Article Summary
You reach for the bowl with ‘sugar’ written on it only to discover, from the bad taste of your coffee, that it contained salt. Mundane experiences like these show that epistemic justification does not necessarily hold stable across possible changes of information. One can be justified in believing a proposition at a certain time (that the bowl contains sugar) and cease to be justified at a later time, as one enlarges one’s epistemic perspective (as one drinks a salty coffee). When this happens, one’s justification has been defeated. An epistemic defeater, broadly speaking, is that in terms of which the defeat of justification proceeds.

The notion of an epistemic defeater is mostly associated with J. Pollock’s work on reasoning and inference. Pollock has provided the canonical definition of an epistemic defeater and proposed an influential taxonomy in terms of the way different types of defeaters induce their characteristic effects. The notion of an epistemic defeater has acquired a broader meaning in current epistemology, and several distinctions have been introduced that have contributed to a more nuanced understanding of the phenomenon of defeat as well as to complement Pollock’s original taxonomy.

The correlated notion of a propositional defeater, defined in terms of would-be defeat of justification, is appealed to by the proponents of the defeasibility theory in the parallel debate about knowledge to explain what prevents a justified true belief from constituting knowledge in Gettier cases.
1. Inference, Justification and Defeat

To understand the notion of an epistemic defeater, it is useful to start from the notion of an inference and from the distinction between deductively valid and inductively strong inferences.

Inference is a transition in thought whereby one comes to believe a proposition, the inference’s conclusion, on the basis of one’s beliefs in different propositions, called the inference’s premises. An inference is deductively valid if and only if, on condition of taking as inputs true beliefs, cannot but give true beliefs as outputs. It is inductively strong if and only if, on condition of taking as inputs true beliefs, it gives probable beliefs as outputs (see inductive inference). Inferences of either kind are especially valuable from an epistemological point of view. Suppose one justifiably believes P, and one is aware that Q can be inferred from P in either of the two senses. In this case, it is intuitive that one has (propositional) justification for believing Q; and that by actually inferring Q from P, one could acquire the (doxastically) justified belief that Q (see justification, epistemic).

The epistemic justification deriving from either kind of inference differs in terms of its stability across possible enlargement of information. If an inference from P to Q is deductively valid, in no situation in which P is true Q can be false, no matter what else is true in that situation. So, no matter what new premises one justifiably believes, if one justifiably believes P one can always add those new premises and still transition via inference to the justified belief that Q. For instance, so long as one is justified in believing that Kant was married, one can transition to the justified belief that he wasn’t a bachelor, whether or not one also justifiably believes that he was left-handed, good looking or a relentless womanizer. Contrast this with the case in which the inference from P to Q is inductively strong. In this case, P can be true also in a situation in which Q is false, and the probability of Q given P is affected by whatever else is true in that situation. For this reason, whether or not one can justifiably infer Q from the justified belief that P also depends on what other premises one justifiably believes. For example, one could not justifiably infer that
Hume wasn’t bald from the justified premise that he was always seen with flowing hair if one also justifiably believes that he used to wear a wig.

When, as with Hume’s wig, one acquires information D that renders an otherwise good inference from P to Q inappropriate, one can retain one’s justified belief that P but one’s (would-be) justification for Q is negatively affected or entirely lost. In this case, it has become customary to call D an *epistemic defeater* of the justification for Q based on P. Moreover, it has become customary to say that inductive inferences are defeasible and that deductive inferences are indefeasible. For if P entails Q, no D can apparently downgrade the justification for Q so long as one retains the justification for P.

Although this is a useful start, the reader must bear in mind that the notion of an epistemic defeater has a broader meaning in contemporary epistemology.

To begin with, there is wide consensus that epistemic defeaters can negatively impinge on *non-inferential* justification, for instance on the justification for believing that (P) it rains based on (EX) a perceptual experience as of rain, as well as on inferential justification. This is a natural extension of the above notion of an epistemic defeater. For instance, when one learns that EX is hallucinatory, one may retain EX but intuitively one loses the justification for believing P based on this experience.

Many epistemologists have also departed from the above notion of a defeater to a greater extent and have contended that deductive inferences are also defeasible.

2. Defining Defeat

We owe the canonical definition of an epistemic defeater to the work of J. Pollock (1974, 1986, 1987). Though Pollock’s interest was primarily in defeasible reasoning, he mainly concentrated on epistemic reasons—that in terms of which reasoning proceeds—and on their defeat.

For Pollock, *A mental state M is a reason for a person S to believe P if and only if it is possible for S to be justified in believing that P on the basis of M* (1987: 484). Reasons can be
beliefs and non-doxastic states (like perceptual experiences), and the transition whereby one forms a belief on the basis of either kind of state counts for Pollock as a kind of inference or reasoning. Pollock thus advocates a broader notion of reasoning and lets the status of M as a reason for P depend on the quality of the reasoning from M to P in this broader sense.

We can now move to Pollock’s notion of a reason’s defeater:

**DEF**

D is a defeater for M as a reason for P if and only if M is a reason for S to believe P, D is logically consistent with M, and M&D is not a reason for S to believe P.

While DEF tells us what it takes for a reason to be defeated, contemporary epistemologists mainly understand defeat as having justification and knowledge as its target, and distinguish between *prima facie*—namely, conditional on there not being defeaters—and *ultima facie*—namely, all things considered—justification for P from M. Given how Pollock understands the notion of a reason, talk of D being a defeater for M as a reason for P can be translated without loss as talk of D being a defeater of S’s justification for P based on M.

Some epistemologists deny that a belief can only be justified if it is supported by the subject’s reasons. These philosophers might object that DEF does not easily accommodate the defeat of this kind of justification. A. Goldman (1979) is a prominent example of this strand in the literature on defeaters. Goldman is an externalist who contends that a belief’s doxastic justification should be understood in terms of the reliability of the belief-forming process that has produced it and suggests that defeat should be understood accordingly (see Reliabilism). So, for Goldman, S’s belief that P is justified if and only if this belief is the output of a reliable belief-forming process, and this justification is defeated when S has available a (conditionally) reliable belief-forming process that wouldn’t produce the belief that P as output. BonJour (1985) suggests that, however it is analyzed, the notion of an epistemic defeater is foreign to the general spirit of externalism (see Internalism and Externalism in Epistemology). Grundmann (2009) criticizes Goldman’s account of
defeat as unmotivated *vis-à-vis* the goal of truth, while Beddor (2015) adduces counterexamples against it.

A second possible limitation of DEF is that it portrays defeat as an all-or-none phenomenon. Many epistemologists however conceive of defeat differently, as a graded phenomenon (e.g., Bergmann 2005). Those epistemologists typically distinguish between *partial defeaters*—which only induce a partial loss of justification—and *full defeaters*—whose effect is that justification is lost entirely.

A third important limitation of DEF is that, by itself, this principle tells us when D has the right *content* for defeating M as a reason for P, but it doesn’t tell us when it actually induces the effects associated to defeat, namely when it becomes a *mental state defeater* in Bergmann (2006)’s sense. Most epistemologists would reply that in order for this to happen, to begin with, there has to be a moment in time when S’s justification for P based on M is in place. These epistemologists mean defeat in the *revisionary* sense, as always involving the revision of an antecedent epistemic status, and not in the *contributory* sense, in which an epistemic status is being *prevented* from being attained (Brown 2018). In addition to that, many epistemologists would add that S must form the *ultima facie* justified belief that D. But this answer is controversial for a number of reasons.

Begin to consider that many epistemologists contend that perceptual experiences and withholdings can be epistemic defeaters. These epistemologists will want to make room for defeaters that do not meet the belief and the justification condition. Moreover, some controversy also surrounds the question whether all believed defeaters must be justified, and all justified defeaters must be believed.

M. Bergmann (2005) has suggested that D can defeat M as a reason for P even if S *unjustifiably* believes D. It is enough, for instance, that S unjustifiably believes that M is unreliable. Along similar lines, J. Pryor (2018) has argued that *hypothetical* defeaters (that he contrasts with *categorical* defeaters) are unjustified considerations that can have a negative impact on doxastic (but not on propositional) justification. On the other hand, J. Lackey (2005) distinguishes between
doxastic defeaters—namely, defeaters S actually believes—and normative defeaters—namely defeaters that one ought (epistemically) to believe and contends that the latter, whether or not they are believed, have the power to defeat. In the same spirit, S. Goldberg (2016, 2017) has contended that evidence one should have had, but that one doesn’t possess, has the power to defeat one’s justification.

In recent years, DEF has also been targeted by criticisms that impugn its correctness more in general. Chandler (2013) has adduced an example in which a reason D appears at the same time to defeat P as a reason for Q and to supply fresh support for Q. A similar situation for Chandler engenders a dilemma that can only be avoided by redefining a defeater D of R as a reason for P, roughly, as a reason to not believe that R is a reason to believe P. More recently, Dutant & Littlejohn (2020) have contended that DEF is troubled by lottery cases, and that this motivates a knowledge-first account of defeat, according to which a defeater D of M as a reason for P is evidence that one is not in a position to know P on the basis of M.

3. Propositional Defeaters

The related notion of a propositional defeater has gained prominence in the post-Gettier debate on knowledge (see Knowledge, Concept of).

Gettier (1963) has shown that a true belief can be justified yet attained at too accidentally to count as knowledge (see Gettier Problem). According to Lehrer & Paxson (1969) and Klein (1981), when this happens it is always because one’s justification is rendered unfit for knowledge by the existence of a propositional defeater.

Suppose S has the true belief that P, and this belief is justified on the basis of M. A true proposition D is a propositional defeater for this belief, and by merely being true it prevents it from constituting knowledge, provided that S is not aware of D and if S acquired the (justified) belief that D, S would have her justification for P based on M defeated.
Mental state defeaters, by defeating a belief’s justification, prevent it from constituting knowledge. The existence of a propositional defeater, on the contrary, is consistent with a belief’s retaining its justification. The capacity of a propositional defeater to impede knowledge is allegedly explained by the fact that knowledge requires justification which wouldn’t easily be lost by enlarging one’s information. This however is exactly what would happen, if S came to believe a propositional defeater.

The defeasibility theory of knowledge is at the center of an ongoing debate. In a classic counterexample, a professor sees Tom Grabit stealing a book from the library. Unbeknownst to the professor, Tom Grabit’s mother has falsely declared that it was Tom’s twin to grab the book. The proposition describing the testimony of Tom’s mother meets the definition above of a propositional defeater, yet it is intuitive that the professor knows that Tom stole the book (Lehrer & Paxson 1969). To handle this and similar counterexamples, defeasibility theorists of knowledge have tried to distinguish between misleading propositional defeaters, which would be unable to prevent knowledge, and genuine propositional defeaters, to which only that capacity should be ascribed.

4. The Mechanics of Defeat

Pollock has identified two main ways in which a reason D may act as a defeater for M as a reason for P. Since, for Pollock, defeat proceeds by making the reasoning from M to P inappropriate, these ways can be introduced by explaining how differently they accomplish this task. A rebutting or opposing defeater D for M as a reason for P compromises the transition from M to P by making available a better (or at least an equally good) transition (from D itself) to a proposition incompatible with P. This happens, for instance, when Arturo believes (P) that it is raining outside based on (M) the testimony of his sister Mafalda, and then looks out of the window thereby (D) seeing the sun shining. An undercutting or undermining defeater D of M as a reason for P compromises the inference from M to P by making available a new premise (D itself) which is a
reason to believe that M is not indicative of P’s truth. In the example above, this would happen if Arturo realized (D’) that Mafalda is clueless about the current weather.

In what precedes, rebutting and undercutting defeaters are illuminated as reasons of a special sort. As such they can themselves be defeated by defeater-defeaters—namely, by additional defeaters of the first defeaters as reasons for the relevant propositions—, in which case the original justification may be reinstated or the initial defeating effect on this justification attenuated. New information may then be added which in turn defeats the defeater-defeater, and so on.

One important difference between rebutting and undercutting defeaters is that the former always affect an old justification for believing a proposition by simultaneously supplying (some degree of) justification for believing a second proposition incompatible with the first. In contrast, the latter merely destroy or downgrade antecedent justification for believing a proposition without providing us with new justification for a second proposition to replace it.

A second important difference explored in the current literature, which Pollock’s characterization of the distinction has been accused to overlook entirely, is this. Some authors have argued that the way rebutting and undercutting defeaters induce their characteristic effects is fundamentally different, because the latter, but not the former, require and interact with higher-order presuppositions pertaining to the basis of the first-order beliefs whose justification they target (Sturgeon 2014; Melis 2014, 2016). According to Sturgeon, this is illustrated by examples in which one is unaware of the real basis M of one’s belief that P, and one’s justification for P based on M isn’t compromised by misleading evidence D that M was not indicative of P’s truth. For Sturgeon, these examples show that D can only undercut M as a reason for P if S believes that her belief that P is based on M. In response, Casullo (2018) has suggested that, even if we accept Sturgeon’s examples, we’re not bound to accept that rebutting and undercutting defeaters work in a fundamentally different way; for the claim that undercutting defeaters do, and that rebutting defeaters don’t, require and interact with higher-order presuppositions is of dubious coherence. More recently, McGrath (2020) has replied to Sturgeon by arguing that D can be an undercutting
defeater of M as an *inferential* reason for P without the need of higher-order help and has suggested that the kind of defeat instantiated in Sturgeon’s examples, while only possible in the presence of the right higher-order commitments, is of a third *sui generis* kind of defeat (see §5).

M. Bergmann (2006) has described a different way in which a reason D can be a defeater for M as a reason for P. A *reason-defeating defeater* defeats the justification for P based on M just when the justification for M is defeated. To illustrate, suppose Carolina justifiably believes that (M) it is snowing in Florence based on (M’) Luca’s testimony. In virtue of justifiably believing (M) Carolina has some justification to believe (P) that the Brunelleschi Dome will soon be covered in snow. Intuitively, when Carolina sees through a live webcam (D) that in Florence it is raining, she loses her justification for believing (M) based on (M’), thereby also losing her justification for believing (P).

A reason-defeating defeater for M as a reason for P is a *standard* defeater for M’ as a reason for M. For this reason, to group them under a different label might be criticized as potentially misleading. It also deserves emphasis that to acknowledge *reason-defeating defeater* as *sui generis* is to *extend* the notion of an epistemic defeater beyond its original boundaries. On Pollock’s characterization of defeat, only inductive doxastic reasons and non-doxastic reasons can be defeated. However, reason-defeating defeaters target inductive and deductive reasons alike. To illustrate, suppose Carolina infers (M*) that *ice crystals are falling in white flakes* on Florence from (M). In this case (D) would be a (reason-defeating) defeater for (M) as a *deductive* reason for (M*).

A different candidate for constituting a genuine alternative category of defeaters has emerged from the debate about *pragmatic encroachment*, the (not uncontroversial) thesis that the amount of epistemic support needed for knowing P varies with the practical stakes. The higher the practical cost of a mistake about P, the greater the amount of support needed to know P (Fantl & McGrath 2002, 2009; Stanley 2005). Pragmatic encroachers acknowledge that a shift in the practical stakes may induce a loss of knowledge that P even if it leaves S’s reasons for P in place.
When this happens, it might appear natural to describe the considerations raising the stakes as a *sui generis* new kind of—*interest-relative* (Weatherson 2011)—defeater.

5. Higher-Order Evidence and Defeat

Recent epistemology has started to explore the hypothesis that S’s justification for believing P can be defeated in a way that is different in kind from more ordinary cases of rebutting or undercutting defeat. While ordinary cases of defeat involve evidence that is first-order, in the sense that it is evidence with a direct bearing on the likely truth of some target proposition P (Kelly 2016), the new kind of defeat would crucially involve *higher-order* evidence, namely evidence about whether the relevant belief is rational on that evidence (Coates 2012) or about whether it is the output of a flawed psychological process (Lasonen-Aarnio 2014). Examples of this kind of defeat would involve evidence that one was under the effect of a drug, mentally too tired, or victim of a condition like hypoxia. A different and widely discussed example involves disagreements with one’s epistemic peers, which is alleged to provide some indication that one has performed badly in processing the available evidence (see Epistemology of Disagreement).

The claim that this sort of higher-order evidence can exert any defeating power is at the center of an ongoing debate. Moreover, those who accept higher-order defeaters as *sui generis* disagree about what keeps them apart from more mundane defeaters, like, in particular, undercutting defeaters.

Let’s start by considering the second issue. Some philosophers subsume higher-order defeaters and undercutting defeaters under one single broad category (e.g. Pryor 2018; Constantin & Grundmann 2020), but the majority treats them as distinct in kind. R. Feldman (2005) has suggested that while the former question the existence of an evidential connection between the evidence and the target proposition, the latter simply indicate that a connection holding in general cannot be relied upon in present circumstances. D. Christensen (2010) has advanced the different suggestion that higher-order defeaters leave the connection between the evidence and the
conclusion intact, and that they “attack” it just in the sense that they require the subject to put it aside, or to bracket it. Van Wietmarschen (2013) suggests that higher-order defeaters can impair the target belief’s doxastic justification by rescinding its basing relation to the original first-order evidence. The different hypothesis that higher-order defeaters supply direct reason for suspending the original judgment is explored in Sylvan & Lord (2020).

As with other types of defeaters, higher-order evidence can be misleading. If higher-order defeat is a real phenomenon, this provides a new reason for contending that deductive inference is defeasible. Suppose one has competently deduced P from E but misleading higher-order evidence D indicates that E doesn’t entail P. In this case, D would defeat justification for P based on the entailing reason E. Precisely for this reason, the idea that higher-order evidence can defeat has met with some resistance. Some epistemologists, to begin with, find it counterintuitive to say that one may fail to justifiably believe P when one has competently deduced P from evidence that entails it! Moreover, cases of this kind have prompted some skepticism about the defeating power of higher-order evidence tout court.

M. Lasonen-Aarnio (2014) has developed an interesting challenge. Lasonen-Aarnio understands justification in terms of conformity to correct epistemic rules. On this view, to credit higher-order evidence D with the capacity to make the belief that P unjustified requires that, when one possesses D, believing P doesn’t conform to correct epistemic rules. But when E entails P, believing P on the basis of E may appear to conform to a correct epistemic rule! This, for Lasonen-Aarnio, forces the proponent of higher-order defeat to opt for a two-tiered theory of justification, according to which justification is a matter (i) of conforming to correct epistemic rules, and (ii) of lacking evidence that one is not conforming to correct epistemic rules. The problem, for Lasonen-Aarnio, is that the two-tiered account generates epistemic dilemmas: one correct epistemic rule may require to believe P, while another correct epistemic rule may require, upon receiving the higher order defeater, suspension of judgment about P.
The debate about whether there are higher-order defeaters is intertwined with the parallel debate about whether epistemic akrasia can ever be rational, namely about whether it can ever be rational for S to believe P on the basis of E even if it is rational for S to believe that E does not adequately support P. Defenders of rational epistemic akrasia have to deny that the higher-order evidence has always the power to defeat first-order justification (e.g. Coates 2012; for a negative appraisal see Horowitz 2014 and Brown 2018). Note that to reject higher-order defeat, one does not have to accept that epistemic akrasia can be rational. One can reject higher-order defeat \textit{and} the rationality of epistemic akrasia by denying that misleading higher-order evidence can ever justify the belief that E does not support P (e.g. Titelbaum 2015).

Bibliography


Bergmann, M. (2006) \textit{Justification without Awareness}, Oxford: OUP (argues against internalism about epistemic justification; contains a chapter on epistemic defeaters where several distinctions are introduced and defended)


Brown, J. (2018) \textit{Fallibilism. Evidence and Knowledge}, Oxford: Oxford University Press (examines the debate between infallibilism and fallibilism about knowledge and devotes to chapters to examining epistemic defeaters)


Dutant, J. & Littlejohn, C. (2020) ‘Defeaters as Indicators of Ignorance’ in Simion & Brown 2020 (proposes a knowledge-first account of defeaters as indicators that one is not in a position to know).


Goldberg, S. (2017) ‘Should Have Known’ *Synthese* 194 (8): 2863–2894 (complements Goldberg 2016 and suggests that evidence one should have known can act as a normative defeater).


Horowitz, S. (2014) ‘Epistemic Akrasia’ Nous 48 (4): 718–744 (rejects the rationality of epistemic akrasia in paradigmatic cases in which misleading higher-order considerations cast doubt on first-order evidence)


Melis, G. (2014) ‘Understanding Undermining Defeat’, Philosophical Studies 170 (3): 433–442. (contrast the way in which justification can be undermined with cases in which it is rebutted).


Moretti, L. & Piazza, T. (eds) (2018) Defeaters in Current Epistemology, S.I of Synthese 195 (contains an introduction that covers important contemporary work on epistemic defeaters and collects several essays by leading philosophers on this topic)

Plantinga, A. (1993) Warrant: The Current Debate, Oxford: Oxford University Press (introduces the notion of warrant as that which, when added to true belief, yields knowledge; examines and criticizes extant conceptions of warrant; contains a discussion of Pollock’s notion of defeat)


Stanley J. (2005), *Knowledge and Practical Interests*, Oxford: Oxford University Press (defends interest-relative invariantism, the view that whether or not someone knows a proposition is partly determined by her practical interests)


