

SPECIAL ISSUE ARTICLE

Objectivist conditions for defeat and evolutionary debunking arguments

Michael Klenk 

Department of Values, Technology and Innovation, Delft University of Technology, Jaffalaan 5, Delft, 2628 BX, The Netherlands

Correspondence

Michael Klenk, Department of Values, Technology and Innovation, Delft University of Technology, Jaffalaan 5, Delft 2628 BX, The Netherlands.
Email: m.b.o.t.klenk@tudelft.nl

Funding information

Research for this article was funded by the Netherlands Organisation for Scientific Research (NWO), project number 360-20-340

Abstract

I make a case for distinguishing clearly between subjective and objective accounts of undercutting defeat and for rejecting a hybrid view that takes both subjective and objective elements to be relevant for whether or not a belief is defeated. Moderate subjectivists claim that taking a belief to be defeated is sufficient for the belief to be defeated; subjectivist idealists add that if an idealised agent takes a belief to be defeated then the belief is defeated. Subjectivist idealism evades some of the objections levelled against moderate subjectivism but can be shown to yield inconsistent results in some cases. Both subjectivisms should be rejected. We should be objectivists regarding undercutting defeat. This requirement, however, is likely to be problematic for a popular interpretation of evolutionary debunking arguments in metaethics as it can be shown that existing objectivist accounts of defeat do not support such arguments. I end by discussing the constraints of developing such an account.

KEYWORDS

defeasibility, epistemic defeat, evolutionary debunking arguments, metaethics, undercutting

1 | INTRODUCTION

Suppose you are at the bookshop and you see Tom Grabit, whom you know to be a notorious thief, come flying out of the door, rushing away with a stack of books barely hidden under his coat. As you walk off in astonishment, believing that Tom stole the books, you meet a trustworthy friend who tells you that Tom's identical twin brother

is in town. What does your friend's testimony do to your belief that Tom stole the books? Arguably, your friend's testimony reduces or even nullifies your epistemic justification for believing that Tom stole the books. After all, taking your friend's testimony seriously means that you can no longer be sure whether Tom stole the books or whether his twin brother did.

The phenomenon that accounts for the loss of epistemic justification for your belief about Tom is *undercutting defeat* (Chisholm, 1964; Pollock & Cruz, 1999). An undercutting defeater takes away or nullifies support for a given belief without showing that the belief is false. Undercutters are usually distinguished from *rebutters*, which imply that a given belief is false (Pollock, 1995, p. 85). In either case, and at the most general level, a defeater causes a belief to cease to be epistemically appropriate (Bergmann, 2006, p. 162). On the assumption that few, if any, of our beliefs are certain and are formed by non-deductive rather than deductive belief-forming methods, the concept of defeat, of both the rebutting and the undercutting kind, is essential in contemporary epistemology and fallibilism in particular (Spohn, 2012, p. 115).

However, there is a lacuna in the current understanding of undercutting defeat. The critical question is this: when does new information undercut a belief such that the believer lacks positive justification for maintaining the belief? Recent discussions relevant to this question focused on the question of whether higher-order commitments (such as beliefs) are required for undercutting defeat (Melis, 2014; Sturgeon, 2014) and the question of whether defeaters must be justified (Alexander, 2017; Casullo, 2018). These discussions imply that defeat is *subjective*: the subject needs to be *aware*, in a sense to be specified, that information d defeats her belief that p for her belief that p to be defeated.

This paper's main contribution is to raise a novel, two-pronged dilemma for subjectivist accounts of defeat. Subjectivists, as defined in this paper, rely on either a necessary or a sufficient awareness condition for defeat. If awareness is necessary for defeat, then unreflective subjects, who form no commitments about the relation between a defeater and the target belief, will be immune from defeat, which is an implausible result. If awareness is sufficient, then either the awareness condition is too weak, because it mistakenly legislates for epistemic ignorance, or, when the condition is strengthened, it can be shown to rely on incompatible intuitions and, moreover, to be inconsistent. In either case, the implication is that subjectivism regarding defeat is false.

In addition to leading to a better understanding of the nature of undercutting defeat, the paper shows how the critical question noted above is of relevance for the recent debate about evolutionary debunking arguments in metaethics, which rely heavily on an often unanalysed notion of undercutting or 'undermining' (Braddock, 2016; Leben, 2013; Lutz, 2018; Silva, 2016). Once the need for an objectivist account of defeat becomes clear, it can be shown that existing objectivist accounts of defeat are inadequate to substantiate an evolutionary challenge and why such arguments face severe limitations.

Section 2 introduces crucial terminological distinctions. Section 3 discusses the subjectivist approach to undercutting defeat and then substantiates the first horn of the dilemma, which is that strong subjectivism is false. I introduce Bergmann's (2006) representative account of moderate subjectivist idealism regarding defeat in Section 4, revealing the second horn of the dilemma: I show that moderate subjectivism evades the objections against strong subjectivism. Section 5 contains my argument against moderate subjectivism, revealing the second dilemma: moderate subjectivism is too weak or unsubstantiated. We should be objectivists regarding undercutting defeat. In Section 6, I argue that a popular interpretation of evolutionary debunking arguments in metaethics relies on giving an account of undercutting defeat, but that existing accounts fail. I close by sketching out two strategies to overcome this problem.

2 | A BRIEF TAXONOMY OF DEFEAT

We need a taxonomy for defeaters to grasp the relevance of the subjectivism/objectivism dichotomy for defeat. This section borrows some terms from biological classification to provide such a taxonomy and examines the genus of undercutting defeat central to this paper.

The class of defeaters encompasses two orders: *mental state defeaters* and *propositional defeaters*. The focus of this paper is mental state defeaters. Typically, mental state defeaters are beliefs with propositional contents, but experiences can be defeaters too (Bergmann, 2006; Pollock, 1995). Propositional defeaters are true propositions that would make some of the subject's beliefs unjustified if they were added to a subject's evidence base (e.g. Lehrer & Paxson, 1969). Some scholars distinguish mental state defeaters from propositional defeaters in terms of an internalist and an externalist account of defeat respectively (Sudduth, 2017). While internalists regarding justification reject the relevance of propositional defeaters (because they are focused on what is, in fact, part of the subject's evidence base), many externalists regarding justification recognise the need to account for the epistemic significance of mental state defeaters (cf. Beddor, 2015; Grundmann, 2009). I am focusing on mental state defeaters, so any criterion for defeat must be a function of the mental states of the subject. Let M_S be the set of all mental states of S (indexed in time). Thus, within the order of mental state defeaters, whether a belief of S is defeated must depend on M_S or some particular mental state within M_S , which I will call M_{Sx} .

Within the order of mental state defeaters, there are two families: undercutters and rebutters (the same families can be found in the order of propositional defeaters). A rebutter for S 's mental state M_{Sx} , where M_{Sx} is a reason for S to believe that p , is a reason for S to believe that $\sim p$. An undercutter for S 's belief that p is a reason for S to doubt or deny that p would be true unless M_{Sx} were true (Pollock & Cruz, 1999, pp. 196, 197). Because I focus on mental state undercutters, S must have among its mental states, M_S at least a belief (B), a belief base for B (BB), and information that is potentially undercutting (U). Whether S also needs a mental state about the relation between U , BB , and B will depend on whether subjectivism regarding defeat is true.

There are two genera within the family of mental state undercutters: undercutters that defeat doxastic justification and those that defeat propositional justification. One has doxastic justification for one's belief that p just in case one not only possesses propositional justification for believing that p but also believes that p on that basis (cf. Turri, 2010). I will focus on mental state undercutters that defeat doxastic justification. This leaves open whether mental state undercutters defeat doxastic justification by defeating propositional justification or whether they are limited to defeating doxastic justification (cf. Alexander, 2017).

Within both genera of mental state undercutters, several species can be distinguished based on the subject's awareness of the relation between the mental state that potentially does the defeating and the target belief. For example, within the set of all of S 's mental states, there is the subset of the occurrent mental states, of whose members the subject is presently conscious. Many more distinctions are possible (e.g. non-occurrent but accessible mental states), but focusing on the occurrent mental states is sufficient for this paper (cf. Klenk, 2017).

Based on this taxonomy, we can distinguish *actual* from *presumed* and *normative* defeaters. Let an *actual defeater* (of S 's belief that p) be a mental state of S that defeats S 's justification for believing that p . Let a *presumed defeater* (of S 's belief that p) be a mental state of S to the effect that mental state M_{Sx} defeats the justification of the belief that p . A presumed defeater need not be an actual defeater, that is, it need not affect the epistemic justification of S 's belief that p . For example, if S has *reason* to take p to be defeated by her mental state M_{Sx} but does not take p to be defeated, then S lacks a presumed defeater but may have an actual defeater. Conversely, if S takes her mental state M_{Sx} to defeat her belief that p , but M_{Sx} does not, in fact, defeat her belief that p , S has a presumed defeater but lacks an actual defeater. Finally, let a *normative defeater* be a mental state M_x to the effect that M_x defeats the justification of the belief that p , where it is true that S lacks M_x and S ought to have M_x (cf. Lackey, 2008, p. 45). Again, whether any normative defeater is an actual defeater depends on the nature of defeat and thus on the question of which form of subjectivism (if any) is true. For example, if awareness of the relation between new information and a target belief is required for defeat, then no normative defeater will be an actual defeater.

The taxonomy allows us to put the key question in slightly different terms and to preview an interesting result of the paper: are presumed defeaters and/or normative defeaters actual defeaters? What we will see is that presumed defeaters are not but also that there are principled reasons against counting *both* presumed defeaters and normative defeaters as actual defeaters, though there have been attempts to do so.

3 | SUBJECTIVISM ABOUT UNDERCUTTING DEFEAT

3.1 | Full-blown subjectivism regarding defeat

The relations between actual, presumed, and normative defeaters depend on the nature of defeat. Let's distinguish between two basic subjectivist accounts of defeat:

Strong subjectivism: If S 's mental state M_{Sx} is a defeater for her belief that p , then S believes that M_{Sx} is a defeater for her belief that p .

Moderate subjectivism: If S believes that her mental state M_{Sx} is a defeater with regard to her belief that p , then M_{Sx} is a defeater for her belief that p .

According to strong subjectivism, you must have a belief about or a commitment to the relationship between the alleged defeater and your existing belief for there to be defeat. There are thus no normative defeaters on this view, and failing to realise the relevance of new information for the epistemic status of your beliefs immunises you from defeat. For example, learning that most of the barns around you are fake barns cannot defeat your belief that what you just saw was a barn if you fail to think that you have lost your justification for believing you saw a barn.

According to moderate subjectivism, any presumed defeater counts as an actual defeater. Therefore, there are unjustified defeaters. For example, S 's mental state M_{Sx} need not be justified to be an actual defeater of S 's belief that p (cf. Alexander, 2017, p. 894).¹ It is underdetermined whether there are normative defeaters from this perspective.

Let *full-blown subjectivism* be the conjunction of strong and moderate subjectivism:

Full-blown subjectivism: S believes that her mental state M_{Sx} is a defeater with regard to her belief that p if and only if S believes that M_{Sx} is a defeater with regard to her belief that p .

While strong subjectivism is at odds with many popular theories in contemporary epistemology, such as process reliabilism or evidentialism, moderate subjectivism is not. However, reasonably general considerations about the nature of undercutting defeat support full-blown subjectivism.

Full-blown subjectivism regarding defeat has not been explicitly defended in the literature. However, it can be shown to follow from the *higher-order view of undercutting defeat*, or the higher-order view for short (Melis, 2014, 2016; Sturgeon, 2014). The higher-order view is the view that for S 's mental state M_{Sx} to defeat her belief that p , S must have a higher-order commitment to the effect that M_{Sx} relates to her belief that p . Higher-order commitments are mental states about one's other mental states. For example, learning about Tom's twin brother being in town will seem epistemically relevant to you only if you are committed to the view that your belief that Tom stole the books depends on your visual apprehension of Tom and that a twin brother will be indistinguishable from Tom from a distance (cf. Sturgeon, 2014, p. 117). Having such commitments about the grounds of one's beliefs seems required for undercutters to work (Melis, 2014; Sturgeon, 2014).

It can therefore be shown, first, that the higher-order view entails moderate subjectivism. Consider Sturgeon's (2014) claim that being committed to the view that your belief that p is undercut is sufficient to undercut your belief that p (Sturgeon, 2014, p. 114). He claims, for example, that a person with a 'firm presupposition that it is not the case that it wouldn't look to her as if a red thing were before her unless a red thing were before her' has, upon seeing what looks like a red thing before her, '[an] undercutting defeat[er] for her visual experience as a reason

¹The reasons for moderate subjectivism have hitherto not been considered explicitly as reasons for the possibility of the existence of unjustified defeaters; see Alexander (2017). Doing so might strengthen Alexander's case.

for her belief' (Sturgeon, 2014, p. 116). Accordingly, merely *taking* a belief to be undercut, without justification, is sufficient for the belief to be undercut, just as moderate subjectivism has it.

Second, it is a consequence of the higher-order view that subjects that fail to form relevant higher-order commitments are immune from defeat, which entails strong subjectivism. Melis (2014), for example, writes that 'unreflective agents cannot suffer undermining defeat' (Melis, 2014, p. 441). Insofar as unreflective agents cannot *take* a belief to be defeated, Melis's higher-order view leads to strong subjectivism because taking a belief to be defeated turns out to be a requirement for defeat.

Whether strong, moderate, or full-blown subjectivism is true depends partly on the higher-order view, whose truth I will assume for the sake of argument. Neither Sturgeon nor Melis explicitly defend either strong or moderate subjectivism, and their accounts could be amended to include, for example, a requirement for higher-order commitments to be justified or rational.

3.2 | Reasons against subjectivism: Casullo's objection

As it stands, however, subjectivism faces serious objections raised by Casullo (2018). This section aims at showing that Casullo's argument falls short of establishing the falsity of moderate subjectivism while retaining an important insight brought out by his argument: strong subjectivism is false.

Casullo (2018) argues that subjectivism is both 'too strong' and 'too weak' (Casullo, 2018, pp. 2901, 2902): it is too strong because it allows unjustified mental states to count as defeaters and too weak because it allows unreflective subjects to evade defeat. Let me briefly recap Casullo's two main points that are relevant to the key question of this paper. First, he notes (Casullo, 2018, pp. 2900, 2901):

[I]f [strong subjectivism] is correct, [then] if S presupposes that others are unreliable, she frees herself from having to consider any evidence from sources other than herself ... Hence, accepting the view that unjustified beliefs retain their power to defeat results in implausibly low standards for justified beliefs.

Second, subjectivism allows epistemically *irresponsible* subjects to immunise themselves from defeat (Casullo, 2018, p. 2902):

If one has a justified belief that some source S is unreliable, one can reject any other belief merely by believing that it is based on source S. Conversely, one can insulate one's beliefs from undercutting defeaters to the effect that some source S is unreliable merely by refraining from believing that any of one's beliefs are based on source S.

In both points, Casullo objects to the existence of unjustified defeaters because they mistakenly legitimise ignorance of new information. There are thus cases that look like cases of defeat, but the subjective account is incapable of explaining why they are cases of defeat.

However, Casullo's objections fall short of establishing the need for a thoroughly non-subjectivist or *objectivist account of undercutting defeat*. Let an objectivist account of undercutting defeat be the view that taking a belief to be defeated is neither sufficient nor necessary for that belief to be defeated. A moderate subjectivism can be amended, as suggested at the end of Section 3.2, to evade Casullo's objections, which can be seen in Bergmann's account of undercutting defeat (Bergmann, 2006, 2009). Casullo (2018) mainly discusses Sturgeon (2014), but in a footnote he asserts that '[Bergmann's] position is open to the objection I present against Sturgeon' too (Casullo, 2018, p. 2900).

In the next section, I take up Casullo's claim and show that it is incorrect. Casullo's objections are efficacious against strong subjectivism only. Bergmann does not defend a strong subjectivist view but a moderate subjectivist

one, which evades the objections levelled against it by Casullo. Although strong subjectivism appears to fail, moderate subjectivism might still be an option.

4 | MODERATE SUBJECTIVISM ABOUT UNDERCUTTING DEFEAT

4.1 | Bergmann's idealist account of moderate subjectivism

The main motivation for Bergmann's account of moderate subjectivism comes from his 'no presumed-defeater condition' (NPDC) for epistemic justification (2006, p. 163):

NPDC: S's belief that p, B, is justified only if S does not take B to be epistemically inappropriate.

If NPDC is true, then a belief is rendered unjustified merely by one's believing it to be epistemically inappropriate, even if one's doing so is unjustified. Bergmann defends NPDC based on the following case (2006, p. 165):

[W]hat happens to the justificational status of your belief that you have hands once you become convinced that you are a brain-in-a-vat? Is that hand belief defeated? It was justified. Does it lose its justification? I think the answer is 'yes' [because this would be] an appropriate response to the rest of your evidence.

Thus, according to Bergmann, unjustified beliefs are part of one's total evidence. Moreover, insofar as what one is justified in believing is influenced by one's total evidence, unjustified beliefs can be defeaters (2006, p. 175 emphasis in original):

It is not only *justified* [presumed] defeaters that count as actual defeaters. All [presumed] defeaters count as actual defeaters.

We can restate Bergmann's idea in the terms outlined above: inconsistencies within M_S , the subject's set of mental states, rationally compel the subject to make adjustments to M_S because the relations between mental states in M_S determine what one is justified in believing. Just as moderate subjectivism has it, any presumed defeater counts as an actual defeater on Bergmann's account.

Bergmann explains the defeating power of presumed defeaters in reference to a *properly functioning system* (cf. Bergmann, 2006, p. 134). Support for the NPDC condition depends on the intuition that a properly functioning system would reject any beliefs the system takes to be undercut by a presumed defeater and that the nature of a properly functioning system thus explains why presumed defeaters are actual defeaters (Bergmann, 2009, p. 344).

In addition, Bergmann counts both presumed and normative defeaters as actual defeaters (Bergmann, 2006, p. 174):

it is not only *presumed defeaters* which are actual defeaters; there are also times when a person doesn't have a presumed defeater but she *should* have one. In that case, she too has an actual defeater.

By counting both presumed and normative defeaters as actual defeaters, Bergmann presents a remarkable version of moderate subjectivism that can be dubbed *subjectivist idealism* because it adds an idealised element to the conditions for undercutting defeat.

Bergmann explains that the defeating power of normative defeaters is based on the nature of a properly functioning system. Whether the belief that p is justified depends on the proper function of the cognitive system that

generated the belief that p . So, depending on which belief a properly functioning human with M_S would take to be undercut (if any) in response to new information determines which belief is indeed undercut (Bergmann, 2006, p. 174).²

This completes Bergmann's account of subjectivist idealism. Setting aside differences about the nature of proper function, Plantinga (1993) shares Bergmann's commitment to moderate subjectivism enriched by an idealist element and, as we have seen, Sturgeon's and Melis's accounts can in principle be amended to be compatible with Bergmann's account too. The next section shows how such subjectivist idealist accounts can escape Casullo's objections.

4.2 | A reply to Casullo on behalf of subjectivist idealists

Casullo's objections do not extend to Bergmann's subjectivist idealism, contrary to Casullo's claim. Moderate subjectivism, therefore, does seem subject to a dilemma after all. To understand how it ultimately falters, and thus how to grasp the second horn of my dilemma, we have to understand how subjectivist idealism avoids Casullo's objections.

Bergmann avoids the charge that he is making justification less valuable by counting normative defeaters as actual defeaters. Ignorant or misinformed subjects that fail to have a presumed defeater may still have a normative defeater, depending on what their proper function demands. So, insofar as normative defeaters are actual defeaters, subjectivist idealists avoid the implausible result that brought down strong subjectivism. For example, even if you failed to make the connection between Tom's twin brother being in town and your belief that Tom stole the books, a properly functioning subject would do so, and this is why you would have an undercutter for your belief that Tom stole the books. The relevant point here is thus that Bergmann's account has the resources, in principle, to evade Casullo's first objection.

Bergmann's idealist moderate subjectivism also evades Casullo's objection that a subject can (illegitimately) immunise herself from defeat if moderate subjectivism is true. The fact that normative defeaters are actual defeaters is sufficient to block Casullo's objection. In a disagreement, for example, subject S might form the mental state, M_{S_d} , that his opponent is unreliable and thus come to have an undercutter for the view he holds based on his opponent's testimony, which is what Casullo rejects. However, M_{S_d} need not be an undercutter. A properly functioning subject might take M_{S_d} to be defeated, in which case S , depending on his proper function, has a defeater for the belief that his opponent is unreliable. In short, Casullo's objection that subjectivism regarding defeat allows epistemic ignorance does not apply to subjectivist idealism: properly functioning subjects would not form beliefs in an epistemically irresponsible way. In turn, subjectivist idealism need not allow actual subjects to be epistemically ignorant.

Therefore, Casullo's objections fail against subjectivism enriched by an *idealist* component. It might seem as if subjectivist idealism is a good way of viewing undercutting defeat. However, this is not so, as the next section shows.

5 | AGAINST MODERATE SUBJECTIVISM

5.1 | The subjectivist idealist approach is unmotivated

Subjectivist idealism would vindicate the epistemic power of *presumed defeaters* because of the primacy of subjectivity:

²See also Bergmann (2006, p. 118).

The primacy of subjectivity: the defeat status of S's mental states depends on what S takes his or her defeat status to be.

The primacy of subjectivity principle explains the intuitions invoked by Sturgeon (2014), Melis (2014), and Bergmann (2006) in response to the cases discussed above: whenever a subject takes a belief to be defeated, it can no longer justifiably maintain the belief. Paraphrasing the subjectivist's insight, one cannot reasonably maintain that one has no grounds for believing that p and p at the same time. As already noted above, this view might cause problems in the broader context of theories about justification, as neither process reliabilism nor evidentialism, for instance, can easily be combined with this view.

The defeating power of *normative defeaters* is defended in reference to an objectivist criterion (which is proper function, in Bergmann's case). There being two different kinds of defeaters is not a problem per se. There might be a purely subjective kind of defeat, applicable in some contexts, such as in a court of law. However, insofar as justification is not entirely subjective, and Bergmann does not think that justification is entirely subjective, we must give a non-subjectivist account of defeat in cases where it seems that S *should* have a defeater. Subjectivist idealism does not do this, as I will show in this section.

Consider that S will have a certain degree of epistemic justification, which might be zero, for any of her beliefs. Bergmann has to say that normative defeaters affect S's epistemic justification, since that is what allows him to escape Casullo's objection. However, Bergmann also claims that presumed defeaters affect S's epistemic justification. What is the relation between both kinds of defeaters? Bergmann maintains that if *either* S has a believed defeater for the belief B or S has a normative defeater for B then B is defeated for S.

I object that the explanation of why a presumed defeater is epistemically relevant is different from the explanation of why a normative defeater is epistemically relevant. These explanations cannot be reconciled because the primacy of subjectivity (which explains the defeating power of presumed defeaters) implies that normative defeaters are not actual defeaters, and the idealist criterion (which explains the defeating power of normative defeaters) implies that presumed defeaters are not actual defeaters.

To elaborate, consider that the primacy of subjectivity captures something intuitively appealing about the epistemic effects of holding some belief vividly. Try imagining that you are a brain-in-a-vat (BIV). Could you still rationally maintain that you have hands? As we have seen, Bergmann thinks that the answer is no (Bergmann, 2005, p. 425).

You might feel a Peircean worry that it is impossible for humans to *really* believe a sceptical scenario as radical as the BIV scenario, and thus discount the relevance of the BIV example. But more mundane scenarios lead to similar results:

Steve enthusiastically starts a new office job. All goes well in the first weeks and, because of his positive self-image, Steve believes that he'll probably make it through the three-month probationary period. However, four weeks into the job, Steve's partner of just two months ends their relationship. Steve sinks into a deep depression. No one, Steve believes, can bear his company for longer than two months. At the same time, nothing changes at work. His colleagues continue to be open and friendly; his boss seems satisfied. In light of his belief that nobody can cope with him, Steve wonders how he could ever believe that he'll make it through the probationary period. After all, who would want to hire an awful person like him? Steve, therefore, no longer believes that he'll make it through the probationary period and, to be on the safe side, he starts applying for a new job.

Steve's break-up does not tell us, or him, anything about his suitability as an office clerk and his chances of keeping the job. Suppose that a properly functioning human would have kept both issues, job performance and private relationship, apart. Still, taking Steve's perspective on the issue seriously, there is a clear sense in which Steve's world view after his break-up makes it reasonable for him to revoke the belief that he'll make it through the probationary

period. So, given Steve's world view, he should no longer believe that he'll make it through the probationary period. This intuition seems triggered by the primacy of subjectivity: Steve must feel a kind of internal pressure, an urge to disbelieve that he will make it through the probationary period.

The distinction between proper function and the internal perspective could be drawn using a distinction between internal and external rationality. According to Plantinga (2000, pp. 110–112), internal rationality has to do with what goes on in belief formation 'downstream from experience', whereas external rationality is broader in that it depends on what goes on in belief formation before experience. Bergmann (2009, p. 337) uses this idea to define a belief as 'internally rational iff it is an epistemically appropriate response to the subject's mental states' and 'externally rational iff the believer's cognitive processing mechanisms are working as they epistemically should be in producing the belief'. Steve's belief might be internally rational but not externally rational.³

However, if we prize the primacy of subjectivity in deciding about presumed defeaters, then it seems mistaken to hold that subjects without a presumed defeater, with a perfectly harmonious inner economy, also have an actual defeater. According to the subjectivist idealists' grounds for defending the relevance of presumed defeaters, the perfectly harmonious inner epistemic economy of a subject who fails to believe a normative defeater (but does not notice that one exists) should take precedence. It is ad hoc to uphold a commitment to idealism. Subjectivist idealism defers to the *actual* agent (such as Steve) when it comes to assessing the effects of held beliefs, but the component containing idealist views defers to the *properly functioning* agent (certainly not Steve) when it comes to assessing the effects of non-existing beliefs. However, we want to know *why* some mental states are defeaters, and subjectivist idealism gives differing explanations for presumed defeaters and normative defeaters.

Bergmann might reply that the explanation for the defeating power of presumed defeaters and normative defeaters is the same: in each case, there is a violation of the requirement that beliefs be produced in accord with the *right kind of proper function*. Bergmann's reply might seem promising. A presumed defeater is an actual defeater *because* any design plan that might confer epistemic justification must require the believer not to believe B when the believer takes B to be defeated. A normative defeater is an actual defeater of S's beliefs *because* it might be the proper function of S to take such information to be a defeater.

However, Bergmann's reply does not resolve the problem but merely moves it. Why believe that proper function can produce justified belief *only if* a design plan specifies proper function such that S does not hold B if S takes B to be inappropriate? The answer to this question must be that the totality of S's mental states determines the defeat status for S's beliefs. Here we look to internal rationality to find out what defeat is. But the reason why normative defeaters are actual defeaters on Bergmann's account is that this is what proper function might require, and here we end up citing *external rationality* as that which determines defeat. We still end up with fundamentally different explanations of the defeating power of presumed defeaters and normative defeaters.

5.2 | The subjectivist idealist approach is unsubstantiated

The differing explanations of undercutting defeat offered by subjectivist idealism lead to a deeper problem, which causes Bergmann's subjectivist idealism either to yield conflicting verdicts about defeat or to succumb to Casullo's objections after all.

To begin with, note that defeaters can themselves be defeated by so-called defeater-defeaters (there are also defeater-defeater-defeaters, and so on). Every case of undercutting defeat by a presumed defeater can therefore also be described as a case of undercutting defeat by way of a normative defeater. Consider a briefer version of Steve's case:

³In a later publication, Bergmann writes that 'justification is equivalent to internal rationality' (Bergmann, 2009, fn. 9). If that were so, however, then Bergmann's account would succumb to Casullo's objections, discussed in Section 3.2.

ACTUAL: Steve believes that he will make it through the probationary period. Later, Steve believes, without evidence, that nobody can cope with him for longer than two months (M_{S1}). Steve takes M_{S1} to defeat his belief that he will make it through the probationary period.

Now consider the same case in which the actual Steve is compared with a hypothetical, properly functioning Steve:

HYPOTHETICAL: Steve believes that he will make it through the probationary period. Later, Steve believes, without evidence, that nobody can cope with him for longer than two months (M_{S1}). Steve*, Steve's properly functioning counterfactual self, believes that the fact that nothing changed at work gives him no reason to believe that he will not make it through the probationary period (M_{S1}^*). M_{S1}^* defeats M_{S1} , leaving the epistemic status of Steve's belief that he will make it through the probationary period unchanged.

Bergmann has to say that in HYPOTHETICAL, we take the totality of Steve's* mental states, M_{S1}^* , to be decisive in deciding whether Steve has a defeater. Otherwise, we end up with the objections raised by Casullo. However, taking M_{S1}^* to be decisive does not fit well with Bergmann's commitment to the primacy of the subjective, which would take M_{S1} to be decisive. So, we seem committed to an objectivist view of defeat after all.

The same point can be made about belief revision. Recall that Bergmann's account implies that a subject with a normative defeater ought to form a mental state (be it belief or commitment) about the relation between the defeater and the target belief (e.g. 'my realisation D shows that I ought to give up B'). In the case of a normative defeater, subjectivist idealists thus ask the subject to *add* something to his belief base. In the case of an actual defeater, however, the situation appears analogous. The subject has formed a mental state, R, about the relation between two other mental states, but R is misleading: the new information does not imply what the subject thinks it does. Still, a properly functioning subject, Bergmann will say, would react to this set of mental states by rejecting the target belief. At the same time, however, the subject has a normative defeater too. Given M_S , the subject better rejects the misleading mental state. So, since normative defeaters are actual defeaters, Bergmann's account will lead to contradictions in some cases: the mental state D will be an actual defeater in virtue of being a presumed defeater, and it will not be an actual defeater in virtue of being defeated by a normative defeater.

Therefore, even setting aside that the explanation for why presumed defeaters and normative defeaters are actual defeaters marks only an arbitrary distinction, that distinction cannot be substantiated. It leaves subjectivist idealists unable to tell whether a belief is undercut or not and thus the view should be rejected as an account of defeat.

6 | LIMITS OF EVOLUTIONARY UNDERCUTTING OF MORALITY AND TWO WAYS FORWARD

6.1 | Evolutionary explanations of morality as undercutting defeaters for moral beliefs

The rejection of subjectivism regarding defeat has repercussions for the debate about the metaethical implications of evolutionary explanations of morality. Several scholars have taken evolutionary explanations to *undercut* objectivist moral beliefs. It is argued that some, if not all, human moral beliefs are the product of evolutionary forces: we were selected for having useful rather than true moral beliefs.⁴ Learning about the evolutionary origin

⁴The empirical details of evolutionary explanations of morality are controversial. For instance, can *all* moral beliefs be explained by evolution in the relevant sense? Since this paper focuses on the epistemic rather than the empirical details of debunking arguments, this question is not addressed here; for further discussion see Klenk (2018b, ch. 1).

of our moral beliefs undercuts their prima facie justification, or so the challenge goes (cf. Klenk, 2018a; e.g. Braddock, 2016; Leben, 2013; Lutz, 2018; Silva, 2016).⁵

On a subjectivist account, that would suffice to undercut moral beliefs that can be explained by evolution. However, in summary, because subjectivist accounts of defeat fail, existing attempts to use evolution to defeat moral objectivism fail, too. At best, we do not know whether or why evolution undercuts objectivist moral beliefs.

A concrete example of moral beliefs explained by evolution concerns beliefs about parental obligations. Simplifying some empirical details, the belief that parents have an obligation to care for their children plausibly motivated caring for children and thereby increased the chance of rearing healthy offspring (cf. Klenk, 2018a, pp. 116, 117). Hence, the story goes, we ultimately believe that parents have an obligation to care for their children *because* of our evolutionary history and *independently* of whether parents really have such an obligation. After all, we would believe that parents ought to care for their children even if the moral truths were different; and we would think differently about parental obligations had we taken a different evolutionary path (where, say, infanticide was evolutionarily advantageous, as in some bee colonies). The evolutionary genealogy of the belief about parental obligations allegedly shows a worrying disconnect between the belief and its truth-makers and thereby undercuts the belief's prima facie justification.

The evolution-as-an-undercutter characterisation, or the *evolutionary undercutting challenge* for short, relies on the controversial assumption that moral beliefs are prima facie justified. It grants that there is some way in which moral beliefs can be justified, which is not a trivial matter. Many objections to moral objectivism turn on arguments to the effect that objective morality is peculiar so it is impossible to have justified beliefs about it in the first place (cf. Sinnott-Armstrong, 2006). Therefore, the evolutionary undercutting challenge promises to be dialectically effective: it may show that moral objectivism fails even if one of its crucial assumptions are granted.

However, to make good on the promise and to pursue the evolutionary challenge against moral objectivism, it must be shown why information about the evolutionary origins of our moral beliefs undercuts moral beliefs. Here the argument of this paper connects with the debunking debate: to make the evolutionary undercutting challenge work, we need to know *when* new information undercuts a belief.

Revealing the inadequacy of a subjectivist account has prevented a straightforward answer: *taking* one's belief that *p* to be defeated is *not* sufficient for one's belief that *p* to be defeated. In light of the inadequacy of a subjectivist account of defeat, we must look for alternative objectivist explanations of defeat. The next section aims at showing that there is currently no objectivist account of undercutting defeat that substantiates the evolutionary undercutting challenge.

6.2 | No adequate account of undercutting defeat currently exists

Recall that an objectivist account of undercutting defeat must explain why *S*'s mental state M_{Sx} takes away the justification of *S*'s mental state *p*, and the explanation must not reduce to *S*'s taking there to be such a justification-reducing relation between M_{Sx} and her belief that *p*. My aim for the remainder of this paper will not be to develop such an account but to show why the previous discussion implies that a successful evolutionary undercutting challenge has serious obstacles to overcome.

Of course, the task of finding a suitable objectivist account of defeat raises a methodological question: how do we know whether or not to accept said account? Though I cannot develop a fuller answer in this paper, my argument requires only one general criterion for accepting an objectivist account of undercutting defeat: any reason to give up a belief must indicate a lack of something relevant for a well-formed belief. To illustrate, learning that you did not form your belief that *p* during a full moon is, all else being equal, not a reason to give up your belief,

⁵Moral objectivism is the view that (at least some) moral truths are metaphysically necessary as well as constitutively and causally independent of human attitudes or beliefs and that justified moral beliefs and knowledge are possible, e.g. Enoch (2011). It is a common target of evolutionary debunking arguments in metaethics.

because forming beliefs during a full moon is not required for reasonable belief. What counts as 'epistemically relevant' is up for debate, but the criterion is useful nonetheless insofar as it rules out accounts that link defeat to something that is epistemically *irrelevant*.

The next two criteria are specific to setting up an evolutionary undercutting challenge. First, evolutionary explanations of morality must count as undercutters on this account, for otherwise we do not get an evolutionary undercutting challenge. Second, the epistemic principle that motivates the account must be compatible with objectivist moral beliefs being justified in the first place. To illustrate, consider the proposal that M_{S_x} defeats p iff the content of M_{S_x} is that the belief that p is not causally connected with p . Such a view relies on a causal requirement regarding justification. If this were true, beliefs in objective moral facts, which are causally inert, could not be justified in the first place. Since we are assuming that moral beliefs are *prima facie* justified, such an account would not do, as there would not be any justified objectivist moral beliefs to undercut.

Given these three constraints on objectivist accounts of defeat, it can be shown that even the most promising recent attempts fall short of substantiating an evolutionary undercutting challenge. The problem has already arisen in Pollock's influential definition of undercutting defeat. Pollock defines undercutters as giving reason to S to 'doubt or deny' that S would be in a state M_{S_x} unless p were true (Pollock & Cruz, 1999, pp. 196, 197). He further clarifies that "' S would not be in state M_{S_x} unless p were true'" can be read more simply as " M_{S_x} does not *guarantee* p " (Pollock & Cruz, 1999, pp. 197, emphasis added). Pollock explicitly rejects interpreting the 'guaranteeing' relation between two mental states as a material or causal relation (*ibid.*), rightly so, for linking undercutting to a causal constraint on justification violates the first, general requirement: causal constraints on justification are widely rejected in epistemology (cf. Lutz, 2018). A truth-functional relation, in contrast, is congenial to sensitivity or safety requirements for knowledge and justification (and thus satisfies our first requirement), but it can be shown to violate the second requirement: arguments to the effect that moral beliefs are trivially sensitive and very likely to be safe can be interpreted in this light (cf. Clarke-Doane, 2015). These observations suggest that Pollock's account of undercutting defeat is inadequate to underwrite an evolutionary undercutting challenge. Relying on Pollock's framework to raise the aforementioned evolutionary challenge against moral objectivism is thus mistaken (e.g. Braddock, 2016; Leben, 2013; Silva, 2016).

Promising improvement, Lutz (2018) argues that the relevant sense of 'guaranteeing' should be understood as an *explanatory* connection. According to Lutz, evidence E 'guarantees' the truth of the belief that p insofar as the best explanation of E entails p . So, an undercutting defeater of your belief that p is based on evidence E is information that fully explains E *without* implying that p is true. However, Lutz's explanatory connections view fails our third requirement: if an explanatory connection were required, in the form that Lutz envisages, then objectivist moral judgements could *not* be justified in the first place. Objectivists assume, and debunkers accept for the sake of argument, that *there are* justified moral beliefs whose truth is *not* implied by the best explanation of them. If Lutz were right, then evolutionary explanations of morality would not defeat moral objectivist beliefs because these beliefs would not be justified to begin with.

Therefore, the case for an objectivist account of defeat presented in this paper implies that we do not know whether or why evolution undercuts our moral beliefs. There is no reason, then, to give them up, for any account implies either that moral beliefs are not justified to begin with (in which case we do not have an undercutting situation) or that evolutionary explanations of morality do not provide a reason to give up our moral beliefs.

6.3 | Two ways forward to raise an evolutionary debunking challenge

The need to develop a suitable objectivist account of defeat to underwrite the evolutionary undercutting challenge is an essential outcome of this paper. Existing discussions of the metaethical implications of evolutionary explanations of morality have taken for granted too much epistemological barrage, notably the applicability of Pollock's account of undercutting defeat. There are two ways to make progress with an evolutionary debunking challenge, which I will briefly sketch out in the remainder of this section.

Naturally, the first option is to develop a suitable objectivist account of defeat to raise an undercutting challenge. This paper has produced significant constraints on developing such an account. Based on the failings of existing accounts of defeat, we can infer that the reason for giving up a belief must be, first, an objective, valid epistemic desideratum that, second, stable beliefs about necessary truths do not vacuously fulfil. Since modal and probabilistic criteria for justification will therefore probably fail, as will causal and explanatory criteria, it may be fruitful to look for virtue-based criteria of justification to underwrite an evolutionary undercutting challenge (cf. Klenk, 2018b). Presupposing virtue-based criteria for epistemic justification, however, requires further defence and argumentation in the context of moral epistemology.

In light of these obstacles, the second option is to re-evaluate the starting assumptions of evolutionary challenges against morality. In attempting to *undercut* objectivist moral beliefs, we must first accept that moral beliefs are *prima facie* justified. However, the failings of existing objectivist accounts of defeat (to substantiate the evolutionary undercutting challenge) give impetus to reconsidering whether this is the best interpretation of the evolutionary challenge after all. Perhaps evolutionary explanations of morality should best be regarded as burden-shifting arguments: rather than taking moral objectivism as the default position to consider in metaethics, they give us reason to consider non-objectivist accounts of morality first. Taking this route requires answering a methodological question about legitimate assumptions in metaethics. Again, the case for objectivist conditions for defeat reveals the importance of this question for evolutionary debunking arguments in metaethics.

7 | CONCLUSION

Subjectivist idealism relating to undercutting defeat may rescue a moderate form of subjectivism from objections that affect strong subjectivism. However, adopting an idealist form of moderate subjectivism is no help. Taking on board an idealist component leads to a bifurcation in the subjectivist idealist's concept of defeat. There is no clear way to reconcile the two concepts. Subjectivism in either form is not the right view of undercutting defeat. This finding has repercussions beyond epistemology. Arguments to the effect that evolutionary explanations of morality undercut moral beliefs rely on an objectivist account of defeat, but no suitable objectivist account exists yet, because such accounts either depend on epistemically irrelevant criteria (such as a causal theory of justification) or they imply that evolution does not undercut objectivist moral beliefs. Therefore, at best, we do not know whether evolution undercuts morality, which shows where the metaethical debate about moral objectivism and the implications of evolutionary explanations of morality depend on open and underexplored epistemological questions about the nature of defeat.

ORCID

Michael Klenk  <https://orcid.org/0000-0002-1483-0799>

REFERENCES

- Alexander, D. (2017). Unjustified defeaters. *Erkenntnis*, 82, 891–912. <https://doi.org/10.1007/s10670-016-9849-z>
- Beddor, B. (2015). Process reliabilism's troubles with defeat. *The Philosophical Quarterly*, 65, 145–159. <https://doi.org/10.1093/pq/pqu075>
- Bergmann, M. (2005). Defeaters and higher level requirements. *The Philosophical Quarterly*, 55, 419–436. <https://doi.org/10.1111/j.0031-8094.2005.00408.x>
- Bergmann, M. (2006). *Justification without awareness: A defense of epistemic externalism*. Oxford, UK: Oxford University Press.
- Bergmann, M. (2009). Rational disagreement after full disclosure. *Episteme*, 6, 336–353. <https://doi.org/10.3366/E1742360009000756>

- Braddock, M. (2016). Evolutionary debunking: Can moral realists explain the reliability of our moral judgments? *Philosophical Psychology*, 29, 844–857. <https://doi.org/10.1080/09515089.2016.1163667>
- Casullo, A. (2018). Pollock and sturgeon on defeaters. *Synthese*, 195, 2897–2906. <https://doi.org/10.1007/s11229-016-1073-5>
- Chisholm, R. M. (1964). The ethics of requirement. *American Philosophical Quarterly*, 1, 147–153.
- Clarke-Doane, J. (2015). Justification and explanation in mathematics and morality. In R. Shafer-Landau (Ed.), *Oxford studies in metaethics: Volume 10* (pp. 80–103). Oxford, UK: Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780198738695.003.0004>
- Enoch, D. (2011). *Taking morality seriously: A defense of robust realism*. Oxford, UK: Oxford University Press.
- Grundmann, T. (2009). Reliabilism and the problem of defeaters. *Grazer Philosophische Studien*, 79, 65–76.
- Klenk, M. (2017). Can moral realists deflect defeat due to evolutionary explanations of morality? *Pacific Philosophical Quarterly*, 98, 227–248. <https://doi.org/10.1111/papq.12207>
- Klenk, M. (2018a). Evolution and moral disagreement. *Journal of Ethics and Social Philosophy*, 14, 112–142. <https://doi.org/10.26556/jesp.v14i2.476>
- Klenk, M. (2018b). *Survival of defeat: Evolution, moral objectivity, and undercutting* (PhD thesis). Utrecht University.
- Lackey, J. (2008). *Learning from words: Testimony as a source of knowledge*. Oxford, UK: Oxford University Press.
- Leben, D. (2013). When psychology undermines beliefs. *Philosophical Psychology*, 27, 328–350. <https://doi.org/10.1080/09515089.2012.731579>
- Lehrer, K., & Paxson, T. (1969). Knowledge: Undefeated justified true belief. *The Journal of Philosophy*, 66, 225–237. <https://doi.org/10.2307/2024435>
- Lutz, M. (2018). What makes evolution a defeater? *Erkenntnis*, 83, 1105–1126. <https://doi.org/10.1007/s10670-017-9931-1>
- Melis, G. (2014). Understanding undermining defeat. *Philosophical Studies*, 170, 433–442.
- Melis, G. (2016). Undermining defeat and propositional justification. *Argumenta*, 1, 271–280.
- Plantinga, A. (1993). *Warrant and proper function*. Oxford, UK: Oxford University Press.
- Plantinga, A. (2000). *Warranted Christian belief*. Oxford, UK: Oxford University Press.
- Pollock, J. L. (1995). *Cognitive carpentry: A blueprint for how to build a person*. Cambridge, MA: MIT Press.
- Pollock, J. L., & Cruz, J. (1999). *Contemporary theories of knowledge* (2nd ed.). *Studies in epistemology and cognitive theory*. Lanham, MD: Rowman & Littlefield Publishers.
- Silva, P. (2016). Etiological information and diminishing justification. *Inquiry*, 55, 1–25. <https://doi.org/10.1080/0020174X.2016.1238779>
- Sinnott-Armstrong, W. (2006). *Moral skepticisms*. Oxford, UK: Oxford University Press.
- Spohn, W. (2012). *The laws of belief: Ranking theory and its philosophical applications* (1st ed.). Oxford, UK: Oxford University Press.
- Sturgeon, S. (2014). Pollock on defeasible reasons. *Philosophical Studies*, 169, 105–118.
- Sudduth, M. C. (2017). Defeaters in epistemology. Retrieved from <http://www.iep.utm.edu/ep-defea/>
- Turri, J. (2010). On the relationship between propositional and doxastic justification. *Philosophy and Phenomenological Research*, 80, 312–326. <https://doi.org/10.1111/j.1933-1592.2010.00331.x>

How to cite this article: Klenk M. Objectivist conditions for defeat and evolutionary debunking arguments. *Ratio*. 2019;00:1–14. <https://doi.org/10.1111/rati.12230>