defeater for Julio's testimony about the weather, since it entails (given our background knowledge) that Julio is just as likely to say that it's raining when it is raining as when it's not. But this information does *not* give us any reason to doubt or deny that *Julio wouldn't say it's raining unless it were raining*; we knew all along that Julio is very unlikely to say that it's raining when it's not raining, and the information in question doesn't change our attitude about that.

How should we fix things up? We need to weaken Undercutting IFF Defeater Has No Independent Effect's condition, since hybrid cases show that it's not necessary for undercutting defeat. And we need the weakened condition to be compatible with Opposing IFF Defeater Has Independent Effect (or whatever alternative condition we formulate to characterize opposing defeat), so that we can get hybrid cases of undercutting and opposing defeat. Pollock and Cruz's characterization of undercutting defeat seemed to get at the idea that undercutting defeaters are such that, when they are assumed to be true, the evidential connection between *E* and *H* is interfered with; however, as argued above, their counterfactual formulation of that idea is flawed.

Intuitively, in the hallucinogen case above, the reason that my reliable friend's testimony serves as an undercutting defeater is that it's some evidence that my visual experience as of Ada donating money was a hallucination. Thus, once I hear the testimony that Ada drugged my coffee, my visual experience as of her donating money no longer confirms the hypothesis that she is morally upstanding as much as that visual experience did before the testimony gave me reason to doubt the veridicality of my visual experiences. In other words, the degree of confirmation that *E* confers on *H* is lower once we assume *D*.

Let $dc(E, H, K)$ be a real-valued function of three variables *E*, *H*, and *K*, which quantifies the degree of confirmation that *E* confers on *H* relative to background information *K*. The above discussion motivates:

> Undercutting IFF Degree-of-Confirmation Lowering: *D* is an undercutting defeater for the evidence that *E* provides for *H* (relative to background information *K*) just in case $dc(E, H, K) > dc(E, H, K \,\&\, D)$.

In other words, an undercutting defeater is such that the amount of confirmation that the evidence confers on the hypothesis is lower when you assume the truth of the undercutting defeater than when you don't.

It's controversial how best to measure degree of confirmation.[9] Here are a few candidates:

$$dc1(E, H, K) = p(H|\, E \,\&\, K) - p(H|K)$$

$$dc2(E, H, K) = \log\left(\frac{p(H|\, E\&K)}{p(H|K)}\right)$$ [10]

[9] See Fitelson (1999) and Eells and Fitelson (2000) for good surveys of various candidates.

[10] One purpose of taking the log of these quantities is so that the measure counts as a so-called "relevance measure," where the measure is positive if *E* confirms *H*, negative if *E* disconfirms *H*, and 0 if *E* is neutral to *H*. Another purpose is to ensure scale-invariance. For our current purposes, the log can be ignored. Since log is a monotone increasing function, it will follow from the fact that

$dc3(E, H, K) = \log\left(\frac{p(H|E\&K\Big)/p(\sim H|E\&K)}{p(H|K\Big)/p(\sim H|K)}\right) = \log\left(\frac{p(E|H\&K\Big)}{p(E|\sim H\&K\Big)}\right)$ (The former is the log of the "Bayes Factor" and the latter is the log of the "likelihood ratio.")

For reasons that are beyond the scope of this chapter to address, I like $dc3(E, H, K)$ as a measure of degree of confirmation.[11] But for my purposes here, I wish to remain agnostic about which measure of confirmation is best. Whichever account is right, we can plug that account into Undercutting IFF Degree-of-Confirmation Lowering to yield a precisified account of undercutting defeat.[12] For example, if you also like $dc3(E, H, K)$, then Undercutting IFF Degree-of-Confirmation Lowering becomes:

> Undercutting IFF Degree-of-Confirmation Lowering*: D is an undercutting defeater for the evidence that E provides for H (relative to background information K) just in case
> $$log\left(\frac{p(H|E\&K\Big)/p(\sim H|E\&K)}{p(H|K)/p(\sim H|K)}\right) > log\left(\frac{p(H|E\&D\&K\Big)/p(\sim H|E\&D\&K)}{p(H|D\&K\Big)/p(\sim H|D\&K)}\right).$$

Two points are worth mentioning here. First, notice that Undercutting IFF Defeater Has No Independent Effect's condition entails Undercutting IFF Degree-of-Confirmation Lowering*'s condition, given the assumptions that I've been making. Undercutting IFF Defeater Has No Independent Effect's condition entails that $p(H|D) = p(H)$, which entails that $p(\sim H|D) = p(\sim H)$, so the denominators of the expressions on both sides of the inequality can be ignored. And we're supposing that $p(H|E\&D) < p(H|E)$, since we're supposing that D is a defeater, which entails that $p(\sim H|E\&D) > p(\sim H|E)$. It follows that Undercutting IFF Degree-of-Confirmation Lowering*'s condition is satisfied. So, Undercutting IFF Degree-of-Confirmation Lowering* entails that Undercutting IFF Defeater Has No Independent Effect's condition is a sufficient condition for undercutting defeat, which is desirable.

Moreover, this result doesn't essentially depend on my selection of $dc3$ as the measure of degree of confirmation. Presumably, any reasonable measure of confirmation is going to be a "relevance measure" (i.e., a measure that is positive if E confirms H, negative if E disconfirms H, and 0 if E is neutral to H) and one such that if $p(H|E_1) > p(H|E_2)$ then E_1 confirms H more than E_2 does. If Undercutting IFF Defeater Has No Independent Effect's condition holds, then $p(H) = p(H|D)$, and hence on any reasonable measure $dc(E, H, K) > dc(E, H, K \& D)$ will hold iff the linear distance between $p(H|E)$ and $p(H)$ (which is equal to $p(H|D)$) is greater than the linear distance between $p(H| E \& D)$ and $p(H)$. But that holds iff $p(H| E \& D) <$

$A > B$ that $\log A > \log B$ (and conversely). So if we want to compare two degrees of confirmation, all we need to do is to compare the argument of the log.

[11] For some reasons to accept $dc2(E, H, K)$, see Milne 1996 (though see also Pollard 1999). For some reasons to accept $dc3(E, H, K)$, see Eells and Fitelson (2000).

[12] Of course, if there is no "one true measure" of degree of confirmation but rather just a plurality of different measures, then my account entails that there will be many different notions of undercutting defeat—one relative to each of the confirmation measures. But I think that this is precisely the right result; if there is no one privileged way to measure evidence, then I don't think that there can be one privileged way to measure undercutting of evidence either.

$p(H|E)$, which is just our condition on defeat (and which Undercutting IFF Defeater Has No Independent Effect's condition entails anyway).

Second, and more importantly, notice that Undercutting IFF Degree-of-Confirmation Lowering*'s condition and Opposing IFF Defeater Has Independent Effect's condition are compatible with each other. A defeater D can simultaneously reduce the amount of confirmation that E confers on H *and* serve as evidence against H in the absence of E. Indeed, in the hallucinogen case, my friend's testimony that Ada drugged my coffee seems to do precisely this. Since the testimony is some evidence that I'm hallucinating, it reduces the degree of confirmation that my visual experience as of Ada writing a check confers on the hypothesis that she is morally upstanding. And since the testimony is also some evidence that Ada has committed a morally bad act, it's also "direct" evidence against the claim that she is upstanding.

Somewhat more concretely: Let E be my visual experience as of Ada donating money, let H be the hypothesis that Ada is upstanding, and let D be my trusted friend's testimony. Clearly, $p(H|E) > p(H)$ and $p(H| E \& D) < p(H|E)$. If we ignore the visual experience as of the donation, the testimony is still evidence against the hypothesis that Ada is upstanding, so $p(H|D) < p(H)$, so Opposing IFF Defeater Has Independent Effect's condition is satisfied. It's a little trickier to see that Undercutting IFF Degree-of-Confirmation Lowering*'s condition is satisfied, but it is. Let's fill in the details of the case a bit to see that. Suppose that before having the visual experience, my credence that Ada is upstanding is .5 (so $p(H) = .5$) and my credence that Ada is non-upstanding is .5. After having a visual experience as of Ada donating money, my credence that Ada is upstanding goes up to .8 (so $p(H|E) = .8$). If I had just heard my friend's testimony that Ada drugged my coffee, without having the visual experience as of Ada donating money, my credence that Ada is upstanding would have gone down to .1 (so $p(H|D) = .1$). The combined effect of acquiring both E and D isn't to push my credence in H all the way down to .1; after all, there is still some non-trivial chance that my friend is mistaken or lying about Ada drugging my coffee, in which case I (probably) really did see her donate money to a worthy charity. But, since I don't know much about Ada and trust my friend, my credence in H after acquiring both E and D should still be significantly lower than $p(H)$; let's suppose that $p(H| E \& D) = .15$. Undercutting IFF Degree-of-Confirmation Lowering* will then hold just in case $\log\left(\frac{.8/.2}{.5/.5}\right) > \log\left(\frac{.15/.85}{.1/.9}\right)$, which will hold just in case log 4 > log 1.588 (approximately). So Undercutting IFF Degree-of-Confirmation Lowering*'s condition holds. Clearly, this argument would have gone through even if I had used slightly different numbers.

14.5 A Second Taxonomy

Now that we've abandoned Undercutting IFF Defeater Has No Independent Effect in favor of Undercutting IFF Degree-of-Confirmation Lowering, should we modify Opposing IFF Defeater Has Independent Effect, Bidirectional IFF E Flips D From a Confirmer to a Disconfirmer, or Defeater IFF Credence-Lowering as well? I see no reason yet to modify Opposing IFF Defeater Has Independent Effect; as argued above, it is consistent with Undercutting IFF Degree-of-Confirmation Lowering's condition, which is desired, and it seems to capture our intuitions about opposing defeat quite naturally.

But once we accept Undercutting IFF Degree-of-Confirmation Lowering, that begins to cast some doubt on Defeater IFF Credence-Lowering's condition. As I'll argue next, D can lower the extent to which E confirms H, and yet can fail to lower a rational agent's evidential credence (so $p(H| E \& D) \geq p(H|E)$). Is this a problem for Undercutting IFF Degree-of-Confirmation Lowering, or rather a problem for Defeater IFF Credence-Lowering? I think it's the latter. Just as there can be hybrid case of undercutting defeat and opposition defeat, I think that there can also be hybrid cases of undercutting defeat and *evidential support*; in some such cases, the evidentially supporting effect of D can be strong enough to neutralize or outweigh D's undercutting effect on E, in which case the net effect of D can be to make it the case that $p(H| E \& D) \geq p(H|E)$, in violation of Defeater IFF Credence-Lowering's condition. But since this is still a case where D lowers the degree of confirmation that E confers on H, I still think that this sort of cases deserves to be classified as a case of (undercutting) defeat.

To see such a case, we can make a few changes to the scenario involving Ada from above. This time, let H be the hypothesis that Ada is a morally *bad* person. And this time, let E be my visual experience as of Ada doing something mildly morally bad—say, cutting someone in line at the grocery store. Clearly, given suitable background assumptions, it's plausible that $p(H|E) > p(H)$. As before, let D be a trusted friend's testimony that Ada put a visual hallucinogen in my coffee this morning. If I have reason (from any source) to believe that I'm visually hallucinating, then clearly my visual experience as of Ada cutting in line does less to confirm the hypothesis that Ada is a morally bad person than it would if I didn't have any such reason, so Undercutting IFF Degree-of-Confirmation Lowering entails (properly, in my view) that D is an undercutting defeater of the evidence that E provides for H in this case. But, in this case the effect of D's "direct" evidential support for H is much stronger than the effect of D's undercutting E's support for H; seeming to see Ada cut in line is only very mild evidence that she is a morally bad person, whereas being told by a reliable friend that Ada drugs people without their knowledge is much stronger evidence that she is a morally bad person. So, it's very plausible in this case that $p(H| E \& D) > p(H|E)$, in violation of Defeater IFF Credence-Lowering's condition.[13] If you agree with me that this is still a case of (undercutting) defeat, then we should abandon Defeater IFF Credence-Lowering.

What about Bidirectional IFF E Flips D From a Confirmer to a Disconfirmer? As I characterized bidirectional defeat above, it is a phenomenon where E and D are each evidence for H *separately*, but if we already know E, then D "flips" to being evidence against H. (The fact that the candidate does Metaphysics is evidence that

[13] Clearly, we could play with the details here so that $p(H| E \& D) = p(H|E)$, which would also violate Defeater IFF Credence-Lowering's condition.

she's from X, and the fact that she does Logic is evidence that she's from X; but if you already know that she does Metaphysics, then the fact that she does Logic is evidence against her being from X.) Bidirectional IFF E Flips D From a Confirmer to a Disconfirmer was a natural characterization of a third kind of defeat in the context of the acceptance of Defeater IFF Credence-Lowering, Undercutting IFF Defeater Has No Independent Effect, and Opposing IFF Defeater Has Independent Effect; bidirectional defeaters, on that taxonomy, were just defeaters that were neither undercutting nor opposing. But once we abandon Undercutting IFF Defeater Has No Independent Effect and Defeater IFF Credence-Lowering, it's no longer clear what to say about the cases (like the philosophy job applicant case) that motivated Bidirectional IFF E Flips D From a Confirmer to a Disconfirmer. Moreover, another clue that focusing on whether $p(H|D) > p(H)$ was the wrong strategy is that it is straightforward to construct cases quite similar to the job applicant case above where $p(H|D) = p(D)$—say, by changing the case so that X produces an *average* number of Logic students, but still no students who do both Metaphysics and Logic. Of course, the issue here is at least partly stipulative; as far as I know, "bidirectional" defeaters aren't discussed anywhere in the literature on epistemic defeat, and I certainly don't think that we have clear intuitions about when a given D is serving as a bidirectional defeater. But, Bidirectional IFF E Flips D From a Confirmer to a Disconfirmer fit into our old taxonomy in a way that made bidirectional defeat look to be a distinctive third kind of defeater, so it is natural to wonder how to fit bidirectional defeat into a taxonomy that accepts Undercutting IFF Degree-of-Confirmation Lowering and Opposing IFF Defeater Has Independent Effect.

I propose that the crucial feature of the cases that motivated Bidirectional IFF E Flips D From a Confirmer to a Disconfirmer is not (as Bidirectional IFF E Flips D From a Confirmer to a Disconfirmer claims) that D goes from confirming H to disconfirming H when we assume E, but rather that E goes from confirming H to disconfirming H when we assume D. This happens in the original job applicant case, where the fact that the candidate does Metaphysics goes from confirming that she's from X to disconfirming that she's from X when we assume that she also does Logic. And this also happens in a modified job applicant case where X produces an average (rather than an above-average) number of Logic students. So I propose to characterize bidirectional defeat as follows:

> Bidirectional IFF D Flips E From a Confirmer to a Disconfirmer: D is a bidirectional defeater for the evidence that E provides for H just in case $p(H|D) > p(H|\ E\ \&\ D)$.

Since we're assuming that $p(H|E) > p(H)$, the satisfaction of Bidirectional IFF D Flips E From a Confirmer to a Disconfirmer's condition entails that E's positive relevance to H turns into negative relevance to H once D is assumed as background information.

As it turns out, Bidirectional IFF D Flips E From a Confirmer to a Disconfirmer's condition entails Undercutting IFF Degree-of-Confirmation Lowering's condition,[14] so this taxonomy has it that all bidirectional defeaters are undercutting defeaters (but not vice versa). Again, bidirectional defeat is a partly stipulative matter, so I'm perfectly happy with this result. On the current taxonomy, undercutting defeaters lower the amount that E confirms H when they're assumed as background. Bidirectional defeaters, then, are just the special subclass of undercutting defeaters that lower the amount that E confirms H so much that that quantity becomes negative; *E disconfirms H* when a bidirectional defeater is assumed as background.

How should we replace Defeater IFF Credence-Lowering as a general characterization of defeat? The answer to this question depends on whether *all* opposing defeaters turn out to be undercutting defeaters or not. I argued in Sect. 14.4 that there are *some* cases of "hybrid" defeaters that are both undercutting and opposing, but that clearly doesn't settle the question of whether *all* opposing defeaters are undercutting defeaters. But if Undercutting IFF Degree-of-Confirmation Lowering is true, then there is some good reason to suspect that they might be. After all, if source S provides me with evidence that q is true, and I then acquire reason to believe that q is actually false, then I seem to have acquired at least *some* evidence that S is an unreliable source.[15] To take the paradigm case of an opposing defeater from Sect. 14.1: when Julio tells me that it's raining out and then I acquire reason to believe that it's not raining out (either from my own observation or from Jill's testimony), it's plausible that I have (at least typically) acquired some new reason to put less stock in Julio's testimony—in other words, that I have acquired information which serves as an undercutting defeater of Julio's testimony. If it turns out that all opposing defeaters are indeed undercutting defeaters, then (since on my taxonomy all bidirectional defeaters are undercutting defeaters too) it follows that all defeaters are undercutting, and therefore that Undercutting IFF Degree-of-Confirmation Lowering characterizes defeat in general, and not just undercutting defeat. "Pure" undercutting defeaters, on this picture, would just be (undercutting) defeaters that do not also count as opposing or bidirectional.

By contrast, it may turn out that not all opposing defeaters are undercutting defeaters. To return to our case of Ada, where E is my experience as of Ada donating money and H is the proposition that Ada is a morally good person, perhaps the information that *Ada surreptitiously put a vision-enhancing drug in my coffee this*

[14] For any relevance measure of confirmation, since $p(H|E) > p(H)$, $dc(E, H, K) > 0$. Similarly, for any relevance measure, if $p(H|D) > p(H|\ E \,\&\, D)$, then $dc(E, H, K \,\&\, D) < 0$. So, if Bidirectional IFF D Flips E From a Confirmer to a Disconfirmer's condition holds, then $dc(E, H, K) > 0 > dc(E, H, K \,\&\, D)$, so Undercutting IFF Degree-of-Confirmation Lowering's condition holds.

[15] Note that while this argument provides some reason to believe that all opposing defeaters are undercutting defeaters, it provides no reason to think that all undercutting defeaters are opposing defeaters.

morning could serve as an opposing defeater of the evidence that *E* provides for *H* without also serving as an undercutting defeater of that evidence.[16] If opposing defeat without undercutting defeat is possible, then I see no obvious or natural characteristic that Undercutting IFF Degree-of-Confirmation Lowering's condition and Opposing IFF Defeater Has Independent Effect's condition have in common. On this picture, undercutting and opposing defeat are more distinct than they may at first have seemed, rather than being two instances of the same naturally characterizable type. Undercutting and opposing defeaters may be similar in that they both "count against" a hypothesis in *some* sense, but there might not be any non-disjunctive way to formalize any shared sense in which they both count against the hypothesis. So, if there are indeed some opposing defeaters that are not undercutting defeaters, then I think that all that we can usefully say is that *D* is a defeater of the evidence that *E* provides for *H* just in case it's either an opposing or an undercutting defeater (we don't need to include bidirectional defeat here, since all bidirectional defeaters are undercutting defeaters, on my taxonomy). Thus, we would accept:

> Defeater IFF Undercutting or Opposing:: *D* is a defeater for the evidence that *E* provides for *H* (relative to background information *K*) just in case either Undercutting IFF Degree-of-Confirmation Lowering's condition or Opposing IFF Defeater Has Independent Effect's condition is met.

One might worry that, just as there are hybrids of *undercutting* defeat and evidential support, so too might there be hybrids of *opposing* defeat and evidential support, which would make problems for Defeater IFF Undercutting or Opposing and Opposing IFF Defeater Has Independent Effect. But I don't think that there really are such hybrids. Take a putative hybrid of opposing defeat and evidential support, such as the information that Julio says it's raining but Mary says it's not; if that's too conjunctive-sounding to you, imagine a light that goes on just in case both Julio says that it's raining and Mary says that it's not, and consider the evidential significance of the light going on. We might be inclined to count this is a hybrid because it is a conjunction (or, in the case of the light, it entails a conjunction) of a proposition that supports the hypothesis that it's raining (the proposition that Julio says it's raining) and a proposition that opposes that hypothesis (the proposition that Mary says it's not raining). But if that's a sufficient condition for such a hybrid, then we will end up with far too many hybrids. Consider the information that Julio said only that it's raining. That's just plain evidential support for the proposition that it's raining, and certainly not any kind of opposing *defeater* for the hypothesis of rain. But *Julio said only that it's raining* is a conjunction of *Julio said it's raining or Julio*

[16] The idea here would be that this information is an opposing defeater, since it's morally bad to put drugs in people's coffee without their consent, even if the drug has a vision-enhancing effect. However, this information wouldn't be an undercutting defeater of the evidence that my visual experience provides for the proposition that Ada is morally good; if anything, this information would tend to *strengthen* the impact of my visual experience on that hypothesis, since it reduces the probability of a visual error.

said that the barometric pressure is falling and *Julio didn't say that the barometric pressure is falling*. Since the former conjunct is evidential support for rain, and the latter conjunct (under appropriate suppositions) is opposing defeat for rain, we end up with the result that *Julio said only that it's raining* is a hybrid opposing defeater for rain, which is unacceptable.[17] I take this to strongly suggest that there can't be a non-trivial account of hybrids of opposing defeat and evidential support. Thus, I think it's a welcome result that Opposing IFF Defeater Has Independent Effect entails that opposing defeaters can't provide overall evidential support for *H*, and that Defeater IFF Undercutting or Opposing doesn't allow for pure[18] hybrids of evidential support and opposing defeat.

14.6 Redundancy and Undercutting

Unfortunately, there is a potential wrinkle with Undercutting IFF Degree-of-Confirmation Lowering. Undercutting IFF Degree-of-Confirmation Lowering says that *D* is an undercutting defeater just in case it lowers the degree to which *E* confirms *H*. The problem is not with the necessity of this condition; as far as I can tell, everything that we would intuitively count as an undercutting defeater does satisfy the condition. Rather, the problem is with the sufficiency; there is a broad class of propositions that do lower the degree to which *E* confirms *H*, and yet may not deserve to be counted as undercutting defeaters.

The class that I have in mind is the class of propositions that are somehow *redundant* of *E*. Suppose, for example, that I know that my friend Rex is a fairly reliable predictor of the weather. Each night, Rex communicates to me his predictions for the weather the following day. Since Rex wants to make sure that I receive his predictions, he *both* sends me an email and *also* leaves me a voicemail with his predictions (the predictions are always identical in the email and the voicemail).

One night, I receive Rex's email with a prediction of rain the following day. Since I take Rex to be reliable about these matters, the email is evidence that it is going to rain. Next, I listen to Rex's voicemail with (of course) the exact same prediction. Intuitively, the voicemail is *redundant* or perhaps *irrelevant* once I've already read the email, but it may be somewhat strained to call it an undercutting defeater of the evidence that the email provided for rain tomorrow. After all, the voicemail doesn't cast any doubt on the reliability of the emailed report; it's not like the voicemail said "The email that I sent you earlier was mistaken!" or anything like that.

But Undercutting IFF Degree-of-Confirmation Lowering entails that the voicemail is an undercutting defeater for the evidence that the email provides for

[17] I'm assuming here that if *p* and *q* are logically equivalent, then *p* is a hybrid opposing defeater iff *q* is a hybrid opposing defeater, but I think that's overwhelmingly plausible here.

[18] By "pure," I mean to refer to defeaters that aren't *also* hybrid undercutting defeaters.

rain. When we ignore the voicemail, the email is good evidence for rain, since I know that Rex is reliable about the weather. Suppose that the email justifies a credence in rain of .9. But when we assume the voicemail as background information, the email is no longer any evidence for rain; if I've already heard the voicemail and thus already increased my credence in rain to .9, the email is completely redundant and thus doesn't do anything to further confirm the hypothesis that it's going to rain. Let E be the email, let H be the hypothesis that it's going to rain tomorrow, and let D be the voicemail. Then, $dc(E, H, K)$ is high, since the email is good evidence for rain relative to my background information K. But $dc(E, H, K \& D)$ is 0; once I've already taken D into consideration, E does nothing further to confirm H.[19] So Undercutting IFF Degree-of-Confirmation Lowering's condition is easily satisfied; if Undercutting IFF Degree-of-Confirmation Lowering accurately characterizes undercutting defeat, this entails that the voicemail is an undercutting defeater for the evidence for rain provided by the email, which is counterintuitive.

What should we say about this sort of case? One option, of course, is to just accept that redundant evidence does undercut, since it reduces the amount of confirmation that the original evidence confers on the hypothesis. Another is to try to formally distinguish undercutting from redundancy. I've tried a number of different ways that this latter project might go; unfortunately, I do not have space here to fully develop the various possibilities or their advantages and liabilities. One possible fix is to modify Undercutting IFF Degree-of-Confirmation Lowering by adding in a term to the right-hand side of the inequality which quantifies the extent to which D confirms H *in virtue of being redundant of E*. When there is no redundancy between E and D this term should simply go to 0, whereas when there is "complete" redundancy (as in the Rex case) this term should be equal to $dc(E, H, K)$ itself, with the value of this term monotonically increasing as the extent of the redundancy increases.

One (though certainly not the only) way to implement this is to compare the values of $p(E| H \& D)$ and $p(E|H)$ to characterize the extent to which D is redundant of E (on the assumption of H). When D is not at all redundant of E (on the assumption of H), $p(E| H \& D) = p(E|H)$, whereas when D is totally redundant of E (on the assumption of H), $p(E| H \& D) = 1$. So one natural way to characterize the extent of D's redundancy of E (assuming H) is $\frac{p(E|H\&D)-p(E|H)}{1-p(E|H)}$; this term can be understood to quantify the fraction of the distance between $p(E|H)$ and 1 that D increases E's probability (on the assumption of H).

New Proposal: D is an undercutting defeater for the evidence that E provides for H (relative to background information K) just in case $dc(E, H, K) > dc(E, H, K \& D) + \frac{p(E|H\&D)-p(E|H)}{1-p(E|H)} \times dc(E, H, K)$.

[19] If dc is a relevance measure (which I've been assuming it is), then $dc(E, H, K \& D) = 0$ when $p(H|D) = p(H| E \& D)$. But it's clear here that $p(H|D) = p(H| E \& D) = .9$.

In the original rain case where Julio's testimony that it's raining and my learning that Julio is an unreliable testifier about the weather are probabilistically independent (on the assumption of rain), $p(E| H \& D) = p(E|H)$, so the third term in New Proposal goes to 0, and New Proposal makes the same predictions as Undercutting IFF Degree-of-Confirmation Lowering. In the Rex case, the voicemail saying that it's going to rain tomorrow makes the email saying that it's going to rain tomorrow certain, so $p(E| H \& D) = 1$, so the third term is equal to $dc(E, H, K)$. The second term, $dc(E, H, K \& D)$, equals 0, since the email doesn't confirm rain on the assumption of the voicemail as background information (which is the whole reason that we're now looking for an alternative to Undercutting IFF Degree-of-Confirmation Lowering). So New Proposal entails that there is undercutting defeat just in case $dc(E, H, K) > dc(E, H, K)$, which of course is false, so New Proposal entails that there is no undercutting in the Rex case, as desired.

However, New Proposal may stumble with certain cases where D has *both* a redundancy effect on *E and also* an undercutting effect. For example, if we let $D = E \& U$, where U is some genuine undercutting defeater, then New Proposal looks to wrongly entail that D doesn't undercut. Since $D = E \& U$, $p(E| H \& D) = 1$, so the third term in New Proposal becomes $dc(E, H, K)$. And because D entails E, $dc(E, H, K \& D) = 0$. So New Proposal's condition again becomes $dc(E, H, K) > dc(E, H, K)$, which is never satisfied, so New Proposal entails that there is no undercutting defeat. But, intuitively, $D = E \& U$ is an undercutting defeater here, since it entails some information (i.e., U), which is by hypothesis undercutting. I leave the project of assessing and addressing this worry to future work.

14.7 Conclusion

In this chapter, I have tried to give a framework for analyzing how notions of evidential defeat that have been deployed mostly in the context of binary belief should be generalized to partial belief contexts. Part of my goal has been taxonomic, but I don't think that the primary issue here is just the terminological one of whether this or that piece of information deserves classification as an "undercutting" or "opposing" defeater. Rather, I think that these notions of defeat play important roles in our epistemological theorizing, and I hope that I've made some progress on tracing the contours of these notions. My approach has been unabashedly Bayesian, and though I share the skepticism of some philosophers about how much of our epistemic lives can be modeled in standard Bayesian terms, I do think that Bayesianism has had some truly remarkable successes and that there is a lot to be learned from seeing how many of our epistemic concepts and norms can be modeled in Bayesian terms. To the extent that the proposals above fall short, I hope that their shortcomings will give us some useful clues about which Bayesian or non-Bayesian approaches we might fruitfully pursue with regard to evidential defeat in the future.

References

Chisholm, R. (1989). *Theory of knowledge* (3rd ed.). Englewood Cliffs: Prentice-Hall.

Christensen, D. (2007a). Epistemic self-respect. *Proceedings of the Aristotelian Society, 107*, 319–337.

Christensen, D. (2007b). Does Murphy's law apply in epistemology? Self-doubt and rational ideals. *Oxford Studies in Epistemology, 2*, 3–31.

Christensen, D. (2007c). Epistemology of disagreement: The good news. *Philosophical Review, 116*(2), 187–217.

Christensen, D. (2009). Disagreement as evidence: The epistemology of controversy. *Philosophy Compass, 4*(5), 1–12.

Christensen, D. (2010). Higher-order evidence. *Philosophy and Phenomenological Research, 81* (1), 185–215.

Eells, E., & Fitelson, B. (2000). Comments and criticism: Measuring confirmation and evidence. *Journal of Philosophy, 97*, 663–672.

Elga, Adam. (Unpublished Manuscript). *Lucky to be rational*. Available at https://www.princeton.edu/~adame/papers/bellingham-lucky.pdf

Fitelson, B. (1999). The plurality of Bayesian measures of confirmation and the problem of measure sensitivity. *Philosophy of Science, 66*(Supplement), S362–S378.

Kelly, T. (2010). Peer disagreement and higher-order evidence. In R. Feldman & T. A. Warfield (Eds.), *Disagreement* (pp. 111–174). Oxford: Oxford University Press.

Klein, P. D. (1971). A Proposed definition of propositional knowledge. *Journal of Philosophy, 68* (16), 471–482.

Klein, P. D. (1976). Knowledge, causality, and defeasibility. *Journal of Philosophy, 73*(20), 792–812.

Lasonen-Aarnio, M. (2014). Higher-order evidence and the limits of defeat. *Philosophy and Phenomenological Research, 88*(2), 314–345.

Milne, P. (1996). log[Pr(H|E∩B)/Pr(H/B)] Is the one true measure of confirmation. *Philosophy of Science, 63*, 21–26.

Pollock, J. L., & Cruz, J. (1999). *Contemporary theories of knowledge* (2nd ed.). Lanham: Rowman & Littlefield.

Pollard, S. (1999). Milne's measure of confirmation. *Analysis, 59*, 335–338.

Pryor, J. (2013). Problems for credulism. In C. Tucker (Ed.), *Seemings and justification: New essays on dogmatism and phenomenal conservatism* (pp. 89–131). Oxford: Oxford University Press.

Pryor, J. (Manuscript). *Uncertainty and undermining*. Available at http://www.jimpryor.net/research/papers/Uncertainty.pdf

Schechter, J. (2011). Rational self-doubt and the failure of closure. *Philosophical Studies, 163*(2), 429–452.