ARTICLES

The *Journal of Ethics and Social Philosophy* aspires to be the leading venue for the best new work in the fields that it covers, and it is governed by a correspondingly high editorial standard. The journal welcomes submissions of articles in any of these and related fields of research. The journal is interested in work in the history of ethics that bears directly on topics of contemporary interest, but does not consider articles of purely historical interest. It is the view of the associate editors that the journal's high standard does not preclude publishing work that is critical in nature, provided that it is constructive, well-argued, current, and of sufficiently general interest.

RATIONALITY, APPEARANCES,
AND APPARENT FACTS

Javier González de Prado Salas

Ascriptions of irrationality typically constitute a form of criticism, while ascriptions of rationality are a form of praise.[1] More specifically, charges of irrationality involve personal criticism, in which the agent is negatively evaluated for having responded in certain ways.[2] It is often thought that being criticizable is evidence that one has done something one ought not to do— something one had decisive normative reasons not to do.[3] Assuming that this is so, the fact that charges of irrationality constitute a form of criticism suggests that there is a close connection between rationality and what we have reasons to do. This provides motivation for a normative, reasons-based account of rationality.[4]

---

1   Broome, "Does Rationality Give Us Reasons?" 336; Southwood, "Vindicating the Normativity of Rationality," 12; Way, "Two Accounts of the Normativity of Rationality," 1, and "Reasons and Rationality"; Parfit, *On What Matters,* 33; Wedgwood, "Rationality as a Virtue"; Kiesewetter, *The Normativity of Rationality*, 38–42; Lord, *The Importance of Being Rational*, 3, 222.

2   Kiesewetter, *The Normativity of Rationality*, 26–27. We can think of acts of criticizing and praising as typically expressing attitudes of blaming and crediting, respectively. However, if we do this, we should not understand these attitudes as necessarily having a moral dimension. For the purposes of this paper, I focus on the evaluations associated with acts of rational criticism and praising. For discussion on the relations between blaming, evaluation, and reactive emotions, see among others Strawson, "Freedom and Resentment"; Hieronymi, "The Force and Fairness of Blame"; Zimmerman, "Moral Responsibility and Quality of Will."

3   Greenspan, "Asymmetrical Practical Reasons"; Lackey, "Norms of Assertion," 594–95; Kiesewetter, *The Normativity of Rationality*, 26–38; Madison, "On Justifications and Excuses."

4   Kiesewetter offers a detailed discussion of how a normative account of rationality is motivated by the relation between criticism and ascriptions of irrationality (*The Normativity of Rationality*, ch. 2). See also Way, "Two Accounts of the Normativity of Rationality," 1; Southwood, "Vindicating the Normativity of Rationality," 12. Kolodny, "Why Be Rational?" and Broome, "Does Rationality Give Us Reasons?" and "Is Rationality Normative?" have famously cast doubt on the idea that rationality is normative. It should be noted, however, that when Broome reaches his skeptical conclusion about the normativity of rationality, he immediately goes on to try to explain the connection between charges of irrationality and criticism ("Does Rationality Give Us Reasons?" 336, and "Is Rationality Normative?"

According to this type of account, being rational is a matter of doing what one has most normative reasons to do.[5]

Reasons-based accounts of rationality face, however, an immediate problem. It seems that an agent can behave rationally despite relying on false considerations (for example, in situations in which the agent is guided by convincing but ultimately misleading evidence). Yet, plausibly, only facts, and not false beliefs, can be objective normative reasons—this sort of view finds widespread support in the literature.[6] So, on the face of it, there are cases in which rationality does not amount to responding to factual normative reasons, but rather to merely apparent reasons. In this way, several authors have developed theories of rationality as sensitivity to those considerations that *appear* as normative reasons to the agent.[7] Given that merely apparent reasons have no normative force, these theories fail to make rationality a genuinely normative notion.[8]

Recent proposals seem to offer a way out for factualist reasons-based accounts of rationality, in the face of the challenge posed by rational actions relying on false beliefs.[9] The way out would be to argue that there are always factual, normative reasons to which the behavior of a rational agent is sensitive and that rationalize such behavior. In particular, in cases in which a rational agent seems to be guided by false beliefs, the reasons that rationalize her behavior would be

---

177–78). This shows that, also for Broome, this connection provides *prima facie* motivation for a normative account of rationality.

5   Kiesewetter, *The Normativity of Rationality*; Lord, "What You're Rationally Required to Do and What You Ought to Do (Are the Same Thing!)" and *The Importance of Being Rational*.

6   Raz, *Practical Reason and Norms*; Scanlon, *What We Owe to Each Other*; Dancy, *Practical Reality*; Parfit, *On What Matters*; Schroeder, *Slaves of the Passions*; Alvarez, *Kinds of Reasons*; Lord, *The Importance of Being Rational*. Gibbons, "Things that Make Things Reasonable," appeals to the connection between reasons and rationality to argue against the view that reasons are facts. In this paper, I assume a factualist conception of normative reasons, and I try to defend a conception of rationality compatible with such a conception.

7   Schroeder, *Slaves of the Passions*; Way, "Two Accounts of the Normativity of Rationality"; Parfit, *On What Matters*; Vogelstein, "Subjective Reasons"; Whiting, "Keep Things in Perspective"; Sylvan, "What Apparent Reasons Appear to Be"; Alvarez, "Reasons for Action, Acting for Reasons, and Rationality."

     As I explain in section 6, in order for a consideration to appear as a reason to an agent, she does not need to *represent* it as such—it may be enough if the agent is in a position to *treat* the relevant consideration as a reason (Sylvan, "What Apparent Reasons Appear to Be").

8   Kiesewetter, "A Dilemma for Parfit's Conception of Normativity."

9   Kiesewetter, *The Normativity of Rationality*; Lord, "What You're Rationally Required to Do and What You Ought to Do (Are the Same Thing!)," *The Importance of Being Rational*, and "Defeating the Externalist's Demons"; see also Ichikawa, "Internalism, Factivity, and Sufficient Reason."

constituted by facts about how things appear to her—that is, by reasons about appearances, rather than by apparent reasons.

This paper has two main goals, one critical and one constructive. The critical goal is to show that these sophisticated reasons-based accounts of rationality are still unsatisfactory, at least when combined with a factualist conception of normative reasons. The constructive goal of the paper is to argue that theories of rationality as responding to apparent reasons manage to do justice to the main intuitions seemingly supporting reasons-based accounts of rationality, in particular the connection between ascriptions of rationality and criticism and praise. The key move will be to abandon the view that being criticizable requires acting against one's reasons, while being praiseworthy involves doing what one's reasons actually recommend.

1. THE REASONS-BASED ACCOUNT OF RATIONALITY

An intuitively attractive idea is that being rational is a matter of being properly sensitive to the reasons one has.[10] This idea leads to an account of rationality along the following lines:

> *Reasons-Based Account* (RB): A response ϕ by an agent S is rational if and only if ϕ is, on balance, sufficiently supported by the reasons possessed by S.[11]

The response ϕ may be the performance of an intentional action or the adoption of some reason-sensitive attitude, such as believing or intending. Moreover, the reasons relevant for RB are normative reasons—that is, considerations that count in favor of the response in question. I will take it that such normative rea-

---

10 Comesaña and McGrath, "Having False Reasons," 61; Kiesewetter, *The Normativity of Rationality*; Lord, *The Importance of Being Rational*, 2, 220; Ichikawa, "Internalism, Factivity, and Sufficient Reason."

11 See Lord, *The Importance of Being Rational*, 2, 10–11, 220. Kiesewetter (*The Normativity of Rationality*, 162) explicitly proposes principles concerning rational belief and rational intention that capture something close to the left-to-right direction of the biconditional RB, even if presented in terms of support by decisive rather than sufficient reasons (this difference will not affect the essence of the arguments below). I will consider rational responses more generally, and I will be interested in discussing whether correctly responding to reasons is both necessary and sufficient for rationality. Lord does not only endorse the biconditional RB; he goes further and defines what it is to be rational in terms of responding to reasons: "Rationality consists in correctly responding to the objective reasons one possesses" (*The Importance of Being Rational*, 220).

sons are facts. These assumptions are shared by the advocates of RB I will be concerned with.[12]

I will further assume that reason possession is a perspective-sensitive relation involving some form of epistemic access to the fact constituting the reason, although I will not specify whether this access amounts to knowing, being in a position to know, believing, or something else. All the defenders of RB I am going to discuss agree that only *possessed* reasons determine what is rational for an agent to do (and also what she ought to do). One way of motivating this restriction is by offering an attractive account of three envelope cases, and of structurally analogous examples such as Parfit's mine-shaft scenario or Jackson's doctor cases.[13] In these examples, it is rational for the agent to ϕ despite her being aware that there exist facts beyond her epistemic ken that constitute decisive reasons to do something else instead. The type of view I am considering here deals with these cases easily, insofar as the relevant facts beyond the agent's epistemic ken do not figure among her possessed reasons, and thereby do not contribute to fixing what is rational for her to do.[14]

I will take it that the reasons possessed by an agent sufficiently support a certain response when the combined possessed reasons in favor of that response are at least as strong as the combined possessed reasons against it. Finally, if one wanted to specify what it is for an agent S to be rational in producing a response ϕ, it would be necessary to add the further condition that S's response is properly based on the sufficiently supporting possessed reasons. This condition sees to it that S produces her response for the reasons that rationalize producing it.

On the face of it, RB plus the standard assumption that normative reasons are facts leads to bad results in cases in which seemingly rational agents are guid-

---

12 Kiesewetter (*The Normativity of Rationality*, 6–8) agrees that normative reasons are typically facts, and argues against the thesis that normative reasons are generally mental states, although he leaves open the possibility that in certain particular cases there might be nonfactual normative reasons (e.g., perceptual seemings). Anyway, my aim is to discuss whether Kiesewetter's proposal can be made to work within the framework of a factualist conception of normative reasons.

13 Jackson, "Decision-Theoretic Consequentialism and the Nearest and Dearest Objection"; Parfit, *On What Matters*; also Kolodny and MacFarlane, "Ifs and Oughts"; Kiesewetter, "Ought and the Perspective of the Agent"; Schroeder, "Getting Perspective on Objective Reasons." A further way of motivating the restriction to possessed reasons is by appeal to guidance considerations (Kiesewetter, *The Normativity of Rationality*, 206–11; Lord, *The Importance of Being Rational,* 194–202; and, for discussion, Way and Whiting, "Perspectivism and the Argument from Guidance").

14 Kiesewetter, *The Normativity of Rationality*, ch. 8; Lord, *The Importance of Being Rational,* 188–94; Zimmerman, *Living with Uncertainty*. For a perspective-sensitive proposal in an expressivist spirit, see Schroeder, "Getting Perspective on Objective Reasons."

ed by false considerations.[15] There is the strong intuition that an agent can be rational despite relying on false beliefs, as long as from her point of view there appeared to be good reasons to behave as she did.[16] If one has extremely convincing evidence that *p*, then it seems rational to take *p* as a reason to produce a certain response, even if, unbeknownst to one, the evidence happens to be misleading and the belief that *p* is false. Think, for instance, of the agent who drinks from a glass of petrol, thinking that it contains gin and tonic.[17] More radically, it is very plausible that an agent deceived by a Cartesian demon, or a brain in a vat, can be as rational as her lucky counterpart in an ordinary environment.[18] In general, it seems that an agent will be rational if she takes proper heed of the considerations that, from her perspective, appear as good reasons, regardless of whether such considerations turn out not to be normative reasons after all. A deceived agent that is internally identical to her rational counterpart should also be regarded as rational (the deceived agent seems to be as blameless and as competent as her non-deceived counterpart).

Cases of deceived rational agents (I will call them "bad cases") put pressure on RB, since in these cases the agents' responses do not seem to be supported by normative, factual reasons and, nonetheless, they count as rational. Recent proposals by Kiesewetter and Lord suggest a reply to this problem on behalf of RB.[19] I will call this reply the *appearance reasons* view of rationality. According to this view, whenever an agent behaves rationally, there will be factual, normative

15 Probabilistic reasoning can also be seen to be problematic for factualist versions of RB. On the face of it, rational responses are often based on probabilistic considerations (e.g., given that it is likely to rain, it is rational for me to bring an umbrella). However, some may doubt that there are facts that correspond to these probabilistic rationalizers. (This problem is pressed by Wodak, "Can Objectivists Account for Subjective Reasons?") Although I do not have space to go into details here, I think that (factualist) RB can overcome this worry. One option is to think of the relevant probabilistic rationalizers as facts about probabilities on the agent's evidence (e.g., the fact that my evidence makes it likely that it will rain). Otherwise, one can resort to expressivist understandings of probabilistic facts (as recently proposed by Schroeder, "Getting Perspective on Objective Reasons"). Moreover, I think that it is plausible that such probabilistic facts constitute normative reasons counting in favor or against responses. (See Kiesewetter, *The Normativity of Rationality*, 177–78.) For instance, the fact that rain is likely on my evidence makes it risky for me not to take an umbrella, and arguably being risky detracts from the choice-worthiness of an option. Anyway, I will focus here on the problem posed to RB by rational behavior based on false beliefs, and I will leave aside the discussion of probabilistic rationalizers.

16 I will consider that how things appear from the agent's perspective is characterized by her nonfactive mental states, including non-doxastic seemings.

17 Williams, "Internal and External Reasons."

18 Cohen, "Justification and Truth."

19 Kiesewetter, *The Normativity of Rationality*; Lord, *The Importance of Being Rational* and "De-

reasons that rationalize her behavior. In the bad cases in which the agent relies on false beliefs, the facts that make her response rational will be, in general, facts about appearances—the fact that things appear to be to the agent as she believes them to be, or more broadly, facts that appear to support the agent's belief.[20] On this view, if p is a reason for a response in a good case, then the fact that it appears to the agent that p is a reason in the corresponding bad case. Thus, the central thesis of the appearance reasons view is that appearance facts can be normative reasons for the agent's response: such facts count in favor of the response; they recommend it.[21]

> *Appearance Reasons View*: If p is a reason to ϕ possessed by an agent S, then the fact that it appears that p is a reason to ϕ possessed by an agent S^* in a situation subjectively indistinguishable from that of S.

So, in the case of the glass with petrol, the fact that the glass appears to contain gin and tonic (the agent is in a bar and has ordered a gin and tonic) is a normative reason for the agent to drink from it, given that she is thirsty and fancies a gin and tonic.[22]

At this juncture, one may ask whether appearance reasons in the bad case are as strong as the reasons given by the appeared fact p in the good case. Kiesewetter thinks that they are, whereas Lord takes them to be weaker. These answers lead to different versions of the appearance reasons view of rationality. I will argue that both versions of the view are problematic.

2. DO WEAKER APPEARANCE REASONS RATIONALIZE?

Let us examine first views that assume that in bad cases the reasons favoring the

---

feating the Externalist's Demons." See also Ichikawa, "Internalism, Factivity, and Sufficient Reason."

20 Kiesewetter claims more generally that in bad cases appearances provide normative reasons, leaving open whether these reasons are constituted by the appearance state or by facts about appearances (*The Normativity of Rationality*, 174). I will interpret his proposal in the latter way, since I am interested in evaluating whether his account of rationality can be satisfactorily combined with a factualist conception of normative reasons.

21 Lord, "What You're Rationally Required to Do and What You Ought to Do (Are the Same Thing!)" and *The Importance of Being Rational*, 190–93.

22 Lord does not claim that agents have occurrent beliefs about facts about how things appear to them, but only that agents can be said to have such beliefs tacitly (*The Importance of Being Rational*, 191–92). According to Lord, this is enough for the agent to count as possessing, in the relevant sense, the reasons constituted by appearance facts. I will not press Lord on these issues.

response (call them *pro-reasons*) are weaker than in good cases. This assumption seems plausible. After all, these reasons are constituted by facts about appearances that *p* (where *p* is a reason in favor of the response), rather than by the fact that *p*, and, on the face of it, if these appearance facts give reasons for the response, it is only to the extent that they reliably indicate that *p* is the case.[23] I further discuss this issue in section 3; for the time being I will just grant that the assumption is right.

Lord concedes that appearance reasons are weaker than the reasons that would be given by the appeared facts. However, he suggests that, on balance, the reasons possessed by the agent in a bad case support her response *as strongly* as in the corresponding good case, at least in an important sense.[24] This is so because it is not only the pro-reasons that are weaker in the bad case. According to Lord, the reasons against the response (call them *con-reasons*) will also be equally weak, since they will be constituted as well by facts about appearances.[25] So on the view advanced by Lord, in bad cases the weights of the pro-reasons and the con-reasons are weakened by the same measure. As a result, the agent's response in a bad case will be as rational (i.e., as strongly supported, on balance, by possessed reasons) as in the corresponding good case. In this way, the agent's response in the bad case is as rational as her lucky counterpart's in the good case.

Lord's proposal is problematic in hybrid bad cases in which the pro-reasons are appearance facts but the con-reasons are the same as in the good case (and therefore are as weighty as in the good case). It may well be that, in a hybrid case, only the agent's pro-reasons are downgraded to appearance reasons due to the deceiving environment, while the con-reasons are left untouched.[26] Think of a situation in which the agent suffers visual illusions but her hearing remains reliable. If the agent's pro-reasons are acquired visually but the con-reasons are

23 See Comesaña and McGrath, "Having False Reasons."

24 Lord, *The Importance of Being Rational*, 193–98.

25 Lord, *The Importance of Being Rational*, 195.

26 Comesaña and McGrath present this sort of argument against the view that possessed reasons are facts about mental states ("Having False Reasons"). These hybrid scenarios are also discussed by Littlejohn ("Being More Realistic about Reasons") and Dutant ("Knowledge-First Evidentialism about Rationality"). Dutant argues that hybrid cases are problematic for a knowledge-first approach in which the only input to rational decision-making processes is the agent's knowledge, including knowledge about appearance facts. Dutant proposes instead that the input to decision-making is given by *what is supported* by the agent's knowledge (where the considerations supported for an agent in a bad case are a counterpart of what the agent in the corresponding good case knows). I take Dutant's positive proposal to be a knowledge-first version of apparent reasons accounts of rationality—note that what is supported may include false considerations. In the second part of the paper I explain why responses to apparent reasons (to what is supported) are linked to criticism and praise.

acquired via her hearing, then it seems that the pro-reasons but not the con-reasons will lose weight. Consequently, the weight of the agent's possessed reasons, on balance, will be different than in the corresponding good case (where her vision is also reliable).

Lord acknowledges that, in these hybrid scenarios, in which deception affects only the pro-reasons but not the con-reasons, his account does not predict that the agent in the bad case is as rational as her lucky counterpart in the good case.[27] However, Lord does not seem to recognize the undermining consequences that these cases have for his proposal. He claims that the agent's response in these hybrid cases is still rational, even if less defensible than the corresponding response of her counterpart in the good case—in the sense that, on balance, the reasons to produce the response are less strong in the hybrid case.[28] But, contrary to what Lord seems to think, there is no guarantee that, according to his account, the agent's action will continue to count as rational in the hybrid case. Indeed, it could be that, in the hybrid case, the con-reasons (shared with the good case) actually outweigh the downgraded appearance pro-reasons, so that producing the response stops being supported by sufficient reasons and therefore, according to RB, stops counting as rational.

The problem, thus, is not just that RB fails to account for the intuition that the agent in a hybrid case is as rational as her lucky counterpart in the good case. Furthermore, there can be hybrid cases in which RB predicts that the agent's response is irrational—even after one takes into account the relevant appearance reasons. However, it seems that the agent in the hybrid case can be rational (this is the intuition that the advocates of RB were trying to accommodate). At any rate, it seems weird to say that the agent in a completely deceiving scenario is rational while the agent in a partially deceiving environment (e.g., an environment in which her visual evidence is misleading but her hearing is reliable) is not. After all, from the agent's point of view both situations are indistinguishable. I take hybrid cases to pose a decisive objection to Lord's proposal. There is not much appeal in a position that holds that a partially deceived agent is irrational whereas her counterpart's in a systematically deceiving environment is not.

3. ARE APPEARANCE REASONS AS STRONG AS REASONS CONSTITUTED BY APPEARED FACTS?

It seems that the only way out for the defender of RB is to endorse the following view:

27 Lord, *The Importance of Being Rational*, 199–200.
28 Lord, *The Importance of Being Rational*, 199.

Equal Weight: The reason provided by the mere fact that it appears that *p* is, in the bad case, as strong as the reason provided by the fact that *p*, in the good case (at least insofar as the agent has no further reasons to suspect that *p* is false).[29]

I take this sort of view to follow naturally from factualism about reasons plus what Kiesewetter calls the Backup View:

Backup View: If *A*'s total phenomenal state supports p, and p would—if true—be an available reason for (or against) attitude *a*, then *A*'s appearances provide an equally strong available reason for (or against) attitude *a*.[30]

Note that the relevant backup reason cannot be constituted by the fact that, given the agent's evidence, it is likely that *p*. This reason is clearly weaker than the corresponding reason given by the fact that *p* in the good case, and therefore will lead to problems in hybrid scenarios, as discussed above.[31] Thus, I will take it that the sort of fact that can generally act as the agent's reason in bad cases is the fact that it appears to her that *p*.

I argue that Equal Weight is in tension with the core motivations for factualism about reasons. Thus, Equal Weight should be unappealing for those who want to combine an account for rationality as responding to reasons with a conception of reasons as facts that favor some response.

I take it as *prima facie* plausible that when the fact that *p* is a reason favoring *φ*-ing, the fact that it appears to the agent that *p* is a reason for *φ*-ing only to the extent that there is some reliable connection between the appearance and *p*. This idea becomes particularly natural when we note that appearance reasons are dependent on worldly reasons, in the sense that it is only because we take *p* to be a reason for *φ*-ing that we would also be inclined to take the fact that it appears

---

29 Littlejohn discusses critically a similar principle, which he calls *Exchangeability*. "Exchangeability: if the fact that *p* constitutes a potent normative reason for *A* to *f* that has weight *W*, the fact that it seems to *A* that *p* constitutes a potent normative reason for *A* to *f* that also has *W*" ("Being More Realistic about Reasons," 10). Note that, taken literally, *Exchangeability* applies just to good cases in which the worldly fact *p* is a reason available to the agent. Equal Weight, by contrast, applies as well to bad cases, in which *p* is not a fact. I am sympathetic to most of Littlejohn's objections to Exchangeability, although the arguments I present here are in general different.

30 Kiesewetter, *The Normativity of Rationality*, 174. See also Huemer, "Compassionate Phenomenal Conservatism."

31 So, when deciding whether to drink from a glass of milk, one's rational behavior will be different on the assumption that the milk is not poisoned and on the assumption that it is likely that it is not poisoned.

that p as a reason for it. Below I will say more about how factualist conceptions of reasons motivate this idea, but for the moment let us assume that it is on the right track.

Now, in general the connection between the bare fact that it appears that p and the obtaining of p will not be perfectly reliable (appearances are sometimes misleading). Moreover, note that in the bad case the appearance reason cannot be bolstered by the further auxiliary consideration that the appearance is accurate or very reliable in the context the agent is in: this consideration would be false in the bad case, so it cannot be part of a normative, factual reason in favor of the response. Without the support of the assumption that appearances are reliable in the context at hand, the (evidential) probability of p given the fact that it appears that p will be generally lower than its probability given the fact that p. The difference is easier to appreciate from a third-person perspective. If you are not assuming that Jane's appearances are reliable, then you will in general assign more probability to p on the basis of your knowing that p than on the basis of the bare fact that it appears to Jane that p. In this way, we will feel more inclined to choose a response ϕ that would be recommended by p if we rely on the fact that p than if we merely rely on the fact that it appears that p. This suggests that the reason in favor of ϕ given by the fact that p is stronger than the reason given by the mere fact that it appears that p.[32]

It is true that, in the bad case, the agent will be as confident that p as her lucky counterpart who, in the good case, knows that p. But, arguably, this is so because the subject in the bad case forms such confidence on the basis of the nonfactual consideration that p (which she takes to be a fact), and not merely on the basis of *the fact that it appears to her that p* (that is, not on the basis of the consideration that it appears to her that p). The deceived agent will mistakenly take herself to know that p, so she will see herself as being in a position to rely on the (apparent) fact that p, and accordingly she will have high confidence that p. Yet, plausibly, if she had relied merely on the fact that it appears to her that p, her confidence in p would have been lower. We do not always form beliefs on the basis of facts; sometimes we rely on appearances that turn out not to be factual. In other words, if reasons are facts about which we are fallible, then it will be possible that an agent is mistaken about what reasons she actually has. In particular, it may be that the agent's actual factual reasons to believe p are weaker than she takes them to be, so that her degree of confidence in p does not match what is actually warranted by her possessed, factual reasons.[33]

---

32   See Littlejohn, "Being More Realistic about Reasons," for considerations in a similar spirit.

33   In a similar way, Williamson insists that, if evidence is taken to be constituted by known facts, then agents may be mistaken about what evidence they have (*Knowledge and Its Limits*).

Actually, in deceiving contexts, it may well be that the connection between appearances that *p* and the fact that *p* is largely severed, so that appearances that *p* are extremely unreliable. In a brain-in-a-vat case, appearances never match the facts. In such scenarios, therefore, the fact that it appears to the deceived subject that *p* is not related at all to the obtaining of *p*—such an appearance fact would not actually be reliably linked with the truth of *p* (even if the deceived agent mistakenly thinks that it does). In this sort of radically deceiving scenario, the putative reasons provided by appearance considerations are undercut (or at least, such reasons are greatly attenuated).[34] This is so because, in such scenarios, the fact that it appears that *p* may be completely disconnected from the sort of features that contribute toward the fittingness of the response ϕ (where ϕ is the response that would be recommended by the fact that *p*). And, arguably, some fact favors ϕ-ing only if there is some suitable link between such a fact and features of ϕ-ing that contribute toward its fittingness or rightness. I am not assuming that the response is always made fitting by the reason-constituting fact itself. But, at least, the reason must reliably indicate that some feature in the response contributes to its fittingness. In other words, there must be some relevant, reliable connection between the fact that constitutes the reason- and fitting-making features of the response. Otherwise, it would not be clear in what sense the putative reason actually favors the response.[35]

Assume, for instance, that the fact that this glass contains gin and tonic is a reason for Theresa to drink from it because it will quench her thirst—and this makes (or at least contributes to making) drinking from the glass a fitting thing to do. In other words, quenching Theresa's thirst is the feature of drinking from the glass that favors doing so. In a radically deceiving scenario, the fact that it appears that the glass contains gin and tonic does not count at all in favor of drinking from it, since such an appearance fact is disconnected from the proper-

34 For the notion of reason attenuation, see Dancy, *Ethics without Principles*, and Schroeder, *Slaves of the Passions*.

35 I remain neutral about whether the favoring relation can be accounted for in terms of the notions of fittingness or rightness, perhaps in combination with other notions such as evidence or explanation (Schroeter and Schroeter, "Reasons as Right-Makers"; Chappell, "Fittingness"; McHugh and Way, "Fittingness First"; Whiting, "Right in Some Respects"; Howard, "The Fundamentality of Fit"), or whether favoring is instead a primitive normative relation that explains which attitudes are fitting (for instance, Scanlon, *What We Owe to Each Other*; Schroeder, *Slaves of the Passions*; Parfit, *On What Matters*). I am only assuming that there is some connection between reasons and fittingness, without taking a stand on their relative primacy. I will use interchangeably terms like "fitting," "right," "appropriate," and "recommendable." Although there may be differences of detail in the usage of these terms, I take it that they all point toward the sort of notion I am interested in. For further discussion, see McHugh and Way, "Fittingness First."

ties that would make drinking from the glass fitting (i.e., the property of quenching thirst). In other words, in radically deceiving contexts the putative reasons provided by appearance considerations are weakened or even undercut, because the favoring relation connecting them with the relevant response becomes undermined.

If one wants to insist that radically deceiving appearance facts provide non-attenuated (normative) reasons for the relevant responses, one would have to postulate that, whenever a fact p favors the response ϕ, the fact that it appears that p also favors that response with the same intensity—even in deceiving contexts in which appearances that p are completely disconnected from the actual presence of p. While this is a possible position, I think it is not particularly attractive for those sharing the externalist intuitions associated with factualism about reasons (it would fit more naturally with an internalist view such as Huemer's).[36] Anyway, this discussion points toward the more general question of whether rational subjects can be reasonably mistaken in taking some fact to favor a given response. It is this more general issue that I address in the next section. I shall argue that it puts a great strain on the view that appearance facts generally provide non-attenuated reasons.

4. REASONABLE MISTAKES ABOUT THE FAVORING RELATION

On the factualist approach I am considering, a normative reason is a fact that actually favors some response. In general, it seems that one can be mistaken about whether a consideration favors a given response, and about the strength of that favoring relation.[37] For instance, Alice may mistakenly (even if perhaps reasonably) treat the fact that there is an electrical fire in the kitchen as giving her decisive reasons to pour water on it. Nevertheless, pouring water on an electrical fire is not actually a fitting thing for Alice to do: it can be extremely dangerous. The presence of an electrical fire does not actually favor pouring water.

In the same way that one can be rational when guided by false considerations, one can also be rational when mistakenly taking some fact to favor a certain response—perhaps it reasonably appeared to the agent that there was such a favoring connection between the fact and the response. I am not assuming that in

---

36 Huemer, "Compassionate Phenomenal Conservatism." Granted, it may be that Kiesewetter's proposal has additional theoretical benefits that justify abandoning the externalist views associated with factualism about reasons. My aim here is to see whether we can satisfactorily account for rationality while preserving factualism and the externalist intuitions that motivate it.

37 See Littlejohn, "The Unity of Reason."

order for it to appear to an agent that p favors φ-ing she needs to form an explicit belief that this is so. Arguably, she just needs to treat p as favoring ϕ-ing.[38] Thus, the sorts of mistakes I want to discuss are mistakes in treating p as favoring ϕ-ing, which need not involve mistaken explicit beliefs about such a favoring relation.

Now, it is not clear what could be said about these cases by those who want to make rationality a matter of responding to normative reasons (i.e., by advocates of RB). Again, note that in these situations the agent is not relying on considerations that favor ϕ-ing; rather, she would be relying on considerations that she *mistakenly treats as favoring ϕ-ing*. But these considerations would not be real normative reasons for ϕ-ing, insofar as the factualist is understanding normative reasons as facts that (actually) favor a response. Therefore, the defender of RB will not be able to say that the agent's response is made rational by such considerations. Moreover, if the fact that p does not favor ϕ-ing, in general the fact that it appears that p will not favor ϕ-ing either, so appealing to this further appearance fact does not seem to help.

So, if proponents of RB want to be able to say that the agent's response in the cases I am considering is rational, they seem committed to denying that an agent can be reasonably mistaken about whether some response is favored by the considerations available to her: if it reasonably appears to the agent that some consideration favors ϕ-ing, then it would actually favor ϕ-ing (or at least, there would be some backup consideration available to her that would actually favor ϕ-ing). I find this position unappealing. I do not see why we should attribute such infallibility to our rational dispositions to treat some consideration as favoring a response. At any rate, this commitment puts a heavy burden on those defending RB.

At this point, one option is to abandon the view that normative reasons are facts that *actually* favor a response. More specifically, one could say that normative reasons are facts that *appear* to favor the response. Yet merely apparent normative support is not actual support. Moreover, once this step is taken, I see no principled reason not to go further and claim that normative reasons are considerations that *appear to be facts* that (appear to) favor the response. If the important thing is not whether the reason actually favors the response but rather whether it appears to do so, why should one demand that the reason is a fact, and not just a consideration appearing to the agent as a fact? Since I am assuming the thesis that reasons are actually facts counting in favor of a response, I will leave aside this possibility.

An alternative strategy would be to suggest that, in good cases, the subject has as one of her possessed reasons the fact that the considerations she is relying

38 Sylvan, "What Apparent Reasons Appear to Be."

on actually favor ϕ-ing. The idea would be to argue that the fact that the favoring relation obtains should be included as an additional available reason. Then, in deceiving scenarios one could replicate Lord's and Kiesewetter's strategy: in bad cases, the further reason would be provided by the fact that it *appears to the agent* that the considerations she is relying on actually favor ϕ-ing.

This proposal, however, is not very promising. As I have been arguing, it is doubtful that in general a response ϕ is favored by p just because of the mere fact that it appears to the agent that ϕ-ing is favored by p. Perhaps the agent was reasonable in treating p as favoring ϕ-ing, but this does not mean that φ-ing was actually favored. In certain contexts, our normative judgments about the favoring relation can be widely unreliable. In this way, the presence of an electrical fire does not favor pouring water just because of the fact that it appears so to Alice. Likewise, making decisions with the guidance of tea-leaf reading does not become a fitting thing to do just because of the fact that it appears to be so in the community of Joe the Fortune Teller. At least for those of us committed to fallibilism about our epistemic capacities, it seems unmotivated to think that rational agents are always reliable about whether some response is favored by certain aspects of a situation.

Moreover, this proposal faces the charge of overintellectualizing rationality. Arguably, in order for considerations about the favoring relation to be part of one's *possessed* reasons, one needs to be able to form beliefs about such considerations. However, beliefs about the favoring relation are rather sophisticated, and it is controversial that all rational agents need to be able to form beliefs with such contents.[39] Thus, it is problematic to require that rational agents always have among their possessed reasons considerations about the favoring relation.

A final line of resistance for the defender of RB would be to bite the bullet and maintain that in the examples I am discussing ϕ-ing is not actually supported by reasons and therefore is not rational, even if it may be rational for the agent to take ϕ-ing to be sufficiently supported by reasons (and, accordingly, it may also be rational for her to believe that φ-ing is rational).[40] Several authors have recently argued that, when agents have misleading higher-order evidence about what attitudes are rational, there can be mismatches or incoherences between the actual rationality of first-order attitudes and the rationality of higher-order beliefs about what first-order attitudes are rational.[41] According to these types of views, it may be rational for agents with misleading higher-order evidence to

39   See McHugh and Way, "Against the Taking Condition."

40   I thank an anonymous referee for bringing this possibility to my attention.

41   See, for instance, Coates, "Rational Epistemic Akrasia"; Wedgwood, "Justified Inference"; Lasonen-Aarnio, "Higher-Order Evidence and the Limits of Defeat"; Christensen, "Dis-

believe (falsely) that φ-ing is rational and supported by reasons in cases where φ-ing is actually irrational and not supported by reasons (these proposals tend to focus on doxastic attitudes, but the arguments motivating them can be generalized to other attitudes and responses).

It should be noted that it is far from uncontroversial whether this type of mismatch between rational first-order and higher-order attitudes can actually take place.[42] Anyway, even if there could be such rational mismatches, I do not think that proponents of RB would be in a position to appeal to them in order to provide an attractive response to the worries discussed in this section. Remember that the advocates of RB I am engaging with are interested in a notion of rationality that captures the intuitive connection between rationality and criticizability and praiseworthiness—this was the main motivation for developing a normative characterization of rationality in terms of reasons.[43] However, authors who defend the possibility of rational mismatches between higher-order and first-order attitudes typically concede that, when it is reasonable for the agent to believe that φ-ing is rational, so that she is not criticizable for having such a higher-order belief, she is not criticizable for φ-ing either, even if φ-ing happens to be actually irrational.[44] In this way, criticizability and irrationality (understood in this sense) would come apart. So, proposals that allow for rational mismatches between higher-order and first-order attitudes seem to rely on an externalist notion of rationality that swings free from criticizability and praiseworthiness. This is not the notion of rationality I am concerned with in this paper, since I am working under the assumption that there is an intimate relation between rationality and criticizability.[45]

To sum up, the possibility of rational mistakes about the favoring relation creates problems for RB. It seems that agents can be rational not just when relying on false beliefs, but also when mistakenly taking some consideration to favor a certain response.[46] More generally, it is not clear why we should think that ratio-

---

agreement, Drugs, etc.”; Worsnip, “The Conflict of Evidence and Coherence”; Weatherson, “Do Judgments Screen Evidence?”

42 For objections, see among others Horowitz, “Epistemic Akrasia”; Littlejohn, “Stop Making Sense?”

43 Kiesewetter, *The Normativity of Rationality*, ch. 2.

44 Lasonen-Aarnio, “Higher-Order Evidence and the Limits of Defeat,” 343.

45 Note, moreover, that Kiesewetter (“You Ought to φ Only if You May Believe that You Ought to φ”) explicitly argues that agents have decisive reasons to φ only if they have sufficient reasons to believe that they have decisive reasons to φ. Thus, he does not seem to be willing to allow for the kind of rational incoherence between higher-order and first-order attitudes discussed above.

46 It is worth observing that such mistakes are compatible with the view that whenever an

nal agents are always infallible about what responses are actually favored by the reasons available to them.

5. APPEARANCE FACTS AND JUSTIFICATION

Before moving forward, I will consider a further awkward consequence of Equal Weight. If this view were right, then subjects would be wise to cite appearance facts as their reasons when asked to justify themselves. After all, reasons constituted by appearance facts would be safer. In good cases, in which the belief that p is true, the fact that p and the fact that it appears to the subject that p would provide the subject with equally weighty reasons, so there is no loss in relying on the latter. Moreover, in a deceiving situation in which the subject falsely believes that p, the fact that it appears to her that p would still constitute a normative reason. Why should one recklessly appeal to the riskier putative reason (i.e., the consideration that p), if there is nothing to gain from it? When two putative reasons are equally strong, it is a sound argumentative strategy to rely on the safest one (i.e., the one that is less likely to be false), in order to limit the chances of error. By doing this, one does not always implicate that the more insecure consideration is false: it is just that one does not need to run the risk of relying on it. Thus, when writing a paper, it is common to announce that, in order to defend some thesis, one is going to focus on an argument for it that involves as few unnecessary commitments as possible (even if there exist further, more controversial arguments). In this way, one avoids unnecessary exposure to objections.

Thus, according to Equal Weight, when asked about one's reasons for φ-ing, one should cite the fact that it appears to one that p, rather than directly appealing to the (less safe) consideration that p. However, this is not how we generally proceed. In most cases, we justify ourselves by referring to those worldly considerations that support our responses, instead of appealing to facts about appearances.[47] Perhaps there are situations in which one is not sure enough and retreats to the fact that things appear to be in a certain way. But, precisely, these are cases in which I will be seen as recoiling to a weaker reason.[48] For instance,

---

agent actually has reasons to ϕ, she has sufficient reasons to believe that she has reasons to ϕ (Kiesewetter, "You Ought to ϕ Only if You May Believe that You Ought to ϕ"). I am only claiming that one can be reasonably mistaken in believing that one has certain reasons to ϕ. In particular, one may actually *lack* reasons to φ despite not having sufficient reasons to believe that this is so.

47 Dancy, *Practical Reality*; Alvarez, *Kinds of Reasons*.

48 Actually, it can be argued that sometimes in such cases I will be giving an excuse rather than a justification. (See the discussion in section 6, below.)

if you ask me why I am taking an umbrella, I may reply that my reason is the fact that it is raining. If I say instead that my reason is the fact that it appears to me that it is raining, you will probably take me not to be in a position to undertake the commitments associated with citing the stronger but riskier putative reason that it is actually raining. According to Equal Weight, by contrast, appealing to appearance facts should be our justificatory policy across the board, and not just in special cases. This, I think, is a distorted picture of our practices of giving and asking for reasons. At any rate, it seems that such a picture should be unappealing for those moved by the arguments supporting the thesis that normative reasons are facts and not mental states—in particular, the argument that, when justifying ourselves, we tend to refer to external facts rather than to features of our mental states.[49]

To be sure, one may try to find ways to resist this line of argument. A possible option is to offer some pragmatic explanation for our tendency to appeal to worldly facts rather than to facts about appearances when justifying ourselves. For instance, it may be argued that, in most situations, our conversational aims are not only justificatory but also informative, and in general citing the fact that *p* is more informative than merely referring to the fact that it appears that *p*. It is not clear, however, that in justificatory contexts (which are often noncooperative) speakers generally have the additional goal to communicate worldly information, even at the expense of the strength of their justificatory position: it may well be that the speaker's only aim is to offer the safest justification possible for her behavior. Anyway, this sort of pragmatic explanation, if successful, would undermine the motivations for factualism about reasons, since a similar story would be available to mentalists in order to explain why speakers typically justify themselves by citing worldly facts rather than mental states.

A second option is to endorse some form of error theory and claim that ordinary speakers are usually unaware that Equal Weight holds. Again, this proposal would diminish the appeal of factualism about reasons, given that mentalists could appeal to an analogous type of error theory. Moreover, it would have the costs generally associated with error theories; in particular, it would be necessary to offer an explanation of why ordinary speakers are systematically mistaken about their justificatory practices.

---

49 Dancy, *Practical Reality*; Schroeder, *Slaves of the Passions*; Hornsby, "A Disjunctive Theory of Acting for Reasons"; Alvarez, *Kinds of Reasons*; Kiesewetter, *The Normativity of Rationality*, 7–8. Gibbons suggests that proposals along the lines of Lord's and Kiesewetter's are mere "notational variants" of a nonfactualist, mentalist conception of normative reasons, according to which reasons are given by features of the agent's psychological states ("Things That Make Things Reasonable," 358–59).

As I see it, the considerations I have been discussing put a great deal of pressure on Equal Weight. The underlying problem is that if we admit (equally weighty) reasons given by appearance facts, then the reasons given by worldly facts become in a sense normatively superfluous: the former, on their own, would be enough to figure out what we ought to do and to justify ourselves. I take it that this conclusion will not satisfy those with sympathies for factualism about reasons. Perhaps there is some way of defending Equal Weight, but it would be *prima facie* desirable to have a satisfactory account of rationality that does not undertake its weighty commitments. I conclude the paper by showing how an account of rationality in terms of apparent reasons manages to do so, while accommodating our intuitions about the normative implications of being rational.

6. RATIONALITY AND APPARENT REASONS

Apparent reasons are considerations that appear as reasons to the agent, perhaps mistakenly.[50] A *merely apparent* reason would not actually be a reason, in the same way that a fake Picasso is not actually a Picasso (even if it may appear to be one). The contents of false beliefs can be apparent reasons for agents that take them to be true. Thus, if rationality is seen as a matter of responding to apparent reasons, we can account for the possibility of being rationally guided by false considerations.[51] This sort of view can be sketched as follows:

> *Apparent Reasons Account* (AR): A response ϕ by an agent S is rational if and only if ϕ-ing is, on balance, sufficiently supported by S's apparent possessed reasons.

We do not need to assume that an agent must believe that p is a possessed reason for ϕ-ing in order for p to appear to her as such a reason. It may be enough if the agent treats (or is in a position to treat) p as a possessed reason for ϕ-ing, for instance in her reasoning and as a guide for action.[52] In this way, p can appear

---

50 Way, "Two Accounts of the Normativity of Rationality"; Parfit, *On What Matters*; Vogelstein, "Subjective Reasons"; Whiting, "Keep Things in Perspective"; Sylvan, "What Apparent Reasons Appear to Be"; Alvarez, "Reasons for Action, Acting for Reasons, and Rationality."

51 Schroeder, *Slaves of the Passions*; Way, "Two Accounts of the Normativity of Rationality"; Parfit, *On What Matters*; Vogelstein, "Subjective Reasons"; Whiting, "Keep Things in Perspective"; Sylvan, "What Apparent Reasons Appear to Be."

52 See Sylvan, "What Apparent Reasons Appear to Be"; Lord, *The Importance of Being Rational,* 171–75.

as a reason to agents that do not have the conceptual resources to form explicit beliefs about reasons, thereby avoiding charges of overintellectualization.

One may think that a problem with AR is that it does not capture the normative dimension of rationality.[53] Arguably, merely apparent reasons do not have genuine normative force. Therefore, it would not always be the case that being rational is a matter of responding to actual reasons, of doing what one ought to or may do. Although I will grant that this last claim is right, I will argue that it does not undermine AR. It is possible for defenders of AR to do justice to the considerations that seem to support a normative account of rationality.

As discussed at the beginning of the paper, I take it that the main motivation for thinking that there is something normative in rationality is that ascriptions of rationality are directly connected with criticism and praise. Thus, ascriptions of rationality have a *hypological* dimension, i.e., they have to do with blame, credit, criticizing, and praising.[54]

Now, it could be that hypological and deontic judgments sometimes come apart. Some views hold that the hypological and the deontic may diverge in several ways.[55] According to these types of views, doing something wrong or impermissible does not always make one deserve criticism. Conversely, one may think that agents can be criticizable without doing something wrong or impermissible. Similarly, it can be argued that being praiseworthy does not always go hand in hand with doing what is right or what one ought to do. If this is on the right track, then the fact that being rational is generally praiseworthy (and being irrational, criticizable) does not entail that rationality amounts to doing what there are most reasons to do. Let me present some examples that motivate the sorts of divergences between hypological and deontological appraisals that I have in mind.

Excuses provide examples of situations in which one does something wrong or inappropriate but is not criticizable for it.[56] Imagine that you accidentally step on someone's toe. You may excuse yourself by claiming that it was an accident. Blameless ignorance can also serve as an excuse. If you offer your friend a glass of milk that, unbeknownst to you, is poisoned, you may excuse your action by

---

53 Kiesewetter, "A Dilemma for Parfit's Conception of Normativity."

54 For the notion of hypological judgment, see Zimmerman, "Taking Luck Seriously"; also Littlejohn, *Justification and the Truth-Connection*.

55 See Austin, "A Plea for Excuses"; Strawson, "Freedom and Resentment"; Wallace, *Responsibility and the Moral Sentiments*; Baron, "Justifications and Excuses"; Gardner, *Offences and Defences*; Capes, "Blameworthiness without Wrongdoing"; Littlejohn, *Justification and the Truth-Connection*, "The Unity of Reason," and "A Plea for Epistemic Excuses"; Srinivasan, "Normativity without Cartesian Privilege"; Williamson, *Knowledge and Its Limits*.

56 Wallace, *Responsibility and the Moral Sentiments*; Gardner, *Offences and Defences*; Littlejohn, "A Plea for Epistemic Excuses."

claiming that you had no way of knowing that the milk was unsafe. In these sorts of cases, by giving an excuse you try to exonerate yourself from blame, while acknowledging what you did is not something that should be done (you are not trying to justify yourself, to claim that you actually had reasons to do what you did).

There are also cases in which the behavior of agents that fail to do what is right seems to be praiseworthy, not merely excusable. More generally, one's performance can be praiseworthy despite not achieving its fitting aim. As long as one behaves competently and does what is in one's power to act appropriately, one can be worthy of praise, regardless of the ultimate success of one's performance. Think of the archer who performs masterfully a difficult shot, but who fails to hit the mark due to the unforeseeable, last moment interference of a hidden magnet. Surely this archer deserves praise, given the skills she has revealed.

Promise-keeping offers another example. An agent who goes far beyond what most people would in order to honor her promise may be praiseworthy as a promise-keeper, despite failing to do as promised. This can be seen as a situation in which an agent is praiseworthy while failing to do what she ought to. By contrast, it can happen that an agent who fulfils her promise merely out of luck is criticizable as a poor promise-keeper if she showed no care for her promise and made no effort whatsoever to keep it. This reckless promise-keeper would be criticizable despite having done what she had most reasons to do. More generally, agents may be criticizable for doing something that is actually permissible (i.e., it is sufficiently supported by the reasons available to the agent) if they do it for the wrong reasons—analogously, an archer may hit the target out of luck despite shooting in an incompetent way that deserves criticism.

The examples I have examined only illustrate some relevant ways in which deontological and hypological statuses may come apart, without covering all possible combinations of such statuses.[57] I do not intend these examples to offer decisive evidence for the view I want to consider but rather to motivate it. To be sure, one could always try to interpret some of the examples as cases in which the agent fails to do what is right to do, but nevertheless is praiseworthy for something else she has done (e.g., the agent was praiseworthy for doing her best to keep her promise). Ultimately, one could retreat to arguing that agents have reasons to *try* to do what they ought to do, or more generally to do their best to pursue their fitting aims. However, I take it that these examples can also be naturally interpreted as revealing a divergence between deontological and hy-

---

57   For instance, one could also imagine an agent who does what she ought to do for the wrong reasons, but who is nonetheless non-critcizable (maybe even praiseworthy) because such reasons appeared to her as sufficiently good.

pological appraisals of the agent's behavior: they would be cases in which agents are praiseworthy for doing something that was actually not right, or in which agents are criticizable for doing something that was the right thing to do. More to the point, this interpretation of the cases would suggest that agents may be praiseworthy while failing to do what they had most reasons to do, and criticizable despite doing what they had most reasons to do. In what follows, I sketch a plausible theoretical framework that makes this interpretation attractive and explains why agents are worthy of praise or criticism in these types of cases. This framework allows for an account of rationality as responding to apparent reasons that does justice to our practices of rational criticism and praise while avoiding the difficulties and problematic commitments of RB.

The ideas I want to explore can be perspicuously expressed from the perspective of virtue theory.[58] Appealing to virtue-theoretical language, praise and criticism could be said to be primarily related to evaluations of the agent's *competence*, rather to evaluations of the success or aptness of her performance (where an apt performance is a performance that is successful or correct due to the exercise of the agent's competence). Competences are a type of reliable disposition to perform well or fittingly, at least in favorable circumstances.[59] No matter how competently an agent behaves, it can always happen that she is unlucky and, due to external interferences and factors out of her control, her performance turns out not to be successful (it fails to achieve its aim). So, someone extremely careful may (blamelessly) end up doing something actually dangerous, despite behaving in a way that manifests a careful disposition. In these cases, the agent's display of competence can still be praiseworthy, as the example of the archer shows.

We can follow Sylvan in thinking of guidance by reasons from such a virtue-theoretical perspective.[60] Rational agents aim to be guided by normative reasons, so that they produce fitting responses. Behaving rationally would

---

58 See Sosa, "How Competence Matters in Epistemology" and *Judgment and Agency*; Sylvan, "What Apparent Reasons Appear to Be"; Boult, "Epistemic Normativity and the Justification-Excuse Distinction"; Miracchi, "Competent Perspectives and the New Evil Demon Problem."

59 Sosa, "How Competence Matters in Epistemology" and *Judgment and Agency*; Sylvan, "What Apparent Reasons Appear to Be." Kelp regards competences as aim-directed dispositions characterized by their etiology, in particular by the selective conditions under which they are formed or sustained ("How to Be a Reliabilist"). By contrast, Sosa takes the notion of competence to be a normative primitive (*Judgment and Agency*, 195–206). We could also conceive of competences as the sorts of dispositions displayed by virtuous agents. For instance, competences could be seen as those dispositions shared with an internally indistinguishable virtuous agent who succeeds in behaving aptly.

60 Sylvan, "What Apparent Reasons Appear to Be." See also Boult, "Epistemic Normativity and the Justification-Excuse Distinction."

amount to behaving in a way that reveals one's competence in responding only to reasons. Competent agents, however, will not always succeed in answering only to normative reasons. A response that is not actually supported by the agent's possessed reasons may still display the agent's (fallible) competence in being guided only by normative reasons. For instance, in deceiving scenarios, the most competent reasoner may end up relying on considerations that are not actually normative reasons for her response. Nonetheless, this does not need to be seen as tarnishing her rational competence.[61] If by relying on such considerations the agent displays her competence in responding to reasons, she may be praiseworthy. Competently responding to one's possessed reasons is in general something praiseworthy, insofar as exercising such a competence demonstrates one's sensitivity to the fitting-making features of one's choices. It could be said the manifestation of this competence reveals a virtuous character trait.[62] Another way of putting it is to say that exercising this competence reflects positively on one's quality of will or mind.[63]

Identifying rationality with mere blamelessness is problematic because being rational is a positive status, rather than just an exculpatory one.[64] However, as we have seen, manifestations of competences aiming at fitting targets deserve positive appraisals, not just exculpatory ones. Arguably, there is derivative (albeit noninstrumental) value in properly valuing and respecting something that is itself valuable.[65] Manifesting competence in being guided only by reasons is a way of showing proper concern for the fitting-making features of one's responses. So, if fitting responses are valuable, then the manifestation of rational dispositions will (derivatively) deserve positive evaluations.[66] By contrast, behaviors that involve a poor exercise of one's competence as a follower of reasons may be subject to criticism, to the extent that such behaviors show a lack of concern for

---

61 See Sosa, *Judgment and Agency*; Sylvan, "What Apparent Reasons Appear to Be"; Boult, "Epistemic Normativity and the Justification-Excuse Distinction"; Littlejohn, "A Plea for Epistemic Excuses."

62 Svavarsdóttir, "Evaluations of Rationality"; Wedgwood, "Rationality as a Virtue"; Boult, "Epistemic Normativity and the Justification-Excuse Distinction."

63 Littlejohn, "The Unity of Reason." For the relation between criticism and the quality of will, see Strawson, "Freedom and Resentment"; Hieronymi, "The Force and Fairness of Blame"; Zimmerman, "Moral Responsibility and Quality of Will."

64 Baron, "Justifications and Excuses"; Gerken, "Warrant and Action"; Boult, "Epistemic Normativity and the Justification-Excuse Distinction"; Littlejohn, "A Plea for Epistemic Excuses"; Madison, "On Justifications and Excuses."

65 Hurka, *Virtue, Vice, and Value*; Sylvan, "Veritism Unswamped."

66 Miracchi, "Competent Perspectives and the New Evil Demon Problem"; Sylvan, "Veritism Unswamped."

what is fitting. It is important to note that an agent can engage in such criticizable behaviors without stopping to be responsible for the exercise of her rational faculties (that is, without being excluded from the realm of rational beings). In the same way, an agent that breaks a promise in a criticizable way may keep having promise-keeping capacities, even if she exercises them poorly in relation to that particular promise. Thus, the idea is not that the agent is criticizable for lacking some virtuous trait, but rather that she is criticizable for failing to manifest a competence she possesses and is in a position to exercise (say, the failure may be due to the interference of some vicious disposition that reflects lack of concern for the fitting-making features of one's choices).[67]

Thus, if AR is spelled out in terms of competences, the picture we get is that rational responses manifest the agent's competence in being guided only by reasons, a competence that involves being sensitive to those considerations that appear to one as possessed reasons.[68] One way to go here is to offer an independent analysis of what apparent reasons are, and then appeal to it in order to further illuminate the notion of rational competence. To a first approximation, an apparent reason could be seen as a consideration believed by the agent and that would be a normative reason if things were as the agent takes them to be.[69] This approach, however, is problematic in cases in which the agent has crazy beliefs that go wildly against the evidence available. One may want to say that such an agent is irrational, even if she behaves as would be rational if things were as she takes them to be.[70] Whiting avoids this problem by requiring that the relevant

---

67 When the agent lacks the relevant competence, her responses are not rational or irrational but rather arational (e.g., compulsive behaviors). Kiesewetter (*The Normativity of Rationality*, 37–38) discusses cases in which an agent who lacks some virtuous trait is not criticizable for failing to do what a virtuous person in her position would do (for instance because she knows that doing so would trigger an uncontrollable response with undesirable consequences). These cases, however, are not problematic for the competence-based approach I am sketching. Insofar as the agent in such cases is not in a position to manifest the relevant competences, she is not criticizable for failing to do so. To be sure, the agent is perhaps criticizable for not having cultivated the sort of faculty that would have conferred her rational control over those responses, or for not avoiding situations in which such responses are to be produced. In general, I think that the competence-based view I am presenting avoids Kiesewetter's objections to virtue-theoretical accounts of criticizability and rationality, although I do not have space to develop this issue in detail here.

68 Sylvan, "What Apparent Reasons Appear to Be."

69 Schroeder, *Slaves of the Passions*; Parfit, *On What Matters*. For more sophisticated accounts, see Vogelstein, "Subjective Reasons"; Whiting, "Keep Things in Perspective." This type of account of apparent reasons in terms of normative reasons has been recently criticized by Wodak, "Can Objectivists Account for Subjective Reasons?" For a possible reply to some of these worries, see Schroeder, "Getting Perspective on Objective Reasons."

70 For this line of criticism, see Kiesewetter, *The Normativity of Rationality*, 31–33.

beliefs are themselves rational (of course, this makes the view unsatisfactory as a reductive account of rationality, but this is not Whiting's goal).[71]

An alternative strategy is to characterize the idea of apparent reason in terms of the notion of competence, which would be taken as more basic. On Sylvan's view, some consideration is an apparent reason for an agent if she is in a position to competently treat it as a reason.[72] In turn, competence in treating considerations as reasons can be understood as a reliable disposition to be guided only by normative reasons, or by reference to the way in which some virtuous or ideal agent would be disposed to respond.[73] An agent that recklessly holds the belief that p against all available evidence will not be competently treating p as a reason (her holding that belief would manifest a very unreliable disposition to be guided only by reasons), so p would not be among her apparent reasons and would not make the agent's behavior rational. Indeed, insofar as the agent is in a position to treat competently her available evidence as providing reasons for some response, such evidence would be included among the agent's apparent reasons (even if she happens to disregard it). Thus, this approach is not affected by the problem mentioned in the previous paragraph.

I will not commit myself to any specific understanding of the notions of apparent reasons and competence. For my purposes here, it is enough that these notions make room for the distinction between being praiseworthy and doing what one has most reasons to do. More specifically, the framework I have sketched allows for the possibility of inappropriate responses that are not only blameless, but perhaps also deserve praise. In this way, one can capture the intuition that "rational" is used as a term of praise, without having to assume that being rational is a matter of conforming to one's reasons.

7. CONCLUSIONS

Let us take stock. Ascriptions of rationality typically involve a form of praise (and attribution of irrationality a form of criticism). This seems to motivate a normative account of rationality, according to which behaving rationally is something one has reasons to do. This account, however, is problematic. In particular, when combined with a factualist theory of reasons, it has difficulties with cases in which rational agents are guided by false considerations.

In the last section, I have suggested that we do not need such reasons-based

---

71   Whiting, "Keep Things in Perspective," 6–7.

72   Sylvan, "What Apparent Reasons Appear to Be."

73   See Sylvan, "What Apparent Reasons Appear to Be"; Williamson, "Justifications, Excuses, and Sceptical Scenarios."

accounts of rationality in order to explain the connection between ascriptions of rationality and praising and criticizing. One may deserve praise as a follower of reasons in doing something there was actually no good reasons to do, as long as one's performance manifests one's competence in responding to reasons. Relying on this idea, I have recommended a view of rationality according to which there is no reason to be rational, but you are rational if you try competently to follow your reasons. You are rational when you do your best to do what you ought to do.[74]

Instituto de Filosofia da Nova, Universidade Nova de Lisboa
Universidad Nacional de Educación a Distancia
jgonzalezdeprado@gmail.com

REFERENCES

Alvarez, Maria. *Kinds of Reasons: An Essay in the Philosophy of Action.* Oxford: Oxford University Press, 2010.

———. "Reasons for Action, Acting for Reasons, and Rationality." *Synthese* 195, no. 8 (August 2018): 3293–310.

Austin, John. "A Plea for Excuses." *Proceedings of the Aristotelian Society* 57 (1956): 1–30.

Baron, Marcia. "Justifications and Excuses." *Ohio State Journal of Criminal Law* 2, no. 2 (2005): 387–406.

Boult, Cameron. "Epistemic Normativity and the Justification-Excuse Distinction." *Synthese* 194, no. 10 (October 2017): 4065–81.

Broome, John. "Does Rationality Give Us Reasons?" *Philosophical Issues* 15, no. 1 (October 2005): 321–37.

———. "Is Rationality Normative?" *Disputatio* 2, no. 23 (November 2007): 1–18.

Capes, Justin A. "Blameworthiness without Wrongdoing." *Pacific Philosophical Quarterly* 93, no. 3 (September 2012): 417–37.

Chappell, Richard Yetter. "Fittingness: The Sole Normative Primitive." *Philosophical Quarterly* 62, no. 249 (October 2012): 684–704.

Christensen, David. "Disagreement, Drugs, etc.: From Accuracy to Akrasia." *Episteme* 13, no. 4 (December 2016): 397–422.

Coates, Allen. "Rational Epistemic Akrasia." *American Philosophical Quarterly* 49, no. 2 (April 2012): 113–24.

Cohen, Stewart. "Justification and Truth." *Philosophical Studies* 46, no. 3 (November 1984): 279–95.

Comesaña, Juan, and Matthew McGrath. "Having False Reasons." In *Epistemic Norms*, edited by Clayton Littlejohn and John Turri, 59–80. Oxford: Oxford University Press, 2014.

Dancy, Jonathan. *Ethics without Principles*. Oxford: Oxford University Press, 2004.

———. *Practical Reality*. Oxford: Oxford University Press, 2016.

Dorsch, Fabian, and Julien Dutant, eds. *The New Evil Demon.* Oxford: Oxford University Press, forthcoming.

Dutant, Julien. "Knowledge-First Evidentialism about Rationality." In Dorsch and Dutain, *The New Evil Demon*.

Gardner, John. *Offences and Defences*. Oxford: Oxford University Press, 2007.

Gerken, Mikkel. "Warrant and Action." *Synthese* 178, no. 3 (February 2011): 529–47.

Gibbons, John. "Things that Make Things Reasonable." *Philosophy and Phenomenological Research* 81, no. 2 (September 2010): 335–61.

Greenspan, Patricia. "Asymmetrical Practical Reasons." In *Experience and Analysis*, edited by Maria E. Reicher and Johann Ch. Marek, 387–94. Vienna: Austrian Ludwig Wittgenstein Society, 2005.

Hieronymi, Pamela. "The Force and Fairness of Blame." *Philosophical Perspectives* 18, no. 1 (December 2004): 115–48.

Hornsby, Jennifer. "A Disjunctive Conception of Acting for Reasons." In *Disjunctivism: Perception, Action, Knowledge*, edited by Fiona Macpherson and Adrian Haddock, 244–61. Oxford: Oxford University Press, 2008.

Horowitz, Sophie. "Epistemic Akrasia." *Noûs* 48, no. 4 (December 2014): 718–44.

Howard, Christopher. "The Fundamentality of Fit." In *Oxford Studies in Metaethics*, vol. 14, edited by Russ Shafer-Landau. Oxford: Oxford University Press, forthcoming.

Huemer, Michael. "Compassionate Phenomenal Conservatism." *Philosophy and Phenomenological Research* 74, no. 1 (January 2007): 30–55.

Hurka, Thomas. *Virtue, Vice, and Value*. Oxford: Oxford University Press, 2001.

Ichikawa, Jonathan Jenkins. "Internalism, Factivity, and Sufficient Reason." In *The Factive Turn in Epistemology*, edited by Veli Mitova, 66–83. Cambridge: Cambridge University Press, 2018.

Jackson, Frank. "Decision-Theoretic Consequentialism and the Nearest and Dearest Objection." *Ethics* 101, no. 3 (April 1991): 461–82.

Kelp, Christoph. "How to Be a Reliabilist." *Philosophy and Phenomenological Research*. Published ahead of print, September 11, 2017. https://doi.org/10.1111/phpr.12438.

Kiesewetter, Benjamin. "A Dilemma for Parfit's Conception of Normativity." *Analysis* 72, no. 3 (July 2012): 466–74.

———. *The Normativity of Rationality*. Oxford: Oxford University Press, 2017.

———. "'Ought' and the Perspective of the Agent." *Journal of Ethics and Social Philosophy* 5, no. 3 (October 2011): 1–24.

———. "You Ought to φ Only if You May Believe that You Ought to φ." *Philosophical Quarterly* 66, no. 265 (October 2016): 760–82.

Kolodny, Niko. "Why Be Rational?" *Mind* 114, no. 455 (July 2005): 509–63.

Kolodny, Niko, and John MacFarlane. "Ifs and Oughts." *Journal of Philosophy* 107, no. 3 (March 2010): 115–43.

Lackey, Jennifer. "Norms of Assertion." *Noûs* 41, no. 4 (December 2007): 594–626.

Lasonen-Aarnio, Maria. "Higher-Order Evidence and the Limits of Defeat." *Philosophy and Phenomenological Research* 88, no. 2 (March 2014): 314–45.

Littlejohn, Clayton. "Being More Realistic about Reasons." *Philosophy and Phenomenological Research*. Published ahead of print, May 20, 2018. https://doi.org/10.1111/phpr.12518.

———. *Justification and the Truth-Connection*. Cambridge: Cambridge University Press, 2012.

———. "A Plea for Epistemic Excuses." In Dorsch and Dutant, *The New Evil Demon*.

———. "Stop Making Sense? On a Puzzle about Rationality." *Philosophy and Phenomenological Research* 96, no. 2 (March 2018): 257–72.

———. "The Unity of Reason." In *Epistemic Norms*, edited by Clayton Littlejohn and John Turri, 135–54. Oxford: Oxford University Press, 2014.

Lord, Errol. "Defeating the Externalist's Demons." In Dorsch and Dutant, *The New Evil Demon*.

———. *The Importance of Being Rational*. Oxford: Oxford University Press, 2018.

———. "What You're Rationally Required to Do and What You Ought to Do (Are the Same Thing!)." *Mind* 126, no. 504 (October 2017): 1109–154.

Madison, Brent J. C. "On Justifications and Excuses." *Synthese*. Published ahead of print, May 18, 2017. https://doi.org/10.1007/s11229-017-1418-8.

McHugh, Conor, and Jonathan Way. "Against the Taking Condition." *Philosophical Issues* 26, no. 1 (October 2016): 314–31.

———. "Fittingness First." *Ethics* 126, no. 3 (April 2016): 575–606.

Miracchi, Lisa. "Competent Perspectives and the New Evil Demon Problem." In Dorsch and Dutant, *The New Evil Demon.*

Parfit, Derek. *On What Matters.* Oxford: Oxford University Press, 2011.

Raz, Joseph. *Practical Reason and Norms.* Oxford: Oxford University Press, 1975.

Scanlon, Thomas M. *What We Owe to Each Other.* Cambridge, MA: Harvard University Press, 1998.

Schroeder, Mark. "Getting Perspective on Objective Reasons." *Ethics* 128, no. 2 (January 2018): 289–319.

———. *Slaves of the Passions.* Oxford: Oxford University Press, 2007.

Schroeter, Laura, and François Schroeter. "Reasons as Right-Makers." *Philosophical Explorations* 12, no. 3 (July 2009): 279–96.

Sosa, Ernest. "How Competence Matters in Epistemology." *Philosophical Perspectives* 24, no. 1 (December 2010): 465–75.

———. *Judgment and Agency.* Oxford: Oxford University Press, 2015.

Southwood, Nicholas. "Vindicating the Normativity of Rationality." *Ethics* 119, no. 1 (October 2008): 9–30.

Srinivasan, Amia. "Normativity without Cartesian Privilege." *Philosophical Issues* 25, no. 1 (October 2015): 273–99.

Strawson, Peter F. "Freedom and Resentment." *Proceedings of the British Academy* 48 (1962): 1–25.

Svavarsdóttir, Sigrún. "Evaluations of Rationality." *Southern Journal of Philosophy* 41, no. S1 (Spring 2003): 121–36.

Sylvan, Kurt. "Veritism Unswamped." *Mind* 127, no. 506 (April 2018): 381–435.

———. "What Apparent Reasons Appear to Be." *Philosophical Studies* 172, no. 3 (March 2015): 587–606.

Vogelstein, Eric. "Subjective Reasons." *Ethical Theory and Moral Practice* 15, no. 2 (April 2012): 239–57.

Wallace, R. Jay. *Responsibility and the Moral Sentiments.* Cambridge, MA: Harvard University Press, 1994.

Way, Jonathan. "Reasons and Rationality." In *The Oxford Handbook of Reasons and Normativity*, edited by Daniel Star, 485–503. Oxford: Oxford University Press, 2018.

———. "Two Accounts of the Normativity of Rationality." *Journal of Ethics and Social Philosophy* 4, no. 2 (December 2009): 1–8.

Way, Jonathan, and Daniel Whiting. "Perspectivism and the Argument from Guidance." *Ethical Theory and Moral Practice* 20, no. 2 (April 2017): 361–74.

Weatherson, Brian. "Do Judgments Screen Evidence?" Unpublished manuscript, 2010.

Wedgwood, Ralph. "Justified Inference." *Synthese* 189, no. 2 (November 2012): 273–95.

———. "Rationality as a Virtue." *Analytic Philosophy* 55, no. 4 (December 2014): 319–38.

Whiting, Daniel. "Keep Things in Perspective: Reasons, Rationality and the *A Priori*." *Journal of Ethics and Social Philosophy* 8, no. 1 (March 2014): 1–22.

———. "Right in Some Respects: Reasons as Evidence." *Philosophical Studies* 175, no. 9 (September 2018): 2191–208.

Williams, Bernard. "Internal and External Reasons." In *Moral Luck*, 101–13. Cambridge: Cambridge University Press, 1981.

Williamson, Timothy. "Justifications, Excuses, and Sceptical Scenarios." In Dorsch and Dutant, *The New Evil Demon*.

———. *Knowledge and Its Limits*. Oxford: Oxford University Press, 2000.

Wodak, Daniel. "Can Objectivists Account for Subjective Reasons?" *Journal of Ethics and Social Philosophy* 12, no. 3 (December 2017): 259–79.

Worsnip, Alex. "The Conflict of Evidence and Coherence." *Philosophy and Phenomenological Research* 96, no. 1 (January 2018): 3–44.

Zimmerman, Michael J. *Living with Uncertainty: The Moral Significance of Ignorance.* Cambridge: Cambridge University Press, 2008.

———. "Moral Responsibility and Quality of Will." In *Responsibility: The Epistemic Condition*, edited by Philip Robichaud and Jan W. Wieland, 219–32. Oxford: Oxford University Press, 2017.

———. "Taking Luck Seriously." *Journal of Philosophy* 99, no. 11 (November 2002): 553–76.

EVOLUTION AND MORAL DISAGREEMENT

Michael Klenk

THE BURGEONING DEBATE about the metaethical implications of the Darwinist view of morality focuses on which epistemic principle(s) allegedly support debunking arguments against moral objectivism.[1] Moral 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.[2] Though objectivists must, of course, explain how objectivist moral beliefs can be justified in the first place, a central question is whether objectivist moral beliefs can be undercut, assuming that they are at least *prima facie* justified.[3]

So, what is that "something" in virtue of which a Darwinist view of morality creates a problem for objectivist moral beliefs? It has been claimed that evolutionary explanations of morality might show that moral beliefs are prone to error or fail to be modally secure, or that the best explanation of moral beliefs does not entail that they are (mostly) true.[4] None of these theses has found widespread support.

1 E.g., Clarke-Doane, "Justification and Explanation in Mathematics and Morality"; Klenk, "Survival of Defeat"; Sinclair, "Belief-Pills and the Possibility of Moral Epistemology"; Vavova, "Evolutionary Debunking of Moral Realism." The "Darwinist view of morality" is shorthand for "an evolutionary explanation of morality"; the view will be specified in section 2, below. Some debunkers take the argument to potentially undercut moral judgments that can be justified, e.g., Sinclair, "Belief-Pills and the Possibility of Moral Epistemology." If you accept this view, you can take the evolutionary challenge to have a wider scope, though I cannot address these further epistemological and metaphysical assumptions about morality in this paper.

2 E.g., Enoch, *Taking Morality Seriously*; Shafer-Landau, *Moral Realism*; Wielenberg, *Robust Ethics*. Objectivism is defended not only by robust realists but also by relaxed realists, such as Scanlon, *Being Realistic about Reasons*, and some naturalists, such as Jackson and Pettit, "Moral Functionalism and Moral Motivation." Arguably, they all face the evolutionary challenge; see Barkhausen, "Reductionist Moral Realism and the Contingency of Moral Evolution"; and Clarke-Doane, "Objectivity and Reliability."

3 See Wielenberg, *Robust Ethics*, and Klenk, review of *Robust Ethics*, for critical discussion.

4 Vavova, "Evolutionary Debunking of Moral Realism" and "Moral Disagreement and Moral

In light of this controversy, a new thesis is quickly gaining currency. A number of philosophers have argued that a Darwinist view of morality is metaethically significant because it shows that moral beliefs are counterfactually subject to disagreement.[5] This line of thought could also explain why the often-noted contingency of our moral beliefs is epistemically problematic.[6] Belief-contingency is problematic, on this view, *because* it reveals a problem with disagreement.[7] So, a Darwinist view of morality could yet play a metaethical role if it piggybacks on the epistemic significance of disagreement. For example, Mogensen writes that any metaethical implications that follow from a Darwinist view of morality "will be due to the epistemic significance of moral disagreement."[8] The disagreement in question is *hypothetical* or *counterfactual* disagreement: had our evolutionary history been different, our moral beliefs would conflict with our actual moral beliefs.[9] The consequence of this counterfactual moral disagreement is that the justification of all affected moral beliefs (objectively construed) is undercut, or so these philosophers argue. Let this be the disagreement view:

Skepticism"; Clarke-Doane, "Justification and Explanation in Mathematics and Morality"; Joyce, *The Evolution of Morality*; Lutz, "What Makes Evolution a Defeater?" See Klenk, "Survival of Defeat," ch. 2, for a discussion of the prospects of these arguments.

5   Bogardus, "Only All Naturalists Should Worry about Only One Evolutionary Debunking Argument"; Joyce, "Arguments from Moral Disagreement to Moral Skepticism"; Mogensen, "Contingency Anxiety and the Epistemology of Disagreement" and "Disagreements in Moral Intuition as Defeaters"; Sinclair, "Belief-Pills and the Possibility of Moral Epistemology"; Tersman, "Debunking and Disagreement"; White, "You Just Believe That Because…." Further suggestions about the relevance of disagreement for debunking, though not explicit endorsements, are provided by Street, "Objectivity and Truth," 314f; Clarke-Doane, "Justification and Explanation in Mathematics and Morality," 100; Ballantyne; "The Problem of Historical Variability," 246–54; and Horn, "Moral Realism, Fundamental Moral Disagreement, and Moral Reliability."

6   E.g., Handfield, "Genealogical Explanations of Chance and Morals"; Leiter, *The Future of Philosophy*; Lillehammer, "Methods of Ethics and the Descent of Man"; Sher, "But I Could Be Wrong."

7   Mogensen, "Contingency Anxiety and the Epistemology of Disagreement."; White, "You Just Believe That Because …."

8   Mogensen "Contingency Anxiety and the Epistemology of Disagreement," 591.

9   According to the ordinary understanding of disagreement, disagreement requires actual disputants and actual disputes. For example, one does not disagree about household chores if one's partner is merely lazy. Thus, on that understanding, whatever is implied by the evolutionary hypothesis seems far removed from disagreement. The relevant idea, however, is that some imaginary disagreements could easily be actual, in which case learning about them seems epistemically significant. I further specify the relevant type of disagreement in section 2.

Disagreement View: Evolutionary explanations of morality imply that there is justification-defeating counterfactual disagreement about all moral beliefs (as conceived of by moral objectivists).

The disagreement view rests on two important claims about the epistemology of disagreement. First, the proper response to genuine peer disagreement is to be concessive: the disputants epistemically ought to reduce confidence in the disputed belief. Second, the same holds for *hypothetical* peer disagreement. The concessive view is controversial, and so is the view that hypothetical disagreement is epistemically significant.[10] For this paper, however, I will assume that the concessive view is true. Another route to attacking the disagreement view would be to deny that hypothetical disagreement is epistemically significant. There are good reasons, however, not to place too much weight on the actual/possible distinction in arguments about disagreement.[11]

Instead, I aim to show that the disagreement view is false by focusing on the requirements of epistemic peerhood, a rather underexplored issue in recent epistemology and uncharted territory in relation to evolutionary debunking arguments in metaethics.[12] My strategy is to raise a dilemma for proponents of the disagreement view. They have to claim that evolution creates counterfactual moral disagreement in nearby or non-nearby scenarios. However, in non-actual nearby scenarios, there will not be disagreement about all moral beliefs. In remote scenarios, there will be disagreement, but not with peers. So, evolutionary explanations of morality do not reveal *epistemically significant disagreement* about all moral beliefs, and the disagreement view is false, or so I argue.

Moreover, if it is true that debunking arguments are epistemically significant if and only if they reveal epistemically significant disagreement, then the argument presented in this paper implies that evolutionary explanations of morality are epistemically insignificant.[13] Independently of that claim, this paper speaks to what we can and cannot learn about counterfactual moral disagreement from

10   On the concessive view see, e.g., Enoch, "Not Just a Truthometer." On epistemic significance, see Kelly, "The Epistemic Significance of Disagreement"; Tersman, "Moral Disagreement."

11   My main worry is that drawing the actual/possible distinction will depend on counterfactual analyses to explain when possible but absent disagreements are significant, and counterfactual analyses have a bad track record in philosophy.

12   King ("Disagreement") and Gelfert ("Who Is an Epistemic Peer?") are notable exceptions, though their arguments do not speak directly to the disagreement view.

13   On epistemically significant disagreement, cf. Bogardus, "Only All Naturalists Should Worry about Only One Evolutionary Debunking Argument," 656; White, "You Just Believe That Because . . . ." See Klenk, "Third Factor Explanations and Disagreement in Metaethics" and "Survival of Defeat," for extended discussions of the epistemic insignificance of evolutionary explanations of morality.

evolutionary considerations. These findings should be of interest to both moral objectivists and their critics. Section 1 clarifies the context and the metaethical significance of the disagreement view. Section 2 reconstructs the argument for the disagreement view in greater detail. Section 3 introduces my argument against the disagreement view and sections 4 and 5 defend the two horns of the dilemma of the disagreement view.

1. COUNTERFACTUAL DISAGREEMENT AND EVOLUTIONARY DEFEAT

Evolutionary explanations of morality maintain that the *capacity* for normative guidance or the *content* of at least some of our most fundamental moral beliefs is a product of human evolutionary history.[14] For example, bravery appears to be evolutionarily useful, and it is evaluated positively across a wide range of societies.[15] So it stands to reason that the positive (moral) evaluation of bravery has an evolutionary origin. Thus:

> *Evolutionary Hypothesis*: A significant proportion of human moral beliefs are the product of human evolutionary history.

For this paper, quite a few critical and controversial issues about the evolutionary hypothesis have to be swept under the rug. That is acceptable, however, because virtually all discussants in the metaethical debate accept two corollaries of the evolutionary hypothesis. First, the evolutionary determinants of our moral beliefs are *contingent*: had human evolutionary history been different, human moral beliefs would have been different. Second, objective moral truths were *causally* irrelevant in the evolutionary genesis of our moral beliefs.[16] The evolutionary hypothesis, with its two corollaries, provides the basis for so-called evolutionary debunking arguments.[17] This paper follows the proponents of the disagreement view and focuses on those variants of evolutionary debunking arguments that aim to conclude that all objective moral beliefs are unjustified.

---

14 Joyce, *The Evolution of Morality*; Street, "Objectivity and Truth" and "A Darwinian Dilemma for Realist Theories of Value."

15 Curry, Mullins, and Whitehouse, "Is It Good to Cooperate?; Curry, "Morality as Cooperation."

16 Enoch, "The Epistemological Challenge to Metanormative Realism"; FitzPatrick, "Debunking Evolutionary Debunking of Ethical Realism"; Street, "A Darwinian Dilemma for Realist Theories of Value."

17 Though the precise role of empirical information in debunking arguments is controversial; cf. Clarke-Doane, "Justification and Explanation in Mathematics and Morality."

2. CLARIFYING THE DISAGREEMENT VIEW

This section reconstructs the argument for the disagreement view. I clarify each premise as we go along. My focus will be on Mogensen's and Bogardus's defenses of the disagreement view. Some steps differ in the details, but they ultimately reach the same conclusion.

First, they argue that the evolutionary hypothesis implies that our counterfactual selves might have had different moral beliefs from us:

> Had our species evolved elsewhere—as easily might have happened—
> and we later formed moral beliefs using the same method we actually
> used, our beliefs may easily have been incompatible with our actual be-
> liefs, false by our own lights.[18]

There is reason to suppose that the moral intuitions of human beings reflect our place on the tree of life: had the conditions for the evolution of moral thought been realized in some distantly related species, their moral outlook would most likely incorporate certain fundamental differences in moral intuition, appropriate to their form of life.[19]

Both quotes reflect the idea that the evolutionary hypothesis implies the contingency of at least some of our moral beliefs. To make that idea more precise, let M_{actual} be the set of moral propositions whose members are the objects of our moral beliefs, where "our" refers to US, the set of all human beings who live or lived in the actual world. Let $M_{counterfactual}$ be the set of moral propositions believed by THEM, where THEM stands for the set of all human beings who live in some fixed counterfactual evolutionary scenario. According to the evolutionary hypothesis, $M_{counterfactual}$ could diverge significantly from M_{actual}.

Second, proponents of the disagreement view claim that the divergence of M_{actual} and $M_{counterfactual}$ amounts to hypothetical *disagreement* with our counterfactual selves. Compare Darwin's famous thought experiment:

> If men were reared under precisely the same conditions as hive-bees, there
> can hardly be a doubt that our unmarried females would, like the worker
> bees, think it a sacred duty to kill their brothers, and mothers would strive
> to kill their fertile daughters, and no one would think of interfering.[20]

---

18 Bogardus, "Only All Naturalists Should Worry about Only One Evolutionary Debunking Argument," 656.

19 Mogensen, "Contingency Anxiety and the Epistemology of Disagreement," 607.

20 Darwin, *The Descent of Man, and Selection in Relation to Sex*, 70.

Following Darwin's conjecture, the contents of $M_{counterfactual}$ might radically conflict with the members of M_{actual}.[21] Then we would find the moral beliefs of THEM "false by our own lights."[22] This constitutes hypothetical moral disagreement.

Third, proponents of the disagreement view claim that the hypothetical moral disagreement with our counterfactual selves is epistemically significant. We have the same type and quality of evidence for our moral beliefs, which, Bogardus claims, shows that there is an epistemically significant "evidential *symmetry*" between us and our counterfactual selves.[23] Mogensen, in contrast, takes the evolutionary hypothesis to show that we and our counterfactual selves have different evidence, which shows that there is an evidential *asymmetry* between us and our counterfactual selves. Such asymmetry is epistemically significant nonetheless because the moral disagreement implied by the evolutionary hypothesis bottoms out in pure conflicts of intuition.[24] Indeed, the claim is that hypothetical disagreement with our counterfactual selves is, all else being equal, as epistemically significant as actual disagreement. Bogardus qualifies this by saying that the hypothetical disagreement is "near enough to cause [epistemic] trouble" such that had we run a different evolutionary course, we would have *easily* ended up disagreeing with our counterfactual selves.[25] To paraphrase, hypothetical disagreement is relevant if it could be *easily present* (Bogardus) or if it is *arbitrarily absent* (Mogensen).

Fourth, the correct response to epistemically significant disagreement is to withhold judgment about the disputed belief. This claim is reminiscent of a *concessive* view about disagreement.[26] Concessive views imply that intractable disagreement about *p* among interlocutors of comparable epistemic standing undercuts their justification for endorsing or rejecting *p*, provided that there is no *independent* evidence for or against *p*.[27]

In conclusion, the evolutionary hypothesis implies that there is justifica-

---

21 Mogensen, "Contingency Anxiety and the Epistemology of Disagreement," 607.

22 Bogardus, "Only All Naturalists Should Worry about Only One Evolutionary Debunking Argument," 656.

23 Bogardus, "Only All Naturalists Should Worry about Only One Evolutionary Debunking Argument," 638, emphasis added.

24 Mogensen, "Disagreements in Moral Intuition as Defeaters," 286ff.

25 Bogardus, "Only All Naturalists Should Worry about Only One Evolutionary Debunking Argument," 657.

26 Bogardus, "Only All Naturalists Should Worry about Only One Evolutionary Debunking Argument," 656; Mogensen, "Contingency Anxiety and the Epistemology of Disagreement," 607.

27 Elga, "Reflection and Disagreement"; Feldman, "Epistemological Puzzles about Disagreement"; Sidgwick, *The Methods of Ethics*. As noted in the introduction, I will assume for the

tion-defeating disagreement about M_{actual}. Given a concessive view about disagreement, and in the absence of independent evidence in favor of M_{actual}, we should give up our belief in M_{actual}.[28]

Before turning to my argument against the disagreement view, two clarifications are in order. First, Bogardus and Mogensen do not specify the nature of disagreement; I suggest understanding disagreement about p as follows:[29]

> *Disagreement*: There is disagreement about p iff there exists a p such that
>   (a) S_1 believes that p and S_2 believes that $\sim p$, or S_1 believes that p and
> S_2 suspends judgment on whether p, or S_1 believes that $\sim p$ and S_2
>   suspends judgment on whether p.
>   (b) S_1 and S_2 have the same understanding of p.

Condition (a) is standard.[30] Condition (b) is sensible to preclude problems with merely apparent disagreement that turns out to be a sort of confusion of tongues.[31] Hence, when I write that there is disagreement about whether or not p, I mean that the disputants are referring to the same thing and that they are not merely talking past each other.

Second, the argument for the disagreement view is not explicitly presented as a case of peer disagreement. Nonetheless, it appeals to cases in which our counterfactual selves appear to be our epistemic peers in the minimal sense that their moral beliefs matter for the evaluation of our epistemic standing regarding morality. Bogardus emphasizes, as we have seen above, the evidential symmetry between us and our counterfactual selves. This affords the interpretation that he accepts what might be called a narrow conception of epistemic peerhood, which can be understood as follows:[32]

Peerhood—Narrow Conception: S_1 and S_2 are epistemic peers in regard to p iff S_1 and S_2 are equals regarding their evidential possession and their evidential processing with respect to p.[33]

Mogensen, in contrast, does not think that we and our counterfactual selves share equal moral evidence. Instead, he believes that we should treat the moral intuitions of our counterfactual selves as equally likely to be an excellent guide to the truth.[34] This sits very well with what might be called a broad conception of epistemic peerhood, which can be understood as follows:[35]

Peerhood—Broad Conception: S_1 and S_2 are epistemic peers in regard to p iff S_1 and S_2 are equally likely to be right about p.

Neither specification of peerhood is entirely satisfactory *as a specification of peerhood*. For example, even on a narrow conception, a full specification of peerhood would doubtlessly require further conditions, such as "similar openness to experience."[36] In the present context, however, my concern is not so much with an entirely accurate specification of the concept of peerhood, but instead with the fixation of our ideas about which interlocutors the proponents of the disagreement view consider to be epistemically relevant. As such, less strict criteria for epistemic peerhood benefit the proponents of the disagreement view, since it would be easier for them to show that there is peer disagreement *on either such conception*.[37]

With these clarifications in place, it is evident that the truth of the disagreement view depends on whether or not the evolutionary hypothesis implies either narrow or broad peer disagreement (or both).

3. THE ARGUMENT AGAINST THE DISAGREEMENT VIEW

So much for the argument in favor of the disagreement view; it is time to introduce my argument against it, which challenges the claim that the hypothetical

---

33 We can distinguish between acknowledged peer disagreement and non-acknowledged peer disagreement; see Kelly, "The Epistemic Significance of Disagreement," 168; King, "Disagreement," 261. In line with an internalist account of defeat, awareness of the disagreement is required to have an effect on justification. In the definition of peerhood, however, we can leave out this criterion.

34 Mogensen, "Disagreements in Moral Intuition as Defeaters," 294ff. Cf. Wedgwood, "The Moral Evil Demons," 241f.

35 E.g., Elga, "Reflection and Disagreement"; Vavova, "Moral Disagreement and Moral Skepticism," 308.

36 Cf. Gelfert, "Who Is an Epistemic Peer?"; King, "Disagreement."

37 See Gelfert, "Who Is an Epistemic Peer?" for problems with these accounts.

moral disagreement implied by the evolutionary hypothesis is epistemically significant (thus, proponents in the disagreement view go wrong in the third step of their argument):

P1. Hypothetical disagreement about p is epistemically significant only if the disputants are epistemic peers in regard to p, according to either a narrow or a broad conception of peerhood.[38]

P2. In non-actual nearby scenarios, the evolutionary hypothesis does not imply hypothetical disagreement about objectivist moral beliefs.

P3. In non-nearby scenarios, the evolutionary hypothesis does not imply hypothetical disagreement with epistemic peers, in the narrow or broad sense, about objectivist moral beliefs.

P4. So, the evolutionary hypothesis does not imply epistemically significant disagreement in either nearby or remote scenarios.

C. So, the evolutionary hypothesis does not imply epistemically significant disagreement.

The argument is deductively valid. P1 specifies two necessary conditions for the epistemic significance of disagreement. P2 and P3 state that neither non-actual nearby nor non-actual non-nearby scenarios exhibit disagreement about the objectivist moral beliefs (P2) or with our peers (P3); that means that there is no epistemically significant disagreement in either case. "Nearness" is a notoriously vague notion. I do not expect to offer a fully satisfactory account of it in this paper. For present purposes, nearby scenarios are those in which our counterfactual selves resemble the members of human societies on the ethnographic record (incidentally, this also implies closeness in time).[39] Non-nearby scenarios are those that depart in more or less extreme ways from the known ethnographic record. P4 is an intermediary conclusion that I will not discuss further. The argument's main conclusion implies that the disagreement view fails.

4. FIRST HORN OF THE DILEMMA:
NO RELEVANT DISAGREEMENT IN NEARBY SCENARIOS

I will first precisify P1 by showing why peerhood is an important criterion for assessing the epistemic significance of hypothetical disagreement. Then I will consider whether the evolutionary hypothesis implies hypothetical disagreement about the truth of the relevant moral beliefs to vindicate P2.

---

38 To wit, relevant for the disputant's epistemic justification for or against endorsing the belief that *p*.

39 See Curry, "Morality as Cooperation."

4.1. Relevant Interlocutors

P1 states, as a necessary restriction on the epistemic significance of hypothetical disagreement in general, that the hypothetical disagreement must be between epistemic peers.[40] The narrow conception of peerhood disjunctively connects with the broad conception, such that two thinkers are peers if they are equals regarding evidential possession or equally likely to get it right (or both). The need to limit the epistemic relevance of possible disagreement through a peerhood requirement is suggested by the potentially devastating effects of combining an uncurbed epistemic relevance of possible disagreement with a concessive view, as suggested in the following example.

Suppose that experts E_1 and E_2 are, before their encounter, defeasibly justified in believing p and $\sim p$, respectively. According to a simplistic version of the concessive view, E_1 and E_2 lose their justification for maintaining either belief once they learn of their disagreement. To maintain their belief, they have to appeal to independent evidence for or against p, or find independent evidence that suggests that their interlocutor is not their epistemic peer, to settle the question of whether or not p is true. Brushing aside thorny issues about the relevant sense of independence here, suppose that E_1 and E_2 do find independent evidence, q, about whether or not p. Usually, that would settle the disagreement. However, with the suggestion about the relevance of possible disagreement on the table, E_1 and E_2 cannot yet stop thinking about whether or not p, because it might be that E_3's belief that $\sim q$ could either be *easily present* or *arbitrarily absent.* In the absence of a reason to think that E_3's disagreement is too modally distant, E_1, E_2, and E_3 would, being diligent adherents of the concessive view, have to consider whether there is independent evidence about whether or not q or about E_3's epistemic status (while E_1 and E_2 remain agnostic about whether or not p), *ad infinitum.* So, on the face of it, a concessive view about disagreement paired with a view about the epistemic significance of possible disagreement leads to a vicious regress that leaves us unjustified in holding any belief at all.

So, lest general skepticism be embraced, the epistemic relevance of hypothetical disagreement must somehow be curbed. Peerhood among the interlocutors is a natural suggestion as a criterion for the epistemic relevance of a given disagreement. More pertinently, in the case of *possible* disagreement, there are countless hypothetical interlocutors, E_n, which might be relevant to the dispute existing between any two disputants E_1 and E_2. Limiting the set of relevant (hypothetical) interlocutors to those who are in equal evidential possession *or* antecedently

---

40 Naturally, proponents of the disagreement view should consider only disagreements that are plausibly implied by the evolutionary hypothesis.

equally likely to get things right would at least partly impede the potential regress that is made possible by invoking hypothetical disagreements. Hence, for hypothetical disagreement to be epistemically significant, it must be among peers.

4.2. Peer Agreement about the Relevant Moral Beliefs

In this section, I will narrow down the range of relevant beliefs that objectivists have to defend, which will, based on plausible evolutionary considerations, vindicate P2. According to proponents of the disagreement view, the evolutionary hypothesis must imply hypothetical disagreement about the contents of moral beliefs. However, objectivists need not defend *all* members of M_{actual} against the evolutionary challenge, and hardly any objectivist aims to do so.[41] This is because M_{actual} certainly does not contain only true and justified moral beliefs. It contains moral beliefs that reflect biases, conceptual errors, and other infelicities. It also contains highly specific beliefs that refer to idiosyncratic sociocultural factors, such as food taboos, which are often moralized.[42] Objectivists do not claim that all of these beliefs are justified.

Rather, defending the justification of some moral beliefs is enough to guard objectivism against the evolutionary challenge. In particular, objectivists defend the justification (and truth) of the following moral beliefs:

1. Survival and reproductive success … is at least somewhat good.[43]
2. Pleasure is usually good, and pain is usually bad.[44]
3. We have rights because we are reflective beings.[45]

These platitudes have a similar structure: some plausibly evolutionarily relevant natural property or event (e.g., being an instance of survival, being painful, being capable of self-reflection, etc.) is related to a moral property such as being good. The normative concept alluded to is always a thin moral concept: GOODNESS, BADNESS, or RIGHT (in the sense of "having a right," rather than being correct).

Let the set of moral platitudes be M_{basic}. M_{basic} is a proper subset of M_{actual}. I do not attempt to outline the contents of M_{actual}. It suffices to distinguish M_{basic} from M_{actual}. The members of M_{basic} are the moral platitudes that are universally accepted (by actual humans). Moral platitudes have two components. First, they are picked out by *thin* moral concepts. Thin moral concepts are evaluative concepts without descriptive content: RIGHT, GOOD, and OUGHT are standard

41 Cf. Shafer-Landau, *Moral Realism*, 17.
42 Cf. Henrich and Henrich, "The Evolution of Cultural Adaptations," 3717.
43 Enoch, "The Epistemological Challenge to Metanormative Realism," 430.
44 Skarsaune, "Darwin and Moral Realism," 232.
45 Wielenberg, "On the Evolutionary Debunking of Morality," 447.

examples. Second, moral platitudes latch onto the nonmoral facts that are evolutionarily relevant. The members of M_{basic} are thus the beliefs that combine thin moral concepts with evolutionarily relevant causal factors, such as pain, procreation, and survival.[46] Judging by the ethnographic record, every society accepts M_{basic}. For example, Curry, Mullins, and Whitehouse found that every (studied) society has moral rules about problem-centered domains of resource allocation, coordination to mutual advantage, exchange, and conflict resolution.[47] They also provide further evidence that people across societies evaluate positively behavior that represents optimal solutions to domain problems (e.g., all appreciate food sharing and bravery). These findings are supported by an overarching hypothesis that suggests that some morals differ across societies, but that there is a fundament of basic moral beliefs, indeed those that seem to belong to M_{basic}, that are held constant and shared across societies.[48]

This characterization of the relevant domain in terms of M_{basic} suggests that in *nonactual nearby scenarios*, we have good reason to suppose that the basic moral beliefs of our counterfactual selves will be like those of individuals in our society or other societies on the ethnographic record. Thus, given the ubiquity of beliefs in the platitudes of M_{actual}, it seems very likely that M_{basic} is a proper subset of $M_{counterfactual}$ too. That is, all the basic moral beliefs, whose contents are in M_{basic}, will also be endorsed by our counterfactual selves in nonactual nearby scenarios.[49] Thus, turning back the wheel of life only a tiny bit will show that there is agreement rather than disagreement about M_{basic}.

We have arrived at an important intermediary conclusion: if we confine ourselves to nearby possible scenarios, then the evolutionary hypothesis implies agreement about some moral beliefs that can be explained evolutionarily, rather than disagreement. While the evolutionary hypothesis might suggest disagreement about *some* moral beliefs in nearby possible scenarios, these disagreements are merely disagreements about the application of thick moral concepts, rather than disagreements about the members of M_{basic}.[50] As such, moral objectivists need not worry about these kinds of disagreement; P2 is thereby vindicated.

46 I take no stance on whether or not the members of M_{basic} stand in deductive or inferential relations to each other.

47 Curry, Mullins, and Whitehouse, "Morality as Cooperation."

48 Cf. Morris, *Foragers, Farmers, and Fossil Fuels.*

49 They need not be explicitly endorsed, as explicit representation is not necessarily required for believing something; cf. Harman, *Change in View*, 13ff. Thanks to an anonymous referee for prompting me to clarify this point.

50 Barkhausen, "Reductionist Moral Realism and the Contingency of Moral Evolution"; Morris, *Foragers, Farmers, and Fossil Fuels.*

However, proponents of the direct approach will probably be unimpressed by the lack of disagreement in nearby scenarios. They might argue that considering only nearby scenarios betrays a lack of imagination. Recall Darwin's thought experiment about the bees, which is supposed to illustrate that "we" could have ended up being very different organisms after all. In that case, it seems that the intersection between M_{actual} and $M_{counterfactual}$ will get smaller and smaller as we replay the tape of life until we arrive at counterfactual selves that do not agree about any member of M_{actual} and thus, by extension, any member of M_{basic}. In other words, P3 might still be false: when our counterfactual selves are like Darwin's human bees, $M_{counterfactual}$ will diverge radically from M_{basic}. So, the proponents of the disagreement view might claim that the evolutionary hypothesis will reveal epistemically significant disagreement in non-nearby scenarios. Let us follow them there.

5. SECOND HORN OF THE DILEMMA: NO DISAGREEMENT WITH MORAL PEERS

In *non-nearby* scenarios, the evolutionary hypothesis would be very likely to yield some disagreement about M_{basic} (that is, M_{basic} would not be a subset of $M_{counterfactual}$). However, to vindicate P3, I aim to show that any disagreement we may find in non-nearby scenarios is not *peer* disagreement.

5.1. Disagreement between Peers on a Narrow Conception

Let us consider *narrow* peer disagreement first. Recall that a narrow conception of peerhood says that two persons are peers if and only if they are in equal evidential possession and their processing of the evidence regarding moral issues is also equal. The crucial question is thus the following: Does the evolutionary hypothesis imply that our counterfactual selves in non-actual, non-nearby scenarios have the same evidence as we do and yet disagree with us about M_{basic}?

In order to assess the crucial question, we first have to take a brief detour back to the evolutionary hypothesis (which said that a significant proportion of human moral beliefs are the product of human evolutionary history). Proponents of the disagreement view must adopt a stringent interpretation of the evolutionary hypothesis, according to which there is a close connection between the ancestral environment, evolutionary forces, and the contents of moral intuitions and moral beliefs.[51] On this view, evolutionary processes influence the raw material based on which we form our moral beliefs to such a degree that if you change the evolutionary environment, you change the raw material and thereby the moral

---

51   Environment is to be widely understood to encompass sociocultural factors.

beliefs that our counterfactual selves will hold.[52] Of course, the details of this story may differ depending on the correct view about the bases of moral beliefs, but the result will be the same for proponents of the disagreement view: a sufficiently different evolutionary trajectory changes *whatever* determines our moral beliefs, and so our counterfactual selves end up with different moral beliefs.[53]

Proponents of the disagreement view need the stringent interpretation of the evolutionary hypothesis to fend off an objectivist objection: objectivists might claim that there is a subset of M_{basic} that is not subject to evolutionary contingency. That is, if moral beliefs are not determined by moral intuitions (and/or mental states with nonmoral content) that, in turn, are determined by evolutionary forces, then moral beliefs might reliably track moral facts after all.[54] For example, objectivists could argue that our beliefs about, say, the (*pro tanto*) badness of pain are not subject to evolutionary contingency and thus our counterfactual selves will *not* adopt diverging beliefs about the badness of pain.[55] The stringent interpretation will, in contrast, entail that our counterfactual selves will have different moral beliefs about the badness of pain *because* our counterfactual selves will have different moral intuitions and different moral intuitions *because* their sensory input is different on different evolutionary tracks.

Importantly, the stringent interpretation of the evolutionary hypothesis implies an inclusive notion of moral evidence (to wit, evidence for the moral truth). According to the inclusive notion of moral evidence, mental states with moral content (such as moral intuitions) and with nonmoral content (such as sensory input) can count as evidence for moral truth.[56] As we have seen, the stringent interpretation of the evolutionary hypothesis implies that both factors

52 Cf. Mogensen, "Contingency Anxiety and the Epistemology of Disagreement," 593.

53 For example, my point applies even when moral intuitions are themselves doxastic states (versus perceptual states) and even when moral beliefs are also based on mental states with nonmoral content (versus exclusively based on moral intuitions). For discussions of these views, see Climenhaga, "Intuitions Are Used as Evidence in Philosophy"; Chudnoff, "What Intuitions Are Like"; Bengson, "Grasping the Third Realm"; and Huemer, *Ethical Intuitionism*. Importantly, even if one thinks, like Bengson and Huemer, that moral beliefs are not based on intuitions as I understand them here, but instead on "direct perception," the stringent interpretation of the evolutionary hypothesis appears to be a threat to the form of moral objectivism they defend, too, as desired by proponents of the disagreement view.

54 FitzPatrick, "Debunking Evolutionary Debunking of Ethical Realism."

55 Objectivists might even use these beliefs, given their untarnished epistemic credentials in this scenario, to set up so-called third-factor accounts in order to vindicate other moral beliefs that are subject to evolutionary contingency; cf. Enoch, "The Epistemological Challenge to Metanormative Realism."

56 Strictly speaking, it is the fact that one has an intuition that may count as evidence, not the intuition, or seeming, itself.

will change on alternative evolutionary trajectories. Both factors are also usually considered to be relevant to the question of which moral beliefs it is rational for one to hold.[57] It is a controversial issue whether both factors or only moral intuitions count as evidence for moral beliefs, but the issue need not be settled here: as long as either factor counts as moral evidence, we will find different evidence on different evolutionary paths (given the stringent interpretation of the evolutionary hypothesis).[58] This means that, on a view of moral evidence charitable to proponents of the disagreement view, moral intuitions count as moral evidence.

I can now compare the input/output relations of our moral faculty with the input/output relations of our counterfactual selves.[59] The options are exhausted by four cases, where *Input* refers to the forces that shaped moral intuitions in THEM and US ($Input_{US}$ and $Input_{THEM}$, respectively) and *Output* refers to the set of moral beliefs (again with the subscript indicating whether they are OUR beliefs or THEIR beliefs):

1. $Input_{US} = Input_{THEM}$; $Output_{US} = Output_{THEM}$
2. $Input_{US} \neq Input_{THEM}$; $Output_{US} = Output_{THEM}$
3. $Input_{US} = Input_{THEM}$; $Output_{US} \neq Output_{THEM}$
4. $Input_{US} \neq Input_{THEM}$; $Output_{US} \neq Output_{THEM}$

Cases 1 and 2 signify agreement (since both outputs are identical) and are thus not relevant here. Cases 3 and 4 signify a divergence of M_{actual} and $M_{counterfactual}$

---

57 Wedgwood, "The Moral Evil Demons," 226.

58 This understanding of "evidence" is unorthodox insofar as it does not signify an epistemic support relation: not every determinant of a moral belief is also an epistemically *good* reason for that belief (for some subject); see Huemer, "The Problem of Defeasible Justification," 376. But such an inclusive notion is required for dialectical reasons. Adopting a restrictive interpretation would be uncharitable for proponents of the disagreement view. For example, the view of Williamson (*Knowledge and Its Limits*) would be restrictive in the present context because, assuming that knowledge requires truth, the evolutionary hypothesis could not, per se, imply peer disagreement about objective moral facts, narrowly construed. Either us or them would have evidence, but not both, and thus there would not be peer disagreement between THEM and US. Of course, this would be one way to argue against the disagreement view, though one that I do not pursue here mainly because an adequate discussion of a theory of evidence is beyond the scope of this paper. Thus, the more inclusive the notion of evidence, the more likely it is that the evolutionary hypothesis implies that there is disagreement with peers that share the same evidence (thus fulfilling the criteria for an epistemically relevant disagreement).

59 I do not place too much weight on the term "information" here. I wish mainly to exploit the thought of an input/output process whose relation between input and output is systematic, since this is what proponents of the evolutionary hypothesis claim; see Harms, *Information and Meaning in Evolutionary Processes*, for a relevant discussion.

and could be relevant for assessing the disagreement view. However, case 3 is not implied by the evolutionary hypothesis and case 4 is not relevant disagreement.

Consider case 3 first. It does indeed appear to be a relevant peer disagreement. Our counterfactual selves disagree about some beliefs in M_{basic}, and given that these beliefs are based on the same input, the narrow conception of peerhood tells us that we have a peer disagreement. However, either case 3 is *not* implied by the evolutionary hypothesis, or our counterfactual selves mean different things when they use moral terms. Consider the first option. Case 3 illustrates that subjects that base their moral beliefs on the very same input will generate differing moral beliefs. In other words, the output is not correlated with the input—a sign of a random process. However, the evolutionary hypothesis does not imply that our moral beliefs are the products of a random process. Indeed, as suggested above, the interpretation of the evolutionary hypothesis that proponents of the disagreement view require implies that moral beliefs are based on moral intuitions to such an extent that changing the moral intuitions would change the organism's moral beliefs.

To illustrate the first interpretation of case 3, suppose that our counterfactual selves live in a world exactly like ours in all nonmoral aspects. Given that they form their moral beliefs in the same way as we do, by relying on their intuition, there is no indication that their intuitions are any different in a world that is just the same as our world. The point is that disagreement is only a problem insofar as differences in output cannot be traced to differences in input.[60] Therefore, case 3 does not follow from the evolutionary hypothesis and thus it does not help the proponents of the direct approach. Of course, the assumption that moral intuitions shaped by evolutionary forces determine the content of moral beliefs is a stark idealization. If evolutionary forces fully determine the bases of moral beliefs, *pace* the evolutionary hypothesis, then organisms subject to the same evolutionary history might have different moral intuitions and correspondingly different moral beliefs. Note, however, that this line of reasoning is no help for proponents of the disagreement view. Pursuing the same thought about the disconnect between evolutionary influences and moral beliefs, objectivists can argue that truth-conducive methods such as reasoning or understanding can lead to true beliefs based on intuitions that are not influenced by evolutionary forces.[61]

Alternatively, when the outputs of our counterfactual selves differ from ours, even though they are based on the very same inputs, we have reason to suspect

60 Wright, *Truth and Objectivity*, 91ff.

61 FitzPatrick, "Debunking Evolutionary Debunking of Ethical Realism"; Huemer, "A Liberal Realist Answer to Debunking Skeptics."

that they are using moral terms differently from us. After all, based on the strin-
gent interpretation of the evolutionary hypothesis and the fact that they had the
same input, we should expect our counterfactual selves to agree with us about
moral matters. The best explanation of their ostensible disagreement would be
that we are not really disagreeing, but merely talking past each other.[62] Conse-
quently, we would not have a genuine disagreement at all (irrespective of wheth-
er it is between peers) on this interpretation of case 3.

Case 4 also shows a disagreement, and the evolutionary hypothesis plausibly
implies it. Suppose, for example, that our counterfactual selves live in a world
where their overall fitness is increased by sacrificing their children. They might
indeed be rather like Darwin's bees. Due to various evolutionary processes, they
might have different intuitions about how to treat their children from those we
have, and consequently, they will form moral beliefs whose contents conflict
with some of the members of M_{basic}. Thus, we certainly have a relevant disagree-
ment in case 4.

However, case 4 does not exhibit peer disagreement, narrowly construed,
because US and THEM do *not* have the same evidence. Our counterfactual selves
had different sensory inputs, and thus they have different moral intuitions: when
they consider, say, whether to sacrifice their children, they might feel a warm
glow of anticipation and a resounding positive feeling toward the thought—
quite unlike our moral intuitions about infanticide. Therefore, US and our coun-
terfactual selves will not be in equal evidential possession: case 4 is not peer
disagreement, narrowly construed.

This line of argument, which says that difference in intuitions prevents peo-
ple from counting as peers, might imply that paradigmatic cases of disagreement
do not count as peer disagreement either.[63] For example, two expert mathema-
ticians who disagree about the truth of Goldbach's conjecture, based on diverg-
ing mathematical intuitions, would not be having a peer disagreement. However,
insofar as this is a problem, it is only a problem for the narrow conception of
peerhood or for the view that intuitions constitute the sole evidence relevant
for beliefs about Goldbach's conjecture (or moral beliefs, in the analogous case).

---

62 This resembles a point made by Davidson (*Inquiries into Truth and Interpretation*) about
    radical interpretation. Davidson argues, roughly, that in cases of radical disagreement about
    a subject matter, the "principle of charity" demands that we should regard the other party as
    talking about a different subject altogether. Since we are concerned with merely conceivable
    disagreement, I suppose we can conceive that there is *no* talking at cross purposes going on
    and so we need not be charitable. Still, in agreement with Davidson, I believe that we should
    not take seriously the disagreement in this case.

63 I am grateful to an anonymous referee for raising this objection and for suggesting a possible
    solution.

However, the validity of the narrow conception is not at issue here (and section 6 aims to show that my argument holds if we adopt the alternative broad conception of peerhood). Moreover, case 4 would still not be a case of peer disagreement if we counted mental states with nonmoral content as moral evidence, too. Mental states with nonmoral content would *also* be different for our counterfactual selves, since this difference explains why they have different moral beliefs, and so we would end up with different evidence again. For example, if our counterfactual selves had different factual beliefs about the effects of, say, fratricide, their moral beliefs about fratricide would plausibly differ from ours, but then again, we would have different evidence and, again, case 4 would not be a case of narrowly construed peer disagreement.

Therefore, on a narrow conception of peerhood that is congenial to the proponents of the disagreement view, the evolutionary hypothesis does not imply relevant peer disagreement in non-nearby scenarios, which partly vindicates P3. Granted, however, the narrow conception of peerhood is not, though congenial to Bogardus's view, the most felicitous conception of peerhood for proponents of the disagreement view. Their argument could yet be saved if there were peer disagreement on the broad conception of peerhood. In the next section, we stay in non-nearby scenarios but consider whether any of our counterfactual selves are peers on a broad conception of peerhood.

5.2. Total Disagreement between Peers on a Broad Conception

On a broad conception of peerhood, you will recall, our counterfactual selves count as our peers insofar as they are equally likely to be right about moral matters. Since relevant disagreement is about M_{basic}, we can distinguish between two relevant cases for analysis: total disagreement and partial disagreement about M_{basic}. Neither case, however, creates a problem for moral objectivism.

Consider total disagreement about M_{basic} first. Total disagreement means that our counterfactual selves disagree about *all* beliefs in M_{basic}, which is a tremendously extreme situation.[64] Our counterfactual selves would not even agree that pain is *pro tanto* bad, that survival is good, that people are fundamentally equal,

---

64 This does not exclude the fact that our counterfactual selves agree about moral logical truths, such as "survival is either good or not good." But tautologies are either part of M_{basic} or they are not. If they are not, then the agreement is irrelevant for my argument. If they are, then we and our counterfactual selves still cannot validly infer points of agreement that lie within M_{basic} that are not tautologies, which suffices for my argument. My view is that platitudes based on thin moral concepts, but not tautologies, are part of M_{basic}. The reason is that the latter, but not the former, are evaluative and action guiding and thus useful to have in an evolutionary sense. Thanks to an anonymous referee for prompting me to clarify this point.

or that we should not cause unnecessary harm.[65] This is noteworthy because *at least some* agreement about some subject matter provides one with *prima facie* reasons to accept someone as a peer. For example, consider whether our counterfactual selves would be our peers if they were like the Neanderthals. Suppose we know nothing about our phylogenetic relatedness, their social habits, or their ventures into early forms of art. Despite ignorance on these matters, a good reason (not necessarily a sufficient reason) to maintain "default trust" in the moral intuitions of Neanderthals would be *some* agreement about M_{basic}.[66] Indeed, Neanderthals plausibly endorsed some of the members of M_{basic}: they believed that it is good to take care of one's family and community, they cherished survival, and they generally avoided pain.[67] Such agreement about M_{basic} would constitute common ground to accept Neanderthals as our moral peers, despite the 30,000 years that separate us from them. Would we have reason to withdraw the default trust we bestowed upon Neanderthals *in the absence of* agreement about M_{basic}? To aid our imaginations, let us conjure up some evolutionary path on which our counterfactual selves do not endorse any member of M_{basic} and let us call that species *Homo sapiens peregrinus*, the strange man.[68]

My claim is that total disagreement about M_{basic} with our Peregrinus-like counterfactual selves gives us reason to reject their default peerhood status (which they do possess, according to proponents of the disagreement view) and so we would not have peer disagreement on a broad conception of peerhood either. I will support my claim by looking at the reasons for granting and withdrawing default trust, which proponents of the disagreement view themselves accept.

---

65 When extreme cases of disagreement, in which disputants share little common ground, are considered in the literature on peer disagreement, even defenders of concessive views concede that disagreement might then lose its epistemic significance; cf. Elga, "Reflection and Disagreement," 495f; Kornblith, "Belief in the Face of Controversy," 50. Most discussions, however, focus on cases in which there is at least *some* agreement, as in disagreements with psychopaths, who disagree about many but not all moral beliefs; cf. Ballantyne, "The Problem of Historical Variability," 254; Sinnott-Armstrong, "Moral Disagreements with Psychopaths," 53. In such cases, Tersman's worry that "the mere fact that a person disagrees with us … cannot itself count as a shortcoming" seems apt (*Moral Disagreement*, 34ff). However, the extremity of rejecting M_{basic} altogether implies that reasons to take even "moral monsters" as our peers do not straightforwardly apply. While psychopaths, for example, recognize the method we use to arrive at moral beliefs and what constitutes good moral reasoning, they are simply unperturbed by it. Total disagreement with our counterfactual peers, in contrast, does not even allow agreement about what good methods of moral reasoning are.

66 Cf. Mogensen, "Disagreements in Moral Intuition as Defeaters," 283.

67 Cf. Wynn and Coolidge, *How to Think Like a Neandertal*, 19–21.

68 It is doubtful that our Peregrinus selves would be plausible products of an evolutionary process. I will return to this biological objection against the disagreement view in section 5.3.

According to proponents of the disagreement view, our counterfactual selves enjoy default trust because of a fundamental symmetry between THEM and US. Insofar as we can (defeasibly) trust our own moral beliefs without independent vindication, we must extend this trust to others.[69] Let us accept this controversial claim for the sake of argument.[70] Second, proponents of the disagreement view claim that default trust is defeasible if considerations *independent* of the disputed proposition let us assign a higher likelihood of us being right about the disputed proposition compared with the disputing interlocutor.[71] For example, suppose that you judge that torture is morally impermissible, but you learn that within a week you will judge that torture is morally permissible. In light of this disagreement with your future self, is your current belief about the impermissibility of torture defeated?[72] No, because you can justifiably reject your future self's peer status based on the privileged access you have to your reasons for thinking that torture is impermissible.[73] Hence, default trust is defeasible if we have reason to think that we are more likely to get things right compared to our peers. For example, given a disagreement, we might know we were, say, not intoxicated when thinking about the disputed issue, but lack such knowledge about our peers. Such cases, where we have *no* or *little* information about our peers, but a lot of (introspective) information about ourselves, are highly relevant for defeating default trust.

It follows from this view of defeating default trust that if we *only* know that our counterfactual selves totally disagree about the truth of the contents of M_{basic}, then our counterfactual selves will lose their status as peers. The crucial question is therefore whether we know *anything* about our Peregrinus-like counterfactual selves, apart from the fact that we disagree, that gives us reason to maintain our trust (our trust, that is, that our counterfactual selves are as likely as us to get moral matters right). In other words, given that (a) we totally disagree about M_{basic} and (b) we know X about our counterfactual selves, are we equally likely to get moral matters right?[74] If nothing can replace X, in combination with the fact that we totally disagree about M_{basic}, suggests that they and we are equally

---

69   Cf. Foley, *Intellectual Trust in Oneself and Others*; Mogensen, "Disagreements in Moral Intuition as Defeaters."

70   See Enoch, "Not Just a Truthometer," 962ff, for critical discussion.

71   Mogensen, "Disagreements in Moral Intuition as Defeaters," 294ff.

72   Wedgwood, "The Moral Evil Demons," 241.

73   Mogensen, "Disagreements in Moral Intuition as Defeaters," 294–95.

74   Again, this is in light of the assumptions of the evolutionary challenge: our moral beliefs are *prima facie* justified, so the methods we are currently using to arrive at M_{basic} are by and large accurate.

likely to get moral matters right, then our Peregrinus-like counterfactual selves do not count as our peers.

My aim is the modest one of showing that two plausible factors for maintaining trust in our counterfactual selves, two factors to "fill in" X, turn out to be mistaken. If that is right, then we have strong reason to judge that cases of total disagreement about M_{basic} are not cases of peer disagreement, which further vindicates P3.[75] We could look, first, to Peregrinus's *cognitive capacities including nonmoral beliefs*, or, second, to their *belief-forming methods*. Neither aspect, however, is convincing. Consider cognitive capacities first.

> X_1 = *Cognitive capacities including nonmoral beliefs.*[76] Our peers are those who, in general, reason as well as us. They are as good as we are in obtaining scientific knowledge. They compose logical proofs, understand physics, and perform as well in standardized intelligence tests as average humans.

Cognitive development might be a necessary condition for peerhood in moral matters, but it is not a sufficient condition. Cognitive development of a certain level might function as a kind of "enabler" for making correct moral judgments. For example, if Peregrinus lacked a theory of mind, similar to very young children, he would be prone to making egoistic decisions and would lack the ability to recognize that other beings have their own plans and wishes.[77] It might thus be thought that cognitive abilities alone provide reason for expecting Peregrinus to be a good moral reasoner. However, the opposite is true. Just because a specimen of Peregrinus can realize that you would be hurt by something he does, this does not imply that he will respect that consideration.

Moreover, people can be experts in one area but still be (systematically) wrong in another area, and it is generally the case that assessments of peerhood seem domain specific.[78] Otherwise, it would make good sense to ask expert chess players to sit on ethics committees and top-notch nuclear physicists to weigh in on Europe's border policies just *because* their cognitive abilities are taken as evidence of their ethical expertise. More generally, an ability to get moral matters right does not seem to be directly inferable from competence in nonmoral mat-

---

75 I say that my goal is modest because I do not provide a conclusive case; there might be other reasons for maintaining trust in our counterfactual selves that I cannot (for lack of space) address here. However, I cannot think of more, nor have proponents of the disagreement view provided such reasons yet.

76 Capacities are understood here as having the ability to function on a certain level.

77 Kohlberg and Hersh, "Moral Development."

78 Goldberg, "Disagreement, Defeat, and Assertion," 169; Weatherson, "Disagreements, Philosophical and Otherwise," 56.

ters, otherwise we could argue, for example, that our ability to ascertain mathematical truth indicates that we are good at grasping moral truth. These proposals do not look promising, and thus the mere fact that we have shared cognitive capacities with Peregrinus (in the absence of any agreement about M_{basic}) gives us no reason to think that they are as likely as us to get moral matters right.[79]

Consider the following alternative instead:

X_2 = *Similarity of belief-forming methods.* Our peers are those who use similar methods of belief formation. Peregrinus is as good as we are at obtaining knowledge about nonnormative matters. Peregrinus also relies on his intuition in forming moral beliefs, and so do we.

The proposal might seem promising, but it just pushes back the crucial question.[80] We wanted to know whether we have reason to believe that Peregrinus is as likely as us to get moral matters right (given total disagreement about M_{basic}), but now we need to know whether we are using the same method of belief formation. However, how can we know what method Peregrinus is using, when the outputs of whatever method he is using are completely different from the outputs of our method?

The problem is one of individuating methods, and there is a metaphysical and an epistemic dimension to it. Metaphysically, the question concerns which method is being used; epistemically, the question concerns which method we have sufficient reason to believe is being used. Consider the metaphysical question first. We can suppose that methods are partly individuated by input, given

79 Cf. Klenk, "Can Moral Realists Deflect Defeat Due to Evolutionary Explanations of Morality?" for an objection to the view that we can use our reliability in a nonmoral domain to vindicate our moral reliability.

80 There is reason to doubt that similarity of belief-forming methods is a good criterion to use to judge whether others are likely to get things right in the first place, though I will not, for dialectical reasons, build on this argument here. The reason is that the proposal presupposes a sound approach for individuating methods (effectively an answer to the generality problem) and no current approach has found widespread support; cf. Bishop, "Why the Generality Problem Is Everybody's Problem," 285. Any description of a method, M, that we are using (e.g., perception, statistical inference) might also fit the description of the methods used by a class of interlocutors that we do not normally regard as peers, such as hallucinating people (in the perception case) or depressed economists (when it comes to making accurate predictions about the economy). According to the present proposal, we would have to regard them as peers *or* find a better principled way to describe the method. But given the problems with individuating methods, this seems unpromising, so we would have to, counterintuitively, accept that they are peers. This worry should be kept in mind as an additional problem for the proposal that similarity of belief-forming methods is a good criterion for peerhood, though I aim to show that the proposal does not help the disagreement view even if it is *prima facie* acceptable.

that individuating methods externally is a widely accepted approach to method individuation among epistemologists.[81] On this view, the causes of one's beliefs (partly) determine the method one is using. For example, a loving mother who assesses the piano skills of her child is not using an objective method, even though she might think she is, when her belief that her child is a prodigy is caused by her love, not by an assessment of her child's skills. However, if inputs partly individuate methods, then given that Peregrinus will have had, according to the evolutionary hypothesis, different input, they will have used different methods. The sensory input and/or the moral intuitions that cause their moral beliefs will be different from ours, and so they will have used a different method. Thus, even though similarity of belief-forming method might be relevant for assessing our counterfactual selves' accuracy, we get the result that Peregrinus is not using the same method. Hence, we have reason to recant our counterfactual selves' default status as peers.

Next consider the epistemic question of how to individuate methods. In this case, it is sensible to suppose that methods are partly individuated by outputs (which is, incidentally, also congenial to individuating methods externally). For example, your friend's genuine belief that you owe $444 for the lunch bill (for a meal that was almost certainly below $100), gives you an abductive reason for thinking that your friend is not using the same method (e.g., arithmetic) that you are to calculate how the bill is shared in the first place. After all, compared with the hypothesis that your generally reliable friend made a mistake, a better explanation is that she is just trying to be funny. More generally, since methods of belief formation ought to be accurate, we should expect that competently using the method yields comparable outputs—a kind of positional objectivity.[82] In the absence of any reason to assume incompetence or insincerity, widely diverging outputs are thus reason to believe that another method was used to form the judgment. Going back to our disagreement with Peregrinus, and the fact that our outputs differ completely, we plausibly get a positive reason to believe that we are *not* using the same method. At best, we have learned nothing that would warrant maintaining trust in THEM (because, at best, we should withhold judgment about whether they are using the same method). In any case, total disagreement about M_{basic} coupled with the, at best, agnostic attitude about whether they are using the same method of belief formation gives us reason to assign less likelihood to our counterfactual selves being accurate in moral matters compared to us. Again, we would have reason to withdraw their peerhood status.

---

81 E.g., Pritchard, *Epistemic Luck*, 152ff. Thanks to an anonymous referee for suggesting this point.

82 Cf. Sen, "Positional Objectivity."

Therefore, plausible alternatives to *agreement* about M_{basic} suggest that there are *no* good reasons to take our counterfactual selves as our peers on a broad conception of peerhood if there is total disagreement about M_{basic}. I have not considered all possible alternatives. However, there are good reasons to think that none will be successful. The crucial point is that we have to presuppose some standard by which we can ascertain what it means to get moral matters right.[83] A necessary component of such a standard seems to be (partial) agreement about M_{basic}. Recall that, before any disagreement, we take ourselves to have reasonable grounds to think that the members of M_{basic} are largely true. Thus, we have reasonable grounds to believe that we are getting moral matters right, insofar as we believe in M_{basic}. Our counterfactual selves do not. So, if the evolutionary hypothesis implies total disagreement about M_{basic}, then objectivists need not be concerned.

In the next section, I consider and rebut a final option on behalf of the disagreement view to close my case for P3: *partial* disagreement about M_{basic}.

5.3. Partial Disagreement between Peers on a Broad Conception

Still staying in non-nearby scenarios, the much more plausible case is that we rewind the wheel of life, but only to a point where there is still *some* agreement about M_{basic}. Let partial disagreement be a case in which our counterfactual selves agree about at least one belief that is a member of M_{basic}. We might, therefore, have reason (though perhaps not sufficient reason) to count them as our peers on a broad conception of peerhood. Of course, there are fuzzy boundaries, and I do not suspect that we can say with precision whether agreement about some percentage of the members of M_{basic} is required for peerhood. However, there could be enough agreement to raise the suspicion that "there is no reason to suppose that either party to the dispute is in an evidentially superior position."[84] So, debunking explanations could reveal *local* disagreement with our counterfactual selves. This might concern only some members of M_{basic}. However, this line of argument does not vindicate the disagreement view for two reasons.

First, if we consider just one counterfactual scenario, in which we end up like, say, Peregrinus, then the most that proponents of the disagreement view could conclude is that the justification of *some* beliefs is challenged. Such a case would not show, however, that all objectivist moral beliefs in M_{basic} are defeated.

For example, it might be that we cannot determine whether it is morally permissible or impermissible to abort fetuses. However, this finding does not imply that all the other moral beliefs in M_{basic} about which there is agreement are also

83 See Elga, "Reflection and Disagreement," 493ff.
84 Cohen, "A Defense of the (Almost) Equal Weight View," 98.

unreliable. To reach that conclusion, proponents of the direct approach would have to appeal to a principle of the following sort:

> *Token-Type*: If there is peer disagreement about a token of a type of proposition, K, then beliefs about that type of proposition are epistemically suspect.

However, the Token-Type principle is undoubtedly false. There may be radical disagreements about matters in physics, but we do not judge all beliefs about physics to be unjustified. Instead, it seems appropriate to judge that the question is beyond our (current) abilities to answer. Objectivists can adopt the same reasoning. There might be peer disagreement about the truth of the contents of some members of M_{basic}—and we might want to suspend judgment about those—but that need not compel us to suspend judgment about all beliefs in M_{basic}.[85]

Second, proponents of the disagreement view might argue as follows: if we consider manifold disagreements with manifold counterfactual selves, we could get cumulative total disagreement about M_{basic}. To illustrate, assume that M_{basic} contains two nonoverlapping proper subsets: A and B. We agree with Peregrinus about A and disagree about B. Now imagine that there is another of our counterfactual species, say *Homo sapiens cerritulus*, the mad man. We agree with Cerritulus about B but disagree about A. As a result, there is peer disagreement about all beliefs in M_{basic}, albeit not with the same interlocutor.

However, that response is only initially plausible because it is unlikely that the evolutionary hypothesis implies that such a situation is possible. For one, the contents of the beliefs in M_{basic} are ecologically related in worlds that are similar to ours.[86] If debunking explanations imply relevant disagreement about M_{basic} with, say, Cerritulus, then Cerritulus's world would be very different from ours. Thus, it would be unlikely that Cerritulus would agree about the beliefs contained in set B. In other words, disagreement about some beliefs in M_{basic} raises the probability of disagreement about other beliefs in M_{basic}, such that it is unlikely that there could be a cumulative disagreement about *all* beliefs in M_{basic}. Moreover, given that mere agreement about bits of M_{basic} can be considered a necessary but not a sufficient condition for peerhood, it is not clear, and is certainly not established on the broad conception of peerhood, that imagining many deviant species with whom we have partial agreement establishes that there is relevant peer disagreement.

These considerations suggest that partial disagreement about M_{basic} is plausi-

---

85   This might imply that there are some moral propositions that are unknowable on an objectivist account of morality (Wright, *Truth and Objectivity*).

86   See Morris, *Foragers, Farmers, and Fossil Fuels*, for an extended argument along these lines.

ble to some extent, but that it does not yield the desired conclusion that all beliefs in M_{basic} are subject to justification-defeating disagreement. This concludes the case for premise P3. As I said, it is not a conclusive case. In particular, there might be other criteria based on which we could take our counterfactual selves to be our peers despite them disagreeing about all members of M_{basic}. Moreover, it might be possible to conjure up scenarios that are biologically possible in which there is a triangulated total disagreement about M_{basic}. However, proponents of the disagreement view have not made that case. As such, the considerations of this section suggest that, no matter how the tape of life is replayed, we will not find relevant, justification-defeating disagreement about objectivist moral beliefs, so moral objectivism has not been refuted by the disagreement view.

6. CONCLUSION

In conclusion, the evolutionary hypothesis does not show that there is relevant hypothetical disagreement that defeats the justification of all our objectivist moral beliefs. As a result, the disagreement view fails. If all alternative interpretations of the epistemic significance of the evolutionary hypothesis fail, as some proponents of the disagreement view claim, then evolutionary debunking arguments fail to have skeptical consequences for moral objectivists.

However, even without assuming the radical claim that the evolutionary hypothesis is relevant only if it implies hypothetical moral peer disagreement, this paper shows that appeals to disagreement do not help the debunker's case, which takes away one possible route for debunkers to press their skeptical conclusion. For defenders of moral objectivity, this means relief on one front. For their opponents, this means that they need to reinforce efforts to find another epistemic phenomenon to undergird evolutionary debunking arguments against the objectivity of morality. In the end, there is strong reason to believe that hypothetical disagreement, no matter how far we rewind the tape of life, does not help the debunker in the case against moral objectivism.[87]

Delft University of Technology
m.b.o.t.klenk@tudelft.nl

REFERENCES

Ballantyne, Nathan. "The Problem of Historical Variability." In *Disagreement and Skepticism*, edited by Diego E. Machuca, 239–58. New York: Routledge, 2013.

Barkhausen, Max. "Reductionist Moral Realism and the Contingency of Moral Evolution." *Ethics* 126, no. 3 (April 2016): 662–89.

Bengson, John. "Grasping the Third Realm." In *Oxford Studies in Epistemology*, vol. 5, edited by Tamar Szabo Gendler and John Hawthorne, 1–38. Oxford: Oxford University Press, 2015.

Bishop, Michael A. "Why the Generality Problem Is Everybody's Problem." *Philosophical Studies* 151, no. 2 (November 2010): 285–98.

Bogardus, Tomas. "Only All Naturalists Should Worry about Only One Evolutionary Debunking Argument." *Ethics* 126, no. 3 (April 2016): 636–61.

Christensen, David, and Jennifer Lackey, eds. *The Epistemology of Disagreement: New Essays.* Oxford: Oxford University Press, 2013

Chudnoff, Elijah. "What Intuitions Are Like." *Philosophy and Phenomenological Research* 82, no. 3 (May 2011): 625–54.

Clarke-Doane, Justin. "Justification and Explanation in Mathematics and Morality." In *Oxford Studies in Metaethics*, vol. 10, edited by Russ Shafer-Landau, 80–103. Oxford: Oxford University Press, 2015.

———. "Objectivity and Reliability." *Canadian Journal of Philosophy* 47, no. 6 (2017): 841–55.

Climenhaga, Nevin. "Intuitions Are Used as Evidence in Philosophy." *Mind* 127, no. 505 (January 2018): 69–104.

Cohen, Stewart. "A Defense of the (Almost) Equal Weight View." In Christensen and Lackey, *The Epistemology of Disagreement*, 98–117.

Curry, Oliver Scott. "Morality as Cooperation. A Problem-Centred Approach." In *The Evolution of Morality*, edited by Todd K. Shackelford and Ranald D. Hansen, 27–51. Dordrecht: Springer, 2016.

Curry, Oliver Scott, Daniel Austin Mullins, and Harvey Whitehouse. "Is It Good to Cooperate? Testing the Theory of Morality-as-Cooperation in 60 Societies." *Current Anthropology* (forthcoming).

Darwin, Charles. *The Descent of Man, and Selection in Relation to Sex.* 1871. Edited by Adrian Desmond and James R. Moore. London: Penguin Books, 2004.

Davidson, Donald. *Inquiries into Truth and Interpretation.* Oxford: Oxford University Press, 1984.

Elga, Adam. "Reflection and Disagreement." *Noûs* 41, no. 3 (September 2007): 478–502.

Enoch, David. "The Epistemological Challenge to Metanormative Realism:

How Best to Understand It, and How to Cope with It." *Philosophical Studies* 148, no. 3 (April 2010): 413–38.

———. "Not Just a Truthometer: Taking Oneself Seriously (but Not Too Seriously) in Cases of Peer Disagreement." *Mind* 119, no. 476 (October 2011): 953–97.

———. *Taking Morality Seriously: A Defense of Robust Realism.* Oxford: Oxford University Press, 2011.

Feldman, Richard. "Epistemological Puzzles about Disagreement." In *Epistemology Futures*, edited by Stephen C. Hetherington, 216–36. Oxford: Clarendon Press, 2006.

Feldman, Richard, and Ted A. Warfield, eds. *Disagreement.* Oxford: Oxford University Press, 2010.

FitzPatrick, William J. "Debunking Evolutionary Debunking of Ethical Realism." *Philosophical Studies* 172, no. 4 (April 2015): 883–904.

Foley, Richard. *Intellectual Trust in Oneself and Others.* Cambridge: Cambridge University Press, 2001.

Gelfert, Axel. "Who Is an Epistemic Peer?" *Logos and Episteme* 2, no. 4 (2011): 507–14.

Goldberg, Sanford C. "Disagreement, Defeat, and Assertion." In Christensen and Lackey, *The Epistemology of Disagreement*, 167–89.

Handfield, Toby. "Genealogical Explanations of Chance and Morals." In *Explanation in Ethics and Mathematics*, edited by Uri D. Leibowitz and Neil Sinclair, 58–82. Oxford: Oxford University Press, 2016.

Harman, Gilbert. *Change in View: Principles of Reasoning.* Cambridge, MA: MIT Press, 1988.

Harms, William F. *Information and Meaning in Evolutionary Processes.* Cambridge: Cambridge University Press, 2004.

Henrich, Joseph, and Natalie Henrich. "The Evolution of Cultural Adaptations: Fijian Food Taboos Protect against Dangerous Marine Toxins." *Proceedings of the Royal Society of London B: Biological Sciences* 277, no. 1701 (2010): 3715–24.

Horn, Justin. "Moral Realism, Fundamental Moral Disagreement, and Moral Reliability." *Journal of Value Inquiry* 51, no. 3 (September 2017): 363–81.

Huemer, Michael. *Ethical Intuitionism.* Basingstoke: Palgrave Macmillan, 2005.

———. "A Liberal Realist Answer to Debunking Skeptics: The Empirical Case for Realism." *Philosophical Studies* 173, no. 7 (July 2016): 1983–2010.

———. "The Problem of Defeasible Justification." *Erkenntnis* 54, no. 3 (May 2001): 375–97.

Jackson, Frank, and Philip Pettit. "Moral Functionalism and Moral Motivation." *Philosophical Quarterly* 45, no. 178 (January 1995): 20–40.

Joyce, Richard. "Arguments from Moral Disagreement to Moral Skepticism." In *Moral Skepticism: New Essays*, edited by Diego E. Machuca, 141–62. New York: Routledge, 2018.

———. *The Evolution of Morality*. Cambridge, MA: MIT Press, 2006.

Kelly, Tom. "The Epistemic Significance of Disagreement." In *Oxford Studies in Epistemology*, vol. 1, edited by Tamar S. Gendler and John P. Hawthorne, 167–96. Oxford: Clarendon Press, 2005.

King, Nathan L. "Disagreement: What's the Problem? or A Good Peer Is Hard to Find." *Philosophy and Phenomenological Research* 85, no. 2 (September 2012): 249–72.

Klenk, Michael. "Can Moral Realists Deflect Defeat Due to Evolutionary Explanations of Morality?" *Pacific Philosophical Quarterly* 98, no. 19 (December 2017): 227–48.

———. Review of *Robust Ethics: The Metaphysics and Epistemology of Godless Normative Realism*, by Erik Wielenberg. *Dialectica* 70, no. 3 (September 2016): 482–88.

———. "Survival of Defeat: Evolution, Moral Objectivity, and Undercutting." PhD Diss., Utrecht University, 2018.

———. "Third Factor Explanations and Disagreement in Metaethics." *Synthese.* Published ahead of print, July 12, 2018. https://link.springer.com/article/10.1007/s11229-018-1875-8.

Kohlberg, Lawrence, and Richard H. Hersh. "Moral Development: A Review of the Theory." *Theory Into Practice* 16, no. 2 (April 1977): 53–59.

Kölbel, Max. "Faultless Disagreement." *Proceedings of the Aristotelian Society* 104 (2004): 53–73.

Kornblith, Hilary. "Belief in the Face of Controversy." In Feldman and Warfield, *Disagreement*, 29–52.

Leiter, Brian, ed. *The Future of Philosophy*. Oxford: Clarendon Press, 2004.

Lillehammer, Hallvard. "Methods of Ethics and the Descent of Man: Darwin and Sidgwick on Ethics and Evolution." *Biology and Philosophy* 25, no. 3 (June 2010): 361–78.

Lutz, Matt. "What Makes Evolution a Defeater?" *Erkenntnis.* Published ahead of print, August 7, 2017. https://doi.org/10.1007/s10670-017-9931-1.

Matheson, Jonathan. *The Epistemic Significance of Disagreement*. Basingstoke: Palgrave Macmillan, 2015.

Mogensen, Andreas L. "Contingency Anxiety and the Epistemology of Disagreement." *Pacific Philosophical Quarterly* 97, no. 4 (December 2016): 590–611.

———. "Disagreements in Moral Intuition as Defeaters." *Philosophical Quarterly* 67, no. 267 (April 2017): 282–302.

Morris, Ian. *Foragers, Farmers, and Fossil Fuels*: *How Human Values Evolve*. Princeton: Princeton University Press, 2015.

Pritchard, Duncan. *Epistemic Luck*. Oxford: Oxford University Press, 2005.

Scanlon, T. M. *Being Realistic about Reasons*. Oxford: Oxford University Press, 2014.

Sen, Amartya. "Positional Objectivity." *Philosophy and Public Affairs* 22, no. 2 (Spring 1993): 126–45.

Shafer-Landau, Russ. *Moral Realism*: *A Defence*. Oxford: Clarendon Press, 2003.

Sher, George. "But I Could Be Wrong." *Social Philosophy and Policy* 18, no. 2 (Summer 2001): 64.

Sidgwick, Henry. *The Methods of Ethics*. 7th ed. 1874. Indianapolis, IN: Hackett Publishing, 1981.

Sinclair, Neil. "Belief-Pills and the Possibility of Moral Epistemology." In *Oxford Studies in Metaethics*, vol. 13, edited by Russ Shafer-Landau, 98–122. Oxford: Oxford University Press, 2018.

Sinnott-Armstrong, Walter. "Moral Disagreements with Psychopaths." In *Challenges to Moral and Religious Belief*: *Disagreement and Evolution*, edited by Michael Bergmann and Patrick Kain, 40–60. Oxford: Oxford University Press, 2014.

Skarsaune, Knut O. "Darwin and Moral Realism: Survival of the Iffiest." *Philosophical Studies* 152, no. 2 (January 2011): 229–43.

Street, Sharon. "A Darwinian Dilemma for Realist Theories of Value." *Philosophical Studies* 127, no. 1 (January 2006): 109–66.

———. "Objectivity and Truth: You'd Better Rethink It." In *Oxford Studies in Metaethics*, vol. 11, edited by Russ Shafer-Landau, 293–334. Oxford: Oxford University Press, 2016.

Tersman, Folke. "Debunking and Disagreement." *Noûs* 51, no. 4 (December 2017): 754–74.

———. *Moral Disagreement*. Cambridge: Cambridge University Press, 2006.

———. "Moral Disagreement: Actual vs. Possible." In *Disagreement and Skepticism*, edited by Diego E. Machuca, 90–108. New York: Routledge, 2013

Vavova, Katia. "Evolutionary Debunking of Moral Realism." *Philosophy Compass* 10, no. 2 (February 2015): 104–16.

———. "Moral Disagreement and Moral Skepticism." *Philosophical Perspectives* 28, no. 1 (December 2014): 302–33.

Weatherson, Brian. "Disagreements, Philosophical and Otherwise." In Christensen and Lackey, *The Epistemology of Disagreement*, 1–17.

Wedgwood, Ralph. "The Moral Evil Demons." In Feldman and Warfield, *Disagreement*, 216–46.

White, Roger. "You Just Believe That Because …" *Philosophical Perspectives* 24, no. 1 (December 2010): 573–615.

Wielenberg, Erik J. "On the Evolutionary Debunking of Morality." *Ethics* 120, no. 3 (April 2010): 441–64.

———. *Robust Ethics: The Metaphysics and Epistemology of Godless Normative Realism*. Oxford: Oxford University Press, 2014.

Williamson, Timothy. *Knowledge and Its Limits*. Oxford: Oxford University Press, 2000.

Wright, Crispin. *Truth and Objectivity*. Cambridge, MA: Harvard University Press, 1992.

Wynn, Thomas, and Frederick L. Coolidge. *How to Think Like a Neandertal*. New York: Oxford University Press, 2013.

OPPRESSION, FORGIVENESS, AND CEASING TO BLAME

Luke Brunning and Per-Erik Milam

W RONGDOING is an inescapable fact of life. We all do wrong and are wronged from time to time and in response we often blame one another. In the broadest sense, moral blame is a personal response to wrongdoing or wrongbeing, which can manifest in a variety of mental states—e.g., judgments, desires, dispositions, and emotions—as well as in behavior. We blame for a variety of wrongs, in a variety of ways, and with a variety of consequences: one expresses disappointment with an unfaithful partner who then apologizes, another rants about injustice thereby alienating part of her Facebook community, a third turns inward in frustration with a neglectful parent who in turn mistakes her withdrawal for indifference. Such conflicts are not the whole or even the greater part of our shared social existence, but they are a defining feature of it.

But if blame is a defining feature of our social lives, so is ceasing to blame. And we cease to blame in a variety of ways, too. Depending on the circumstances, we might excuse, justify, or forgive an offender, or we might simply let the offense go. Each of these ways of ceasing to blame is a social practice with characteristic norms, whether formal or informal, that influence when and how we do it, as well as how it is received. We are wary of those who let go too easily; we find it difficult to forgive an offender who has yet to show remorse; and sometimes learning more about the offender demands that we excuse their offense. In short:

Claim 1: Whether and how we cease to blame depends on a variety of circumstances, not all of which are under our control.

While not obvious, we think this point is plausible. However, it has some surprising implications. Like any norm-governed practice, one can cease to blame appropriately or inappropriately, successfully or unsuccessfully. Indeed, one can fail altogether to overcome blame. This suggests that:

Claim 2: Social and material circumstances can compromise one's ability to successfully cease to blame in the manner one would prefer.

Moreover, the possibility of failure implies that one may lack access to particular ceasing-to-blame practices, because one can be in a position to be regularly prevented from successfully overcoming blame. In order to participate in some practices, one's action must be done for the right reasons and secure uptake. This raises a further, political question: Does everyone have equal (or adequate) access to the various ceasing-to-blame practices? We will argue that they do not. In particular:

Claim 3: The circumstances of oppression can systematically undermine one's ability to successfully perform some ceasing-to-blame practices.

Our argument proceeds as follows. In section 1, we present a taxonomy of different ceasing-to-blame practices and describe their distinctive roles in our moral lives (*Claim 1*). We also explain the value of overcoming blame and, thereby, the harm of not being able to do so. Our subsequent discussion focuses on forgiveness, though our arguments also apply to other ways of ceasing to blame. We focus on forgiveness because it is a complex practice about which much has been written and because many regard forgiving as a matter of personal fiat, a way of ceasing to blame that is elective, unconditional, or otherwise independent of social circumstances. As such, our argument faces the strongest opposition and is most interesting in the case of forgiveness. In sections 2 and 3, we argue that circumstances can conspire to compromise an individual's ability or opportunity to forgive. We argue that forgiveness is reason guided and that lacking good reasons or the right kind of reason can undermine one's ability to forgive (*Claim 2*). In addition, circumstances can be such that victims' attempts to forgive are not recognized. We make the case that recognition—or "uptake"—is necessary for forgiveness. We note, however, that even if forgiving does not require uptake, communicating forgiveness does, and communicating forgiveness is itself an important social practice. In section 4, we argue that the circumstances of oppression systematically compromise the ability of oppressed people to forgive and that this deprivation constitutes a significant but neglected harm (*Claim 3*). In section 5, we address two particularly forceful objections to our view: that forgiveness is always open to the victim and that we overlook the ability of the oppressed to shape their own practices. Finally, in section 6, we explore an important implication of our account of ceasing to blame as a socially scaffolded set of moral practices.

1. BLAMING AND CEASING TO BLAME

1.1. Blame

How we understand ceasing to blame depends, of course, on what it means to blame. While everyone acknowledges and tries to capture a broad set of paradigmatic cases, there is significant disagreement about the nature of blame—i.e., the attitudes and activities involved in blaming.[1] On "cognitive" accounts blame consists in evaluations or judgments of the offender.[2] On "conative" accounts blaming requires a judgment, but also a change either in how one is disposed to feel and act toward an offender or how one perceives one's relationship with him.[3] Finally, on "affective" accounts blame requires, or is constituted by, a negative emotion or hard feeling.[4]

Whatever its nature, though, most accounts agree that the purpose of blame is, roughly, to communicate a response to mistreatment. How and what exactly blame communicates is disputed, but there appears to be broad agreement about the communicative view.[5] Ceasing to blame is an essential part of this communicative enterprise because it signals that blame's demand, whatever it was, has been met—or, in some cases, is no longer being pursued. Given the general agreement on the point and purpose of blame, the arguments of this paper do not depend on which particular account of blame is correct. They remain plausible on any account that captures the paradigmatic cases of blame.

1.2. Ceasing to Blame

People cease to blame in four main ways: by excusing, justifying, forgiving, and letting go.[6] They can be distinguished by the kinds of reasons they require. To

---

1   For a helpful taxonomy, see Coates and Tognazzini, "The Nature and Ethics of Blame."

2   Hieronymi, "The Force and Fairness of Blame"; Smith, "Moral Blame and Moral Protest."

3   Sher, *In Praise of Blame*; Scanlon, *Moral Dimensions*.

4   Wallace, *Responsibility and the Moral Sentiments* and "Dispassionate Opprobrium"; Wolf, "Blame, Italian Style."

5   Moral blame identifies and responds to wrongdoing (McKenna, *Conversation and Responsibility*), protests mistreatment (Hieronymi, "The Force and Fairness of Blame"), and demands recognition of past mistreatment and better treatment in the future (Wallace, *Responsibility and the Moral Sentiments*; Walker, "Third Parties and the Social Scaffolding of Forgiveness"; Smith, "Moral Blame and Moral Protest"; Macnamara, "Blame, Communication, and Morally Responsible Agency"). There are criticisms of communicative accounts of blame—e.g., on the grounds that it cannot explain private blame (Driver, "Private Blame")—but also a number of replies, including Fricker, "What's the Point of Blame?" and McKenna, "Quality of Will, Private Blame and Conversation."

6   Murphy ("Forgiveness and Resentment") makes the same distinctions in his seminal treat-

track these differences, consider the following case. While at a party, Anna tells Boris a joke about recent migrants to their country. The joke seems crude and offensive, and Boris reasonably blames Anna for her remarks. Consider the different ways that Boris might cease to blame Anna.

Boris may excuse Anna. Although the joke is clearly offensive, Boris can excuse Anna if he judges that she is not fully responsible for what she said. For instance, he might realize that Anna lacked the background knowledge to see how her joke could be offensive, and that her ignorance was reasonable—perhaps she was unaware that the joke was a modern riff on one historically made at the expense of another despised group.

Boris may justify Anna's remarks. He might realize that Anna was using the structure of the joke to ridicule people who fear migrants unreasonably. He realizes, that is, that Anna was not doing wrong by making the joke.

Boris may forgive Anna. He might believe both that Anna reasonably could have known that it was offensive and that it was not justified. However, if Anna realizes how hurtful her comment was, expresses remorse, and resolves to refrain from such "humor" in the future, then Boris might decide that she lacks any deep ill will and forgive her, thereby relinquishing blame.

Finally, Boris can let go of his blame. For example, he may distract himself in the company of other friends. Alternatively, as a migrant himself, Boris may have been so browbeaten by such "jokes" in the past that it is no longer worth it to him to continue blaming, perhaps because blame is too emotionally fatiguing.[7]

Since this paper will focus primarily on forgiveness, let us clarify how we understand that concept. Our aim is not to argue for or challenge any particular account of the nature of forgiveness, at least not directly, but rather to identify and explain obstacles to forgiveness and other ceasing-to-blame practices that can arise in circumstances of oppression.

The arguments we develop below are compatible with many of the leading accounts of forgiveness.[8] Most accounts accept some version of the following

---

ment of forgiveness.

7   Many philosophers distinguish forgiving from letting go, though the label itself is uncommon (Hieronymi, "Articulating an Uncompromising Forgiveness," 530; Griswold, *Forgiveness*, 70; Allais, "Wiping the Slate Clean," 43–44, n26; and Pettigrove, *Forgiveness and Love*, 4, 97.) However, depending on how exactly the distinction is drawn, some would view them simply as different forms of forgiveness (e.g., Bennett, "Personal and Redemptive Forgiveness," or Nelkin, "Freedom and Forgiveness"). Whether and how one draws the distinction does not substantially affect our argument in this paper. On our view, both forgiveness and letting go can be compromised by the circumstances in which one ceases to blame. (We develop a full account of letting go in other work.)

8   See, for example, Hieronymi, "Articulating an Uncompromising Forgiveness"; Griswold, *For-*

conditions. In order to forgive one must: i) overcome one's negative attitude toward an offender, ii) about their offense, iii) for the right reasons. For example, Zora ceases to blame Toni for her betrayal because Toni apologizes.[9] We will argue that oppression can undermine one's ability to meet even these minimal conditions. Of course, our account is not compatible with every conception of forgiveness and some could argue that it is not forgiveness but something else that is compromised in the scenarios we describe.[10] However, the vulnerabilities we identify are not idiosyncratic features of our conception, but stem from widely held commitments about how forgiveness typically works.

The different ceasing-to-blame practices are similar but distinct. Justification requires that one cease to view the purported offense as wrong. Excuse requires that one cease to view the offense as one for which the offender is responsible. Forgiveness requires that one continue to view an offender as a culpable wrongdoer and is therefore incompatible with justification and excuse. Letting go is also a response to culpable wrongdoing, but in doing so one ceases to blame for different reasons.[11] These distinctions are used in ordinary discourse in roughly the way we have outlined and people seem to police their appropriate use. For example, people are wary of premature forgiveness and of those who appear to

giveness; Allais, "Wiping the Slate Clean"; Garrard and McNaughton, *Forgiveness*; and Pettigrove, *Forgiveness and Love*.

9 Different views may disagree about the details of these broad requirements. Some argue that a forgiver must cease to resent the offender (Garrard and McNaughton, "In Defence of Unconditional Forgiveness"), others that one must come to view the offender as better than their offense warrants (Allais, "Wiping the Slate Clean"), and still others that one need only overcome excessive blame (Butler, *The Works of Bishop Butler*; Garcia, "Bishop Butler on Forgiveness and Resentment"). Nicolas Cornell ("The Possibility of Preemptive Forgiving") even argues that we can forgive preemptively and, in these cases, may never experience any negative attitude toward the offender. Likewise, there are more and less strict notions of what counts as a reason to forgive (Murphy, "Forgiveness and Resentment," and Nelkin, "Freedom and Forgiveness," respectively). We discuss reasons to forgive in section 2.

10 William Neblett ("Forgiveness and Ideals") suggests that one can forgive without a change of attitude toward the offender. Dana Nelkin ("Freedom and Forgiveness," 170, 182) suggests that one can forgive for any reason or no reason at all. Leo Zaibert's account of "pure forgiveness" ("The Paradox of Forgiveness") also lacks a right-reasons requirement, though such a condition is not strictly inconsistent with his view. The same is true of accounts according to which forgiveness is fundamentally an exercise of a normative power (Warmke, "The Economic Model of Forgiveness" and "The Normative Significance of Forgiveness"). The arguments of this paper do not dispute such accounts directly, though the internal plausibility of our account and its fit with other plausible conceptions of forgiveness may pose an indirect challenge to any view that cannot accept our conclusions.

11 Some offenses warrant partial excuse, justification, and forgiveness. Angie may realize that Bell's behavior was not *as bad* as she thought (justification), that she was not *entirely* responsible for what happened (excuse), but that there is nonetheless *something* to forgive (or let go).

lack self-respect.[12] They are also wary of people who are reluctant to forgive under seemingly ideal conditions (e.g., when someone has shown much remorse and made amends). This wariness suggests that forgiveness is not a *sui generis* type, but is best understood as one form of a broader practice of ceasing to blame.

1.3. The Value of Ceasing to Blame

This paper will argue that the circumstances of oppression can undermine a person's ability to cease blaming in the way she would prefer and that this constitutes a harm that has thus far been neglected in discussions of oppression. It is not the only harm imposed on oppressed persons—they are also more likely to be injured and disrespected in the first place—but it is a significant one. We will argue that offenders and bystanders can behave—whether purposefully, knowingly, or negligently—in ways that compromise victims' abilities and opportunities to forgive. Suppose for now that this is true. Our aim in this section is to show that overcoming blame can be good, that forgiving is sometimes preferable to other ways of overcoming blame, and that it is a harm to force a victim to choose between deficient alternatives. In order to understand this neglected harm, we must first explain the value of being able to overcome blame and of doing so in particular ways.

First, the ability to blame and overcome blame are part of our capabilities of emotion and affiliation.[13] Nussbaum only mentions blame (or "justified anger"), but the ability and opportunity to *overcome* blame is just as important to developing and maintaining attachments. Developing this dimension of our capability of emotion may also help cultivate moral agency.[14] Likewise, the ability to hold one another responsible is part of the capability of affiliation, understood as the ability to "engage in various forms of social interaction" and "having the social bases of self-respect and non-humiliation."[15] Holding ourselves and others responsible—e.g., through blame, apology, forgiveness, and other practices of moral address, accounting, and reconciliation—is how we maintain supportive, trusting, and cooperative relationships and communities. (The capability of emotion also contributes to this end.) Undermining a person's ability or op-

---

12   Novitz, "Forgiveness and Self-Respect."

13   Nussbaum, *Women and Human Development*, 79. Having a capability requires an ability and the opportunity to exercise it. For example, one must have both the right to vote and the opportunity to exercise that right, or the capacity for play and opportunities to engage in play. Undermining the capability in either respect harms the agent.

14   Vargas, *Building Better Beings*.

15   Nussbaum, *Women and Human Development*, 79–80.

portunity to overcome blame deprives them of valuable capabilities and of the benefits of exercising them.

Second, people often want to overcome blame because it feels bad. This is not always so—righteous anger may feel empowering—but it often is. It can be painful to blame those we love and care about, even when we think blame is appropriate. We want to reconcile with them and want to be able to overcome our blame for this reason. It can also be both painful and frustrating to blame someone who does not wish to reconcile or who continues to bear us ill will. Here, too, though for different reasons, we may want to overcome our blame and move on.

Third, all else being equal, ceasing to blame is good when blame is no longer fitting. Blame represents the world in a particular way and when it ceases to represent it accurately, it is good to cease blaming. Susan Wolf calls this the value of "living in accordance with the facts."[16] It is arguably one of the goods sacrificed by plugging into the Experience Machine. Moreover, by misrepresenting the world—the offender, the offense, or their relevant context—blame can cause injury and disrespect, and be unfair.

Such is the value of overcoming blame and the harm of not being able to do so. However, these points do not imply that it is good to be able to overcome blame *in a particular way*—e.g., that depriving one of the ability or opportunity to *forgive* harms that person. While blaming can be painful, there are different ways of avoiding it, so inability (or lack of opportunity) to forgive is a harm only if there are no other options or if the other options are themselves harmful or otherwise inadequate. Since we recognize other options—one can forgive a culpable wrongdoing or let go of blame without forgiving—we must explain why they are inadequate. Likewise, one might accept that the ability to overcome blame partly constitutes a valuable capability, but deny that this capability is undermined if there is another way to overcome blame. Thus, in order to show that a particular way of overcoming blame is valuable, we must explain why the alternatives are (sometimes) inadequate.[17] One way to do this is to compare victims and non-victims.

Consider Vicky and Imogen, both of whom believe that they have been betrayed by a close friend. Vicky actually has been betrayed, and her friend's behavior and subsequent attitude are making it difficult to forgive. Imogen, however, has not been betrayed. She is also finding it difficult to forgive, but this is because her friend, reasonably enough, has denied any wrongdoing and is questioning the legitimacy of her blame. Both Vicky and Imogen have been prompted to

---

16 Wolf, "The Importance of Free Will," 399.

17 Another option would be to refrain from blaming in the first place. However, while it is sometimes possible to refrain from blaming others for their blameworthy behavior, it is not a genuine option. It is not always possible, often difficult, and rarely (if ever) a reasonable expectation.

blame, and let us suppose that both are burdened by their anger and the task of overcoming it. However, they have different claims against their supposed betrayers. We can see the value of being able (or having the opportunity) to forgive, in particular, by comparing the choices imposed on Vicky and Imogen.

Vicky's betrayer deprives her of the ability to forgive and burdens her with a hard choice, while Imogen's innocent friend does not. The former forces Vicky to choose between letting go of appropriate blame and bearing an unjustly imposed burden of continuing to blame. This forced choice is a harm insofar as both options impose burdens and Vicky has a reasonable claim not to be burdened in either way. Vicky could choose not to stand up for herself, which would threaten her self-respect and risk condoning the wrong (or appearing to others to do so). Or she could continue to blame and suffer the hardship of doing so. Things are different for Imogen. She bears a similar burden as a result of her misunderstanding, but it has not been imposed by the offender. Perhaps her friend should try to disabuse her of her mistaken belief since it is causing her to suffer, but Imogen has no more claim on his help than she does on anyone else who could explain her mistake. Vicky's friend has imposed a burden that only he can remove; Imogen's friend has not.[18]

This account explains why having the ability to forgive and the opportunity to do so is good for a victim. The victim can choose how to respond to an offense. When a victim cannot forgive in a particular case, as we will argue can happen, she is forced to choose between two unreasonable options. The offender imposes this hard choice; he deprives the would-be forgiver of the option to forgive by making it difficult or impossible—e.g., by lacking remorse, not apologizing, or demonstrating continuing disregard and lack of care. The ability to overcome blame by forgiving is a good and being deprived of it is a harm. We will argue that members of oppressed groups are particularly susceptible to this kind of harm, which can be imposed by individuals or by a society that fails to recognize their victimization.

2. THE RIGHT REASONS

Particular ceasing-to-blame practices require the right kinds of reasons. If one lacks the right kind of reason, one cannot perform the practice; it is not "on the table," so to speak. Philosophers defend right-reasons requirements for various

---

18   This argument applies beyond the context of blame and forgiveness. A person you have to trust harms you by not giving you a reason to trust them. Vicky would be harmed by being put in the position of having to trust the friend who betrayed her because she has no other choice of confidante. Or, for example, a driver who offers a ride to a stranded motorist harms her by giving her no reason to entrust her safety to him.

practices, including love, trust, and forgiveness.[19] For example, one might make the following kind of argument about love: *R* is not a reason to love, so, if *A* has strong feelings for *B* solely for reason *R*, then *A* does not love *B*. Parallel claims can be made for trust or forgiveness. Such arguments are contested, but widely accepted. In this section, we argue that forgiveness has a right-reasons requirement.

First of all, some reasons just seem to be the wrong kind. One can cease to blame a culpable wrongdoer without having forgiven. At the very least, one cannot forgive for the same reason one blames. Boris can blame and forgive Anna for her betrayal, but he cannot blame her *because she betrayed him* and also forgive her *because she betrayed him*. This is the wrong kind of reason. But this is not the only restriction. The following are all reasons for Boris to cease blaming, but none seems like a reason to forgive: because Anna has the same favorite movie as him; because Anna is likely to promote him; or because doing so will contribute to slightly better cardiovascular health.[20] It seems wrong to describe what Boris has done in these cases as forgiving because he did not decide in virtue of some fact about Anna or her offense. Imagine that a coworker who has wronged you asks you to cease blaming him but can offer no reason to do so other than that he would recommend you for promotion. This case seems best described not as your coworker giving you a reason to forgive but rather your coworker suggesting a reason to cease blaming despite having *no reason* to forgive. If any of these restrictions is plausible, then the question to ask is not *whether* there is a right-reasons requirement on forgiveness, but *which* reasons are the right kind.

Most forgiveness theorists implicitly or explicitly accept a right-reasons requirement. Murphy explicitly argues that forgiveness must be done for particular moral reasons.[21] McGary defends a wider but still limited set of reasons.[22] The requirement appears to support Allais's distinction between forgiving and therapeutic self-management and Pettigrove's distinction between forgiving and just getting over it.[23] Even Garrard and McNaughton, who defend unconditional forgiveness, seem to accept a right-reasons requirement.[24] These accounts

19 Abramson and Leite, "Love as a Reactive Emotion"; Smuts, "Normative Reasons for Love, Part 1"; Hieronymi, "The Reasons of Trust"; Murphy, "Forgiveness and Resentment."

20 Larsen et al., "The Immediate and Delayed Cardiovascular Benefits of Forgiving."

21 Murphy, "Forgiveness and Resentment."

22 McGary, "Forgiveness."

23 Allais, "Wiping the Slate Clean," 43n26; Pettigrove, *Forgiveness and Love*, 97.

24 Garrard and McNaughton's view suggests an important clarification (*Forgiveness*, 114). The claim that forgiving requires the right kind of reason does not imply that forgiveness is not elective or unconditional, though it does propose limits on forgiving. Claims about electivity and conditionality are about the rational and/or moral permissibility of forgiving

are attempting to capture the intuition that one can cease to blame a culpable wrongdoer without forgiving.

The right-reasons requirement is not uncontested, though. To our knowledge, Dana Nelkin is the only philosopher to argue explicitly that one can forgive for any reason (or none at all).[25] On her view, to forgive is to release the offender from a moral debt. Other proponents of debt-release models of forgiveness may also reject a right-reasons requirement, though none does so explicitly.[26] However, these accounts seem incomplete without such a requirement. If one cancels a moral debt because one judges that amends have been made, then one seems to have forgiven. However, to do so because one views the wrongdoer or the offense as beneath one's notice does not seem to be forgiveness.[27] Again, it is not enough to explain how the forgiver changes when she forgives; one must also explain what prompts that change.

Further intuitive support for the right-reasons requirement comes from the need to distinguish between forgiving and condoning. One condones blameworthy behavior if one withdraws warranted censure in order to, for example, avoid social awkwardness or curry favor. But this too seems different from forgiving. To condone in this way is not merely to forgive for morally bad reasons; it is to cease blaming for the wrong kind of reason—namely, personal comfort or gain.

Finally, we can argue for a right-reasons requirement by drawing an analogy to trust. Pamela Hieronymi has argued that many of the reasons *in favor of trusting* are not reasons *to trust* because they do not support a trusting belief.[28] We can tell these reasons apart in virtue of the fact that they answer different questions. Reasons in favor of trusting answer the question, "Would it be good or valuable to trust Anna?" Reasons to trust answer the question, "Will Anna do what I ask?" It is a reason in favor of trusting Anna to keep his secret that he can relieve the stress of keeping it to himself. However, this fact does not bear on whether Anna can be counted on not to tell others. We claim that forgiveness is similar. Boris's cardiovascular health is a reason to think that ceasing to blame Anna would be good, but not a reason to believe that he should view Anna differently with respect to her misconduct.[29] Further, as with trust, if one

---

when it is possible. Calhoun ("Changing One's Heart") and Allais ("Elective Forgiveness") defend elective forgiveness, and Milam ("Against Elective Forgiveness") rejects it without taking up right kinds of reasons.

25   Nelkin, "Freedom and Forgiveness."

26   Twambley, "Mercy and Forgiveness"; Warmke, "The Economic Model of Forgiveness."

27   Griswold, *Forgiveness*, 12–15.

28   Hieronymi, "The Reasons of Trust," 231–32.

29   Allais makes a similar point ("Wiping the Slate Clean," 39n12).

could forgive for any reason that spoke in its favor, then one could change one's attitude insincerely. But, while a person might express forgiveness insincerely, the attitudinal change itself cannot be insincere. This too suggests that there are limits on which reasons can count as reasons to forgive.

So what are the right kinds of reason? Philosophers disagree about this, but for our purposes it is sufficient to point to commonly identified reasons. One set of reasons includes indications of a change of heart on the part of the offender: apology, remorse, repudiation of the offense, making amends, confession, atonement, and repentance. For example, Boris might cease to blame Anna for her hurtful joke because Anna has shown remorse, sincerely repudiated her action, or tried to make amends for the slight. Murphy suggests that one may have reason to forgive if an offense was well-intentioned—e.g., if a colleague speaks down to you while attempting to clarify their argument.[30] And Garrard and McNaughton argue that solidarity with morally fallible offenders is a reason to forgive.[31] We believe that some of these are more plausibly counted as reasons to forgive than others, but we cannot enter that debate here.[32] Instead, we will simply note that the narrower the set of right reasons, the more likely it is that one could lack such a reason and be unable to forgive.

This is the upshot of our argument. If one lacks the right kind of reason, one might try to perform the practice but fail to do so. One might fail to do what one wants and intends and do something else instead. For example, Boris might cease to blame because he sees that Anna's offense was not her fault, without realizing he has actually excused rather than forgiven. Of course, despite the above argument, some may still reject the right-reasons requirement and deny that facts about an offender or society generally can render a victim unable to forgive by depriving her of the right kinds of reasons. However, this commitment is only one strand of our argument, one way in which the ability to forgive can be compromised. Even if there are no right kinds of reasons, there are better and worse reasons to forgive. So, even if forgiveness remains "on the table" despite lacking the right kind of reason, one may nonetheless be less able to forgive if one lacks a good reason.

3. THE RIGHT WAYS

In the previous section, we argued that some practices, including forgiveness, have a right-reasons requirement. This requirement is conceptual; it is a matter of which ceasing-to-blame practices are "on the table" so to speak. We showed

30   Murphy, "Forgiveness and Resentment."

31   Garrard and McNaughton, *Forgiveness*.

32   For one account, see Milam, "Reasons to Forgive."

that one's ability to overcome blame in the way one prefers can be compromised if one lacks the right kind of reason. This applies as much to forgiveness and excuse as it does to, say, love. However, even when the right kinds of reasons are available, participation in a practice can still be unsuccessful or inappropriate. In this section, we argue that the circumstances within which one attempts to overcome blame can undermine one's ability and complicate the morality of doing so. In particular, we will argue that one's ability to cease blaming in the way one prefers can be compromised if one does not receive the necessary recognition or uptake of one's attempt—e.g., of one's proffered forgiveness.

3.1. Three Ways to Fail

Even when one has the right kinds of reasons to forgive, one can be prevented from doing so in one of three ways. First, one might be psychologically unable to overcome one's blame and have a different attitude toward them. For example, you might want to forgive a friend who routinely behaves badly, but find your motivation is undercut by his obnoxious comments. Or one might be able to change one's attitude toward an offender, but not about the offense in question. For example, perhaps you grudgingly admire a colleague's professionalism after a personal conflict, despite still holding a grudge about your mistreatment. These are simple cases of recalcitrant attitudes, in which one struggles to forgive, and which result from contingent and uncontrollable psychological facts about oneself, the offender, or the context of the offense.

Second, one might overcome one's blame for the right reasons and have one's forgiveness recognized, only for it to fall flat—i.e., fail to have the desired and expected effect. This can happen in two ways. One's forgiveness may be recognized and accepted, but be ineffective or counterproductive. For example, while forgiving an unfaithful partner might typically lead to a change in their behavior, sometimes it may encourage further cheating. Something similar seems to have happened in the case of Ronald Carlson and Karla Faye Tucker. Tucker was convicted of murdering Carlson's sister and sentenced to death by the U.S. state of Texas. While in prison she claimed to have found God, repented, and repudiated her crime. Carlson responded by forgiving her. He seems to have hoped that, by forgiving Tucker for murdering his sister, the rest of his community would accept her claims to have repudiated her crime and committed herself to making amends. Unfortunately, his doing so may have merely drawn further indignation on himself. In these cases circumstances diminish or redirect the force of forgiveness, and we will argue that oppressive social norms can have this effect (section 4). Alternatively, one's forgiveness may be recognized but rejected, as one might reject a proffered apology. For example, an offender might reject for-

giveness because they believe themselves unworthy. Lee, the protagonist of the film *Manchester by the Sea*, is wracked with guilt after causing the deaths of his children and struggles to accept his wife's forgiveness. We typically expect people to accept forgiveness in cases in which we have reason to forgive—e.g., because they have expressed remorse or tried to make amends. Difficult cases arise when others act remorsefully and give one reasons to forgive and yet continue to reject one's forgiveness. This mismatch is part of the tragic aspect of Lee's situation in the film.

Third, social circumstances can prevent one from securing uptake altogether when one tries to forgive. The offender or third parties may simply fail to recognize one's overture *as forgiveness.* For example, members of a community might construe a victim's attempt to forgive as an act of condoning or a misplaced excuse. Or one's attempt to forgive might not be recognized at all. In this respect, attempts to forgive resemble attempts to refuse. One cannot successfully refuse an offer if one's refusal is not recognized. Imagine, for example, that when getting a haircut, the stylist asks whether you want your hair shampooed first. Responding with, "That's okay" is only a successful refusal if it is received as such, and not if it is interpreted as "Okay, go ahead." Likewise, one's attempt to forgive the hairdresser for causing your allergic reaction to the shampoo must be received as such. If it is instead received as an excuse—something like "It wasn't your fault, I said 'Okay'"—then, despite meeting the personal conditions on forgiving and having the right kind of reason, one's attempt has gone awry.[33]

3.2. Uptake: A Weak View and a Strong View

Failure to secure uptake of one's attempt to forgive is perhaps the most harmful of all failures canvassed above. However, the nature of this failure is contested. Two different views are possible, depending on whether one distinguishes between public expressions of forgiveness (or other ways of ceasing to blame) and private forms of forgiveness.

According to what we will call the weak view of this distinction, failures of uptake only undermine the expression of forgiveness, while one's ability to forgive is

---

33 In addition to unsuccessfully ceasing to blame, one can do so successfully but inappropriately. One way to inappropriately cease blaming is to misunderstand what the circumstances give one reason to do. For example, one might fail to notice that a proffered apology is less a reason to forgive than a subtle attempt to justify the misconduct or avoid responsibility. This can happen through a mistake of perception or of interpretation. Another way to inappropriately cease blaming is to do so carelessly or for bad moral reasons. For example, some find it careless (or even reckless) to forgive immediately in the aftermath of a tragic event, like a bombing or shooting. Similarly, it seems impermissible to forgive out of a desire to belittle an offender.

unaffected. Alternatively, one may reject this distinction, holding instead that illocutionary uptake is a necessary condition on forgiveness itself. According to this strong view, forgiving someone requires recognition in the same way that warning, consenting, refusing, and, as Miranda Fricker has argued, blaming do.[34] Failure to secure uptake when one attempts to forgive prevents one from forgiving at all.

We can highlight the differences between these two views by considering attempts to forgive absent or dead offenders. Suppose that Pavel writes a letter telling Lin that he forgives her for some past offense. Some would say that he has successfully forgiven her before she ever receives the letter. (A similar case can be imagined for forgiving the dead.) This would be to endorse the weak view. However, others would say that such cases actually put pressure on the weak view. This is our position. Supposing that Pavel has overcome his blame about the offense and done so for the right reasons, there is nonetheless a sense in which the fact that Lin has not, may not, or cannot (if she is dead) recognize Pavel's attempt to forgive, means that he has not forgiven. Note also that the plausibility of the letter case depends in part on the assumption that our attempts to forgive will be received in roughly the way we intend them. Pavel's case looks different if we stipulate that, upon receiving his letter, Lin rejects his offer of forgiveness or fails to recognize it as such. Likewise, if we imagine that it could go either way—be recognized or not—we can see how plausible it is to claim that the attempt to forgive is not (yet) successful when the absent or dead offender cannot receive it. We think the strong view captures both the sense in which Pavel succeeds (he overcomes blame about the offense) and the sense in which he falls short (he does not receive the necessary uptake). Understanding forgiveness in this way also explains why attempts to forgive the absent and the dead are often unsatisfying or even tragic—we need the offender to acknowledge the legitimacy of our blame and to recognize our offer of forgiveness.

Moreover, even private forgiveness requires uptake of a sort. Imagine a would-be forgiver who is deeply ambivalent or conflicted in his attitude toward the offender. An employee in a dysfunctional and hostile workplace may harbor doubts about his own disposition to forgive that someone in a more supportive workplace might not experience. For example, he might wonder whether it is really appropriate for him to have forgiven his boss, even if only in his heart, or what his forgiveness says about his commitment to protesting the kinds of mistreatment he and others in his office have experienced.

However, our conclusion is significant whether one endorses the strong or the weak uptake condition. Even if it is only the ability to *express* forgiveness

---

34 Fricker, "What's the Point of Blame?" 172.

that can be undermined, this ability is a valuable feature of our shared social and moral lives.

So, to summarize, one's ability to forgive can be compromised in a number of ways: one may lack the right kind of reason to forgive; one may have the right kind of reason, but be psychologically incapable of overcoming one's blame; one's forgiveness may fail to have the desired or intended effect; and, finally, one can fail to forgive by failing to secure uptake. In describing this phenomenon, we have focused on forgiving, but our conclusion plausibly generalizes to other ceasing-to-blame practices, like excuse, justification, and letting go.[35]

4. OPPRESSION AND CEASING TO BLAME

The argument thus far has had two stages. First, we argued that our ceasing-to-blame practices are sensitive to reasons (section 1). Next, we argued that an individual's ability to participate in these practices can be compromised if they lack access to the relevant reasons or if their attempt does not secure uptake (sections 2 and 3). In this sense, forgiving resembles other social practices, from proposing marriage to paying taxes, that are reason guided and require uptake to be successful.

The upshot of the previous section was that social life can be capricious and unpredictable. Anyone can find oneself in circumstances in which the proclivities of those around them impede their ability to blame and to cease to blame. For most people, these failures are local instances of pressure or manipulation, exceptions to the comparatively untroubled flow of social interaction. Others, however, may find that their ability to forgive is systematically undermined. In this section, we argue that members of oppressed groups often struggle to forgive successfully. This fact is a significant harm of oppression, but one that is not commonly recognized.

In section 1, we explained why the ability to overcome blame in the manner one prefers is good and being deprived of it is a harm. However, we have since focused on how the circumstances of oppression can undermine the ability to forgive. While those points also apply to forgiveness, in order for this particular inability to be a harm, the ability to forgive must have unique value that other forms of ceasing to blame do not. If the aim is only to avoid the unpleasantness of blaming, this can be done in many ways. However, one often desires both to overcome blame and relate to the wrongdoer in a particular way. For example, a victim may want to reconcile with a repentant offender who shows due respect and care for them. It may be unsatisfying, frustrating, or saddening to recognize

---

35   Calhoun, "Responsibility and Reproach."

that one's only option is to resign oneself to the offender's callous indifference and hope to avoid similar mistreatment in the future. Similarly, even when one has the option to forgive, it is painful, degrading, and unfair when the circumstances surrounding the offender and their offense—especially hatred, prejudice, or disregard—make forgiving difficult or require an extra sacrifice beyond the usual burden of overcoming reasonable blame.

The circumstances of oppression restrict the kinds of life one can lead and the opportunities for flourishing one has, and shape the development of one's character and agency. According to Iris Marion Young's influential analysis of oppression, "all oppressed people suffer some inhibition of their ability to develop and exercise their capacities and express their needs, thoughts, and feelings," but there are many interlocking facets to this inhibition, and they affect different social groups in distinct ways.[36] For Young, oppression has five faces: exploitation, marginalization, powerlessness, cultural imperialism, and violence. Groups are exploited when they are isolated from what they produce in a way that benefits another social group.[37] Exploited people, whether sharecroppers in the post-Reconstruction American South or tobacco pickers in present-day Malawi, are typically materially disadvantaged and potentially desperate, or uninformed about better options, or both.[38]

Groups are marginalized when they are denied access to the public goods their society provides. Marginalization may be formal, as when a group is denied the right to vote, or the consequence of systematic disincentives, as when tuition costs deter potential university applicants from a particular group. It can manifest itself physically in terms of barriers, border policing, or racial/class segregation, or psychologically in the policing of language, fashion, or other norms.[39] It generates material deprivation, as when women are excluded from higher-paying jobs and fields, and prevents groups from exercising their capacities in "so-

36 We use Young's account, but our argument is broadly compatible with other influential analyses of oppression, including Frye, *The Politics of Reality*; Haslanger, "Oppressions"; and Cudd, *Analyzing Oppression*. For Young ("Five Faces of Oppression," 55), oppression applies to groups, like lesbians or African Americans. Groups differ from agglomerations of individuals in that groups are sustained through processes of identification. For example, contrast a group of queer residents with an agglomeration of citizens of a particular legislative voting district.

37 This need not be a material product. Poor whites can be exploited to produce racial tension that helps maintain racial power structures that benefit rich whites. The United States and South Africa are notable examples, but not the only ones.

38 See Palitza, "Child Labour," for an account of tobacco companies' exploitation of child labor.

39 hooks, *Feminist Theory*, xviii.

cially defined and recognized ways," as when housing discrimination prevents members of some group from renting in a particular area.[40]

Groups are subject to cultural imperialism when their society refuses, or fails, to recognize their collective experiences, perspectives, and proclaimed identities. This may happen when a group is forced to assimilate into a dominant culture, when their traditional identities are stereotyped by those in privileged groups, or when the institutions within which they live and work deny their significance. For example, the fact that academic schedules in Western universities are structured around Christian holidays communicates lesser concern for Jews, Muslims, and members of other religions.[41] Similarly, enforcing norms of dress and appearance associated with a particular race or class can undermine other groups' ability to express themselves.[42]

Groups are powerless when they are dominated by other groups and their members lack social status and struggle to be autonomous.[43] Political disenfranchisement is an obvious example, but caste systems, poverty, and extensive managerial authority over workers can also create degrees of powerlessness.[44] Finally, groups are often subject to systematic violence, which is legitimized by prevailing norms and ideals, and sometimes even by public institutions. The legacy of racial violence in the United States, for example, includes the practice of slavery, the terrorism of the Ku Klux Klan during Jim Crow, and (arguably) the routine use of excessive force by police against black suspects.[45]

This account captures the different dimensions of oppression, and lays the foundations for a nuanced analysis of how specific groups experience oppression. Each dimension of oppression can intersect with, and be strengthened or weakened by, the other dimensions.[46] It should also be clear that oppression does not simply impede access to resources or rights; the various dimensions of oppression can also influence one's beliefs, desires, and emotions as well as one's ability to communicate, express, and articulate them. This is demonstrated by

---

40 Young, "Five Faces of Oppression," 63.

41 Cudd, *Analyzing Oppression*, 173–74.

42 Rhode, *The Beauty Bias*.

43 Lukes, *Power*; Patterson, *Slavery and Social Death*.

44 See, *inter alia*, Anderson, "What Is the Point of Equality?" and *Private Government*.

45 These are examples from the United States, but members of oppressed groups around the world are subject to repressive violence—e.g., religious and ethnically motivated violence (e.g., against the Rohingya in Myanmar), violence against gay and trans people, intimate partner violence against women, and violence against indigenous populations.

46 Crenshaw, "Demarginalizing the Intersection of Race and Sex"; hooks, *Feminist Theory*.

the phenomenon of adaptive preference, and is visible in the internalization of racial hatred and anger.[47]

Oppression has many consequences. We examine several dimensions along which oppression compromises the ability and opportunity of those within oppressed groups to successfully forgive wrongdoers. We begin with a fictionalized example of wrongdoing. In doing so, we recognize that the extent to which specific groups are oppressed, and the dimensions along which they are oppressed, is subject to debate. Moreover, we acknowledge that the dynamics of oppression change over time; a group may accrue power while suffering under tightening forms of cultural imperialism or remain marginalized despite being less subject to violence. Finally, we recognize that members of dominant groups are also harmed by oppression, including stereotyping and testimonial injustice—e.g., through the enforcement of restrictive masculinity norms and the dismissal of men's experiences of sexual harassment.

Ceasing to blame can be seen as the final stage of a typical blame scenario. Following Michael McKenna's conversational model of blame, we conceive of such scenarios as typically having four stages.[48]

1. Offense (X wrongs Y)
2. Address (Y addresses X about their wrongdoing)
3. Account (X responds to Y's address and accounts for their behavior)
4. Response (Y responds to X's accounting)

These stages might extend over time and involve multiple encounters. Consider the following concrete case of wrongdoing.

> *Maria's Arrest*: Maria is black American woman. In recent months, several members of her community have died at the hands of the police, without provocation, often in highly publicized contexts. Along with others in her community, Maria participates in peaceful protests of their mistreatment by the police. During one protest she is arrested, purportedly for obstructing traffic. In the course of her arrest she is physically abused, insulted, and not informed of her rights as a suspect, despite her calm insistence to be treated respectfully like privileged citizens.

Maria has been severely wronged. The consequences of her arrest could unfold in various ways, some better than others. The following describes a best-case scenario.

---

47 Stoljar, "Autonomy and Adaptive Preference Formation"; Lorde, "Eye to Eye," 147.

48 McKenna, *Conversation and Responsibility*, 89.

Best-Case Scenario: Maria's arrest and peaceful resistance were filmed and the videos are shown on the news and shared on social media. After her release, buoyed by the solidarity of her community, she publicly denounces her mistreatment and files legal charges against the arresting officer. With his case pending, the police officer contacts Maria via his lawyer. He has been following the media response to Maria's arrest and his actions, including the comments from members of her community, and he acknowledges that he acted wrongfully and shamefully. The officer offers a remorseful apology and makes clear that he accepts any legal consequences of his actions. Confident in his sincerity, Maria publicly forgives him and is lauded by her community.

We can describe the example in the vocabulary of McKenna's model:

1. Offense: Maria's arrest is injurious, disrespectful, and a violation of her rights.
2. Address: Maria denounces her treatment by the police officer and files charges.
3. Account: The officer accepts culpability, apologizes, and expresses remorse.
4. Response: Maria forgives the officer.

This best-case scenario seems unrealistic. The realities of oppression are likely to shape each of the scenario's four stages, producing outcomes markedly different from the best case. While our focus is on the ability of individuals to forgive—which takes place in the fourth stage—it is important to see how the different dimensions of oppression can shape how the whole episode unfolds. We will show that the ability to forgive can be compromised in various ways.

4.1. Offense

The circumstances of oppression affect the form and frequency of mistreatment that oppressed persons suffer. In turn, the nature of an offense can affect one's ability to forgive the offender. An oppressed person like Maria is more likely to suffer particular forms of wrongdoing in the first place. She is a member of a group that has historically been marginalized, exploited, culturally dominated, disempowered, and subjected to violence. The circumstances of oppression amplify the harm of such offenses—as when racial prejudice amplifies the harm of hostile or threatening speech.[49] Worse offenses are harder to forgive, especially when motivated by hatred or prejudice. We know that victims are less able or

---

49   Lawrence et al., "Introduction."

less willing to forgive more severe offenses.[50] And Hieronymi reminds us that forgiveness does not come easily in the best of circumstances: "The wrong is less 'let go of' or washed away than it is digested and absorbed."[51] People like Maria must digest and absorb both the ill will behind the offense and the attitudes that sustain such ill will and that perpetuate the frequency and regularity of such offenses—whether disregard, contempt, or hatred. The psychological difficulty of this task can directly compromise the ability of oppressed persons to forgive.

4.2. Address

Oppression can affect one's ability to blame. An oppressive society is one in which one's response to victimization will itself be shaped by oppressive norms and the apparent inevitability of mistreatment. We imagined Maria publicly denouncing her treatment and bringing charges against the officer involved. Blame of this sort can be a powerful response. However, the circumstances of oppression can compromise one's ability and willingness to blame. It can do this in a number of ways, but most importantly by making blame difficult and by requiring a sacrifice from the would-be blamer.

The historical legacy of oppression—especially the fact that the oppressed are more likely to be victimized and more likely to be victimized in particular ways—informs and influences Maria's responsiveness to offense.[52] People like Maria may come to expect treatment that a more privileged person would find unacceptable and, as a result, may fail to register an offense or may interpret it differently from a privileged person in the same circumstances. In some cases it will be difficult or costly to express blame. As a member of an oppressed group, Maria might face obstacles to using the criminal justice system, risk reprisal and further harm by blaming publicly, or be forced to communicate her blame in less effective ways. For example, if her employment is precarious, taking time off to pursue justice through the courts or through public protest may require too great a sacrifice. And oppressed persons are more likely to have precarious employment, in part because they are more likely to be uneducated and to work in sectors in which labor is informal, unskilled, and replaceable.[53] Maria may have an incentive to avoid being regarded as a "troublemaker" in order to keep her job.

Expressing blame may also be ineffective. The object of her blame may be

---

50   Fehr et al., "The Road to Forgiveness."

51   Hieronymi, "Articulating an Uncompromising Forgiveness," 551n39.

52   Young, "Five Faces of Oppression."

53   Matthew Desmond (*Evicted*) documents the difficulty of poor black tenants in the United States to oppose unjust evictions in court and the obstacles he identifies could easily apply in other contexts.

unclear or uncertain, directed at the police generally, or privileged whites, or her oppressive society, rather than at the arresting officer in particular. If Maria does blame publicly, her tone and manner may be cautious or subdued. On the one hand, she may be aware that she risks being stereotyped—e.g., being perceived as oversensitive, emotional, aggressive, or "uppity."[54] On the other hand, if she tries to minimize being seen in these ways, perhaps through the projection of extreme calm, she may struggle to make the force of her blame felt. More generally, the marginalization of her cultural perspective, idiolect, and values can hinder her ability to articulate to a privileged audience the meaning and significance of the wrongdoing she suffers.[55] Thus, she may struggle to communicate her blame authentically because she faces a trade-off between blaming in a manner that truly expresses her attitudes and doing so in a way that is likely to secure uptake.[56] Such double binds are a central feature of oppression and of dissent under oppression in particular.[57]

But expressing blame is not the only problem. Maria's ability to blame may itself be compromised. For example, she may be so numbed by the routine violence experienced by members of her community that she is unsurprised at being arrested on spurious grounds. As a result, she may not blame at all, responding instead with resignation. She might even reject her inclination to blame as unwarranted. Individuals from oppressed groups often internalize prevalent stereotypes and Maria might doubt the appropriateness of her anger or the force with which she feels it, perhaps because she has internalized a conception of objectivity according to which anger, or emotion in general, indicates an unreliably subjective response to others' actions.[58] These psychosocial effects of oppression make blaming more difficult and require a greater sacrifice by the victim in virtue of their greater psychic toll. Thus, Maria's status as a marginalized and relatively powerless individual shapes her ability to successfully blame, her willingness to do so, and the form her blame will take.

These systematic constraints mean that Maria is less able to blame, even if she has good reasons for doing so. If she does blame, she may do so privately, expressing her views to supportive members of her community. Of course, inability to blame does not imply inability to forgive except in the trivial sense that forgiving, understood as ceasing to blame, requires blaming. However, the

---

54   Banaji and Greenwald, *Blindspot*; McRae, "Anger and the Oppressed."

55   Fricker, *Epistemic Injustice*.

56   McRae, "Anger and the Oppressed," 109.

57   Frye, *The Politics of Reality*; McGowan, "Responding to Harmful Speech."

58   Lorde, "Eye to Eye"; Steele, *Whistling Vivaldi*; Banaji and Greenwald, *Blindspot*; Lugones, "Hard-to-Handle Anger"; Jaggar, "Love and Knowledge."

circumstances of oppression compromise would-be forgivers in similar ways as would-be blamers. Moreover, as we shall see, potential forgivers can fail to receive uptake of their attempts at forgiveness.

4.3. Accounting

Oppression can affect how an offender accounts for their offense. The beliefs and attitudes that give rise to particular offenses will also shape responses to blame for those offenses. Moreover, as we just described, the victim's response, or lack thereof, to mistreatment may also influence whether and how an offender accounts for their offense. How they do so can undermine the ability of the victim to forgive. First, the offender may give the victim no reason to forgive. They may fail to recognize the victim's blame or fail to adequately respond to the victim's moral address. The offender may refuse or fail to apologize, apologize for a different offense, or give a partial or limited apology. For example, the Japanese government recently apologized for its treatment of Korean "comfort women" during World War II, but only for the "involvement of Japanese military authorities." They thereby failed to take responsibility for the misdeeds of others acting on their behalf.[59] Likewise, when American radio personality Rush Limbaugh apologized to Sandra Fluke, a Georgetown law student, for calling her a "slut" after she testified before Congress about access to contraception, his public statement was closer to an excuse rather than an apology.[60] In both cases, the victims could reasonably view themselves as having no reason at all to forgive.

People like Maria are likely to find themselves in this kind of position. Members of oppressed groups are more likely to be victimized in particular ways, but their blame is less likely to be recognized and accepted.[61] An oppressive society is one in which the anger one feels in response to mistreatment is more likely to be ignored, doubted, or criticized, both by the offender and by society. An offender's response to blame can be shaped by the circumstances of oppression,

---

59   Tolbert, "Japan's Apology to South Korea Shows What Public Apologies Should (Not) Do."

60   For Limbaugh's public statement, see https://www.rushlimbaugh.com/daily/2012/03/03/a_statement_from_rush.

61   Members of otherwise privileged groups can be harmed in the same way. A man who has been sexually harassed may, as a result of sexist gender norms and stereotypes about male sexual desire, find himself unable to blame or forgive for the same reasons as an otherwise oppressed person. The offense may be minimized and his blame discounted or rejected. However, it is important to note that some dimensions of the harm will likely be absent for the otherwise privileged victim. For example, in the case of sexual harassment, the offense will not bear the weight of a history of sexual objectification and the experience of frequent harassment since adolescence. That said, intersectionality complicates this kind of case, too. Being a trans person or a male prisoner may increase a man's vulnerability to sexual harassment and to additional harms from it, including inability to blame or forgive.

especially (though not only) if the offender is comparatively privileged.[62] Ideal-ly, Maria's blame would prompt a sincere and unequivocal apology. The police officer would acknowledge his wrongdoing, not downplay its nature or extent, apologize, show remorse, and try to make amends. In fact, though, he may be unapologetically hostile, deny wrongdoing entirely, or simply (but culpably) fail to recognize the nature and significance of his actions. Or, if the police depart-ment recognizes or anticipates public outrage, it may manage his apology, there-by obscuring its sincerity. These alternative scenarios are more likely in contexts structured by implicit and explicit racism and norms that legitimize police vio-lence or minimize its significance. If the officer responds in these ways, Maria may have no reason to forgive him and thus be unable to do so.[63]

Moreover, inadequate responses can make it harder to forgive. As with the initial offense, even if she does have reason to forgive, the experience of being dismissed, discounted, or silenced may make forgiveness more difficult or emo-tionally burdensome.

4.4. Response

We have described how oppression can affect how Maria is treated, her experi-ence of mistreatment, whether and how she blames, and how others respond to her blame. And we have suggested that, at each stage, the realities of an op-pressive society can directly compromise Maria's ability to forgive—in particu-lar, her ability to meet the internal conditions on forgiving. In this section, we focus on Maria's ability to forgive her arresting officer and argue that oppression directly affects her ability to obtain uptake of her attempts to forgive.

In the ideal case, Maria has good reasons of the right kind to forgive. The offi-cer's remorse, apology, and willingness to make amends demonstrate a relevant change of heart. However, as we argued in section 3, this is not sufficient for suc-cessful forgiveness because one may fail to secure uptake. Maria may fail to se-cure uptake in two ways. First, because she is responding to an instance of wrong-doing that is perpetrated widely within her society, an instance of wrongdoing in which people typically lack good reasons to forgive, those in her community as well as privileged allies may mistakenly view her as condoning the police offi-cer's actions. Even if she has good reasons, Maria's wider community—plagued by endemic police violence and social marginalization—may fail to recognize the reasons she has and thus fail to acknowledge her forgiveness as forgiveness.

---

62 Medina, *The Epistemology of Resistance.*

63 She may still have reason to make room for forgiveness in the future—e.g., by explaining the extent and nature of the wrongdoing she suffered, or making it clear why the offender should have known better—but this is neither her responsibility nor a reason to forgive.

Second, others in comparatively privileged communities might recognize that she wants to forgive the officer, but reject her attempt as inappropriate. For example, privileged third parties may assume that police violence is justified or believe, on the basis of negative stereotypes of the criminality of black Americans, that in many cases police officers were justified. As a result, they may reject the possibility of situations in which it is reasonable to forgive a police officer for actions performed while on duty.

The same circumstances that impede uptake by others may also affect Maria's own attitude toward her forgiveness. Oppressed people often internalize inaccurate stereotypes of their character and social standing. They may also be burdened by frequent slights and signs of implicit prejudices, as well as by the struggles associated with pursuing their lives while being socially marginalized.[64] Claudia Card argues that, as a result:

> The oppressed are liable to low self-esteem, ingratiation, affiliation with abusers (for example so-called female masochism), as well as to a tendency to dissemble, fear of being conspicuous, and chameleonism—taking on the colours of our environment as protection against assault.[65]

Such factors may undermine Maria's confidence in how she relinquishes blame, even if she has good reasons to do so and is aware of those reasons. In other words, oppression might undermine self-uptake. One may express forgiveness but doubt whether one was right to do so; or one may excuse another's behavior only to question one's judgment in doing so. This uncertainty is distinct from, albeit potentially exacerbated by, failures of uptake within one's broader community. We might think that one could combat uncertainty through sustained consciousness raising and education, or by group activism, protest, and solidarity. However, the growing awareness of one's oppression can often spark doubt and doublethink rather than enhancing confidence.[66] One reason for this is that oppressed people are often blamed for their lack of standing and the wrongdoing they suffer.[67] Maria must not only weigh whether she has reasons to forgive,

---

64 See, among others, W. E. B. Du Bois's description of "double consciousness" (*The Souls of Black Folk*, 8).

65 Card, *The Unnatural Lottery*, 53.

66 Maria might also experience what Sandra Bartky ("Toward a Phenomenology of Feminist Consciousness") calls "double ontological shock," in which, for example, a woman may realize that sexism is pervasive, but remain unsure whether any particular behavior they experience is an instance of sexism.

67 Donovan, "To Blame or Not to Blame."

within a hostile context; she must also wrestle against the counter-narrative that she is responsible for her mistreatment and suffering at others' hands.

These narratives further complicate Maria's situation. She may have a good reason to relinquish blame and secure uptake from those around her, and yet still feel that the force of her interaction is somewhat diminished. In oppressive circumstances, wrongdoers need to do more when responding to blame and engaging with the victims' subsequent changing attitudes toward their blame than when the victim is not oppressed. The moral status of an oppressed victim is habitually called into question. Thus, while they may be able to blame and cease blaming for good and publicly accepted reasons, the responsiveness of wrongdoers can still seem inadequate unless they make special efforts to convey their understanding of the harms they have caused and remain sincere about making amends. For example, if it is likely that a normal apology will be interpreted as insincere, self-interested, or institutionally managed, the officer may need to apologize in a way that unequivocally accepts blame and expresses remorse.

We have shown how oppression can deprive individuals of reasons to forgive. Maria lives in a community that has been routinely exploited, marginalized, disempowered, and subjected to cultural domination by others, so she is likely to inherit strong social reasons to approach wrongdoing in certain ways. In particular, her victimization and the victimization of others like her, coupled with the experience of voicelessness—of being unable to secure the same political, cultural, or moral recognition as other groups—may give her reasons to let go rather than forgive. She may cease to blame simply in order to survive, as a mode of extended self-therapy. As we have shown, when offenders are not encouraged to think about their conduct toward members of socially disregarded groups, remorse, apology, and reasons to forgive will be rarer.

Finally, even if she is willing and able to forgive, oppression may also undermine its effectiveness. Her forgiveness may be ineffective or even counterproductive if her community believes that she has made a mistake or forgiven wrongfully. For example, Maria might misinterpret the significance of the officer's response. The officer may apologize, but do so in a superficial or equivocating way. Against a background of sustained oppression, such a tepid response may seem significant or even exceptional. In short, the significance of the officer's response to Maria's blame is likely to be distorted. Her assessment of his response and her response to it, in the concluding phase of the blame scenario, will be influenced by all of these factors and more, perhaps leading her to relinquish blame when she should not or in ways that others in her community would condemn or refuse to support.

Our previous points notwithstanding, we do not claim that people like Maria

are always irrevocably compromised in their ability or attempts to blame and cease blaming. Rather, we have argued that forgiving requires particular reasons and that oppression influences both the availability and significance of such reasons for members of some groups. As an oppressed person, Maria's life has been shaped by entrenched social norms and narratives about her marginal status. She has experienced frequent and persistent injustices and frustrations. A neglected dimension of this experience is that her attempts to participate in reason-guided practices of blaming and forgiving have often been significantly impeded. Moreover, because stereotypes and norms are often internalized, she may struggle to secure uptake from members of her own community and may even doubt herself. Finally, even when her reasons to relinquish blame are recognized, her actions may lack the force of comparable actions undertaken by the privileged. Her ability to respond to wrongdoing is compromised by the very norms and social structures that make her more likely to be a victim of wrongdoing in the first place.

5. OBJECTIONS AND REPLIES

We have argued that some people are systematically excluded from a variety of ceasing-to-blame practices, including forgiving. Before exploring the implications of this view, two objections demand attention. First, one might argue that, even if the circumstances of oppression can undermine some ways of ceasing to blame, forgiving is voluntary and personal and therefore open to anyone, irrespective of their social context. This objection says that whether a person forgives is genuinely up to her—that forgiveness is, by its nature, equally open to all.

We think that this view fails to acknowledge the realities of our actual practices. There are two ways to interpret the idea that forgiveness is "up to us" and on neither interpretation is it entirely so. First, forgiveness must be voluntary; forced or coerced forgiveness is not forgiveness.[68] But forgiveness is not entirely voluntary. Important aspects of the practice are beyond our control. We have only partial control over the emotional dimension of our blame, e.g., how angry we get when wronged and how persistent our anger is. It is also beyond our control whether and how others respond to our blame, e.g., whether our anger is ignored or judged excessive or otherwise held to be inappropriate. Nor can we control the social forces that shape forgiveness norms and interfere with our attempts to forgive, e.g., the power of forgiveness depends on how it is likely to be perceived by the offender and the broader community. As we argued above, all this implies that we cannot control the reasons we have to forgive, how strong those reasons are, or how they will be received.

68 Griswold, *Forgiveness*; Allais, "Wiping the Slate Clean"; Pettigrove, *Forgiveness and Love*.

Nor is forgiveness a purely personal exercise. Consider the claim that, independent of its supposed voluntariness, forgiveness is the exercise of a "normative power."[69] On this view, an offender acquires an obligation to her victim simply in virtue of having wronged him and forgiveness is a purely personal choice whereby the victim releases the offender from that obligation. But this conception fails to capture how our practice really works and how this "power" can be affected by one's circumstances.[70] One cannot forgive, by fiat or not, unless one has the right sort of reason for doing so.

Moreover, the power of one's forgiveness to alter the normative landscape depends, in part, on whether and to what degree it is recognized. Implicit biases and the ideological undercurrents of oppressive societies can structure the outlook of the privileged so that oppressed individuals are expected to forgive—perhaps, in some cases, religious ideas provide the relevant social pressure.[71] A form of ceasing to blame that is seen as supererogatory for the privileged can be normalized for the oppressed. And the result is a less powerful practice, one less able to alter the normative landscape surrounding the victim and her offender. A society that expects forgiveness drains it of its power in the same way that a society of liars drains promises of their power. Indeed, the act of forgiveness may have no power at all if those to whom it is addressed continue to deny any wrongdoing. One can commit, by fiat, to forgive someone, but commitment is only a first step in the forgiveness practice.

One might respond that a philosophical account of forgiveness should admit of versions of the practice that are not sensitive to reasons—that one can forgive, for instance, simply when one wants to. But this approach only serves to glorify an idiosyncratic experience of forgiving, which, for many people in oppressive societies, may lack the aura of power, personal choice, and moral status attributed to it by the privileged—or have the aura of power but not the substance.

This suggests a second objection. One could argue that our approach is insufficiently critical because we are describing practices that have been formed and sustained in contexts of oppression. Are we not taking existing practices and norms for granted and ignoring or denying the possibility that oppressed people can shape or create their own? One response would be to note that our task is not prescriptive. Our aim has been to describe the contexts in which people blame and cease to blame. We would go further, though. While we discuss the nature of

---

69 Warmke, "The Normative Power of Forgiveness."

70 MacLachlan, "Moral Powers and Forgivable Evils."

71 Moss and Watkins ("What Christians Get Wrong about Forgiveness") and Patton ("Black America Should Stop Forgiving White Racists") both point to instances of this phenomenon.

existing practices, we do not deny that oppressed people can shape these practices. Indeed, one way of opposing oppression is to resist cultural domination, one dimension of which is the inability to engage with the perspectives and forms of life of other groups.[72] It is a mistake to think that the ways people address blame, and the reasons they adduce to each other in the course of doing so, are immutable. And it would therefore be a mistake to suggest that the oppressed cannot acquire what the privileged already have—namely, full access to the variety of ceasing-to-blame practices, and especially to a truly powerful practice of forgiveness.

But it will not be easy. If we are right, then we have identified some of the ways oppression diminishes or distorts the ability of the oppressed to approach blame as the privileged do. They are hampered by social forces that shape the reasons available to them and the impact of those reasons on others in their society. Thus it is plausible to think that any attempt to reshape how people approach blame will itself be shaped by those same forces. Indeed, it is hard to imagine a context in which the attempt to forge new practices of dealing with blame—admitting new reasons to let go of blame, for instance—does not reinforce the divisions we have already described, thereby amplifying the assumptions and biases of the privileged, and making it less likely that they will recognize or accept attempts by oppressed individuals to forgive, excuse, justify, and let go. Consider the resistance to adopting traditionally plural pronouns to express gender neutrality or trans identity, e.g., "They [singular] went to the dance without a date."

It is nearly a truism that once a power imbalance is established it is easier to perpetuate the disparity than to reverse it. Power reproduces itself. Various instances of oppression support the hypothesis that it is typically easier to perpetuate the social institutions that oppress a group than to dismantle them, in part because the existing state of affairs is supported by a web of forces that are both entrenched and mutually reinforcing.

6. IMPLICATIONS

We have argued that oppression shapes our ceasing-to-blame practices, especially forgiveness. Understanding this matters for how we perceive and perform them as well as for how we develop and reform them. We conclude by discussing two ways in which our analysis bears on our perception of and plans for the practice of forgiveness.

First, our analysis highlights a neglected dimension of oppression and the political contexts in which oppression is generated and maintained. Not only does

---

72 For one account of how the circumstances of oppression can both undermine agency and create opportunities for novel exercises of agency, see Bierra, "Missing in Action."

oppression expose some groups to systematic forms of wrongdoing, it also compromises their ability to respond to such wrongdoing by blaming and forgiving.

Our analysis may seem trivial, a philosophical curiosity describing a single straw on the backs of the burdened, but it has an important practical consequence. If we are right that the various ways of ceasing to blame constitute a set of reason-guided practices, then those who aim to facilitate and develop ceasing-to-blame practices—especially those who advocate developing forgiveness institutions—must reassess their approach. For example, within the burgeoning literature on political forgiveness, theorists often discuss the possibility and usefulness of developing institutional frameworks to help foster forgiveness. Truth and reconciliation commissions and procedures aimed at interpersonal restorative or transformative justice are two such approaches.[73] These projects recognize the power of forgiveness and explore whether political forgiveness could enable a society to better address entrenched social conflicts and their related injustices.

Forgiveness can often be a powerful response to wrongdoing, including wrongdoing in an oppressive context. However, our argument suggests that attempts to develop forgiveness institutions may generate similar problems to the interpersonal ones we describe above. These attempts may amplify the effect of oppression by further exaggerating the distorting norms and ideologies that underpin privilege. It may be harder to secure uptake for other ways of ceasing to blaming if one is expected to forgive, and harder still if there is an institutional practice that disproportionately valorizes forgiveness.

More importantly, though, even if forgiveness institutions are not harmful in the ways we suggest, our argument implies that if we are truly interested in helping people forgive (or excuse, or justify, or let go), then we must address their wider ability and opportunities to relinquish blame. Someone is better placed to forgive if they can access other forms of redress. They must have access, in ways that many plausibly do not, to the wider nexus of institutions that the privileged utilize when evaluating blame and responsibility—e.g., the formal legal system—as well as informal blaming and ceasing-to-blame practices. Forgiveness is not a panacea, but it can be good and useful. The most effective forgiveness institution is a non-oppressive society in which everyone has adequate and equal ability and opportunity to blame and cease to blame when they suffer wrongdoing.

Second, our account illuminates a further parallel between forgiveness and other practices shaped by oppression. The practice of forgiveness is, among other things, a tool. It is a means by which a victim can resolve a moral conflict, just as blame is a means of protesting mistreatment and demanding due regard. And

---

73 Amstutz, *The Healing of Nations*; Johnstone, *Restorative Justice*; MacLachlan, "The Philosophical Controversy over Political Forgiveness."

those who forgive recognize its function and power. Indeed, we view it as a powerful response to those who have earned our blame. However, this tool has been blunted by oppression with the result that, for those who need it most, those most likely to be systematically victimized, it will be least effective.

The ability to grant or withhold forgiveness from those who mistreat them may appear to be one of the few forms of power and moral high ground not eroded by oppression. But this is not the case. Our account shows that even this most personal form of moral interaction can be compromised. Forgiveness is difficult in the best of circumstances—coming to see an offender as better than his wrongdoing warrants entails first seeing him as the person who chose to wrong you—and this burden is increased when combined with (on the one hand) prejudice and (on the other) norms that urge forgiveness. The result is that forgiveness can even be turned against those for whom it represents a significant exercise of moral agency.

This phenomenon is not unique to forgiveness. It is a familiar feature of oppression that the tools (i.e., concepts, norms, practices) employed by the privileged do not work for, and sometimes work against, the oppressed. Consider so-called feminine virtues like chastity, patience, and obedience. These virtues—separate from but supposedly equal to masculine virtues—were promoted and accepted as essential for the good life of a woman. In fact, valuing, instantiating, and policing the feminine virtues was and is, at least in a male-dominated world, a way of perpetuating the oppression of women.

The same can be said of freedom. Nominal freedom, of the sort delineated by a constitution or bill of rights, is equally available to all citizens, but the circumstances of oppression quickly show how material conditions determine the value and power of those freedoms.[74] The freedoms guaranteed to citizens by such documents do not serve all citizens equally well, as is made clear by the history of disenfranchisement, housing discrimination, and police violence suffered by blacks in the United States, the perpetuation of caste norms in India, and continuing struggles for equal rights by LGBTQ communities in many democracies. These freedoms are the tools with which the disadvantaged are expected to escape their material deprivation, but they are being asked to climb higher with a shorter ladder. The valorization of forgiveness involves a similar bait and switch. Forgiveness has an aura of power because, for the privileged, it is powerful. But for those living under oppression, what seems like a powerful tool—something with which to defend oneself and assert one's self-respect, virtue, and strength—is found to be inadequate. This mismatch between the ideal and reality is dangerous. It makes misuse and misunderstanding more likely.

74 Nussbaum, *Women and Human Development*.

Imagine an oppressed person, like Maria, considering whether to forgive a racist offender. Whether she recognizes it or not, her ability to forgive is likely to be compromised—perhaps the offender views his offense as insufficiently bad to warrant blame. If she does not recognize this, registering only a vague distaste about forgiving him, then she may unwittingly harm herself (e.g., by undermining the demand for respect signaled by her blame) and undermine the practice itself (e.g., by perpetuating the norm of forgiving the unrepentant). But even if she does recognize that she cannot forgive or that forgiving is somehow inappropriate and therefore likely to be less effective, she may try to forgive anyway simply because it drives her point home as powerfully as she can. The fact that it further undermines the power of the practice by distorting it is unfortunate but unavoidable. To turn a screw, one wants a screwdriver. Lacking one, a knife may do, but it is unlikely to sink the screw properly, and liable to chip or bend.

Forgiveness can seem like a powerful practice to which the oppressed, as regular victims of wrongdoing, have unimpeded access. In fact, the opposite is the case. The oppressed are less able to forgive and less able to do so successfully. Moreover, because of the mismatch between its perceived power and its actual effectiveness, forgiveness is likely to be misused by those who need it and misunderstood by those who "police" it, whether as recipients of forgiveness or bystanders to the practice. Rather than helping oppose oppression, forgiveness may stem from, reflect, and exacerbate it.[75]

University of Oxford
dr.luke.brunning@gmail.com

University of Twente
per.milam@gmail.com

REFERENCES

Abramson, Kate, and Adam Leite. "Love as a Reactive Emotion." *Philosophical Quarterly* 61, no. 245 (2011): 673–99.

Allais, Lucy. "Elective Forgiveness." *International Journal of Philosophical Studies* 21, no. 5 (2013): 637–53.

———. "Wiping the Slate Clean: The Heart of Forgiveness." *Philosophy and Public Affairs* 36, no. 1 (Winter 2008): 33–68.

Amstutz, Mark. *The Healing of Nations: The Promise and Limits of Political Forgiveness*. New York: Rowman and Littlefield, 2004.

Anderson, Elizabeth. *Private Government: How Employers Rule Our Lives (and Why We Don't Talk about It)*. Princeton: Princeton University Press, 2017.

———. "What Is the Point of Equality?" *Ethics* 109, no. 2 (January 1999): 287–337.

Banaji, Mahzarin R., and Anthony G. Greenwald. *Blindspot: Hidden Biases of Good People*. New York: Bantam Books, 2013.

Bartky, Sandra. "Toward a Phenomenology of Feminist Consciousness." *Social Theory and Practice* 3, no. 4 (Fall 1975): 425–39.

Bennett, Christopher. "Personal and Redemptive Forgiveness." *European Journal of Philosophy* 11, no. 2 (August 2003): 127–44.

Bierra, Alisa. "Missing in Action: Violence, Power, and Discerning Agency." *Hypatia* 29, no. 1 (Winter 2014): 129–45.

Butler, Joseph. *The Works of Bishop Butler*. London: Macmillan, 1900.

Calhoun, Cheshire. "Changing One's Heart." *Ethics* 103, no. 1 (October 1992): 76–96.

———. "Responsibility and Reproach." *Ethics* 99, no. 2 (January 1989): 389–406.

Card, Claudia. *The Unnatural Lottery: Character and Moral Luck*. Philadelphia: Temple University Press, 1996.

Coates, D. Justin, and Neal A. Tognazzini. "The Nature and Ethics of Blame." *Philosophy Compass* 7, no. 3 (March 2012): 197–207.

Cornell, Nicolas. "The Possibility of Preemptive Forgiving." *Philosophical Review* 126, no. 2 (April 2017): 241–72.

Crenshaw, Kimberlé. "Demarginalizing the Intersection of Race and Sex: A Black Feminist Critique of Anti-Discrimination Doctrine, Feminist Theory and Antiracist Politics." *University of Chicago Legal Forum* 1989, no. 1 (1989): 139–67.

Cudd, Ann E. *Analyzing Oppression*. Oxford: Oxford University Press, 2006.

Desmond, Matthew. *Evicted: Poverty and Profit in the American City*. New York: Broadway Books, 2017.

Driver, Julia. "Private Blame." *Criminal Law and Philosophy* 10, no. 2 (June 2016): 215–20.

Donovan, R. A. "To Blame or Not to Blame: Influences of Target Race and Observer Sex on Rape Blame Attribution." *Journal of Interpersonal Violence* 22, no. 6 (June 2007): 722–36.

Du Bois, W. E. B. *The Souls of Black Folk*. 1903. Edited by Brent Hayes Edwards. Oxford: Oxford University Press, 2008.

Fehr, Ryan, Michele J. Gelford, and Monisha Nag. "The Road to Forgiveness: A Meta-Analytic Synthesis of Its Situational and Dispositional Correlates." *Psychological Bulletin* 136, no. 5 (2010): 894–914.

Fricker, Miranda. *Epistemic Injustice: Power and the Ethics of Knowing*. Oxford: Oxford University Press, 2007.

———. "What's the Point of Blame? A Paradigm Based Explanation." *Noûs* 50, no. 1 (March 2016): 165–83.

Frye, Marilyn. *The Politics of Reality: Essays in Feminist Theory*. Trumansburg, NY: Crossing Press, 2007.

Garcia, Ernesto V. "Bishop Butler on Forgiveness and Resentment." *Philosophers' Imprint* 11, no. 10 (August 2011): 1–19.

Garrard, Eve, and David McNaughton. *Forgiveness*. Durham, UK: Acumen Publishing, 2010.

———. "In Defence of Unconditional Forgiveness." *Proceedings of the Aristotelian Society* 103 (2003): 39–60.

Griswold, Charles. *Forgiveness: A Philosophical Exploration*. Cambridge: Cambridge University Press, 2007.

Haslanger, Sally. "Oppressions: Racial and Other." In *Resisting Reality: Social Construction and Social Critique*, 311–40. Oxford: Oxford University Press, 2012.

Hieronymi, Pamela. "Articulating an Uncompromising Forgiveness." *Philosophy and Phenomenological Research* 62, no. 3 (May 2001): 529–55.

———. "The Force and Fairness of Blame." *Philosophical Perspectives* 18 (2004): 115–48.

———. "The Reasons of Trust." *Australasian Journal of Philosophy* 86, no. 2 (2008): 213–36.

hooks, bell. *Feminist Theory: From Margin to Center*. New York: Routledge, 2015.

Jaggar, Alison M. "Love and Knowledge: Emotion in Feminist Epistemology." *Inquiry* 32, no. 2 (1989): 151–76.

Johnstone, Gerry. *Restorative Justice: Ideas, Values, Debates*. London: Routledge, 2011.

Larsen, Britta, Ryan S. Darby, Christine R. Harris, Dana Kay Nelkin, Per-Erik

Milam, and Nicholas J. S. Christenfeld. "The Immediate and Delayed Cardiovascular Benefits of Forgiving." *Psychosomatic Medicine* 74 (2012): 745–50.

Lawrence, Charles R., Mari J. Matsuda, Richard Delgado, and Kimberlé Williams Crenshaw. "Introduction." In *Words That Wound: Critical Race Theory, Assaultive Speech, and the First Amendment*, 1–16. New York: Westview Press, 1993.

Lorde, Audre. "Eye to Eye: Black Women, Hatred, and Anger." In *Sister Outsider*, edited by Nancy K. Bereano, 124–33. New York: Crossing Press, 2007.

Lugones, Maria. "Hard-to-Handle Anger." In *Pilgrimages/Peregrinajes: Theorising Coalition Against Multiple Oppressions*, 103–18. Lanham, MD: Rowman and Littlefield, 2003.

Lukes, Steven. *Power: A Radical View*. London: MacMillan, 1974.

MacLachlan, Alice. "Moral Powers and Forgivable Evils." In *Evil, Political Violence, and Forgiveness: Essays in Honor of Claudia Card*, edited by Andrea Veltman and Kathryn J. Norlock, 135–57. New York: Lexington Books, 2009.

———. "The Philosophical Controversy over Political Forgiveness." In *Public Forgiveness in Post-Conflict Contexts*, edited by Bas van Stokkom, Neelke Doorn, and Paul van Tongeren, 37–62. Cambridge: Intersentia Press, 2012.

Macnamara, Colleen. "Blame, Communication, and Morally Responsible Agency." In *The Nature of Moral Responsibility*, edited by Randolph Clarke, Michael McKenna, and Angela M. Smith, 211–36. Oxford: Oxford University Press, 2015.

———. "Holding Others Responsible." *Philosophical Studies* 152 (January 2011): 81–102.

McRae, Emily. "Anger and the Oppressed: Indo-Tibetan Buddhist Perspectives." In *The Moral Psychology of Anger*, 105–22. Lanham, MD: Rowman and Littlefield, 2018.

McGary, Howard. "Forgiveness." *American Philosophical Quarterly* 26, no. 4 (1989): 343–51.

McGowan, Mary Kate. "Responding to Harmful Speech: The More Speech Response, Counter Speech, and the Complexity of Language Use." In *Voicing Dissent: The Ethics and Epistemology of Making Disagreement Public*, edited by Casey Rebecca Johnson, 182–99. New York: Routledge, 2018.

McKenna, Michael. *Conversation and Responsibility*. Oxford: Oxford University Press, 2012.

———. "Quality of Will, Private Blame and Conversation: Reply to Driver, Shoemaker, and Vargas." *Criminal Law and Philosophy* 10, no. 2 (June 2016): 243–63.

Medina, José. *The Epistemology of Resistance: Gender and Racial Oppression, Epis-

temic Injustice, and the Social Imagination. London: Oxford University Press, 2013.

Milam, Per-Erik. "Against Elective Forgiveness." *Ethical Theory and Moral Practice* 21, no. 3 (June 2018): 569–84.

———. "Reasons to Forgive." *Analysis*. Published ahead of print, June 15, 2018. https://doi.org/10.1093/analys/any017.

Moss, Candida, and Christa Grace Watkins. "What Christians Get Wrong about Forgiveness." *Daily Beast*, October 4, 2015. https://www.thedailybeast.com/what-christians-get-wrong-about-forgiveness.

Murphy, Jeffrie. "Forgiveness and Resentment." *Midwest Studies in Philosophy* 7, no. 1 (September 1982): 503–16.

Neblett, William R. "Forgiveness and Ideals." *Mind* 83, no. 330 (April 1974): 269–75.

Nelkin, Dana K. "Freedom and Forgiveness." In *Free Will and Moral Responsibility*, edited by Ishtiyaque Haji and Justin Caouette, 165–88. Newcastle: Cambridge Scholars Press, 2013.

Novitz, David. "Forgiveness and Self-Respect." *Philosophy and Phenomenological Research* 58, no. 2 (June 1998): 299–315.

Nussbaum, Martha C. *Women and Human Development*. Cambridge: Cambridge University Press, 1999.

Palitza, Kristin. "Child Labour: The Tobacco Industry's Smoking Gun." *Guardian*, September 14, 2011. https://www.theguardian.com/global-development/2011/sep/14/malawi-child-labour-tobacco-industry.

Patterson, Orlando. *Slavery and Social Death*. Cambridge, MA: Harvard University Press, 1982.

Patton, Stacey. "Black America Should Stop Forgiving White Racists." *Washington Post*, June 22, 2015. https://www.washingtonpost.com/posteverything/wp/2015/06/22/black-america-should-stop-forgiving-white-racists/.

Pettigrove, Glen. *Forgiveness and Love*. Oxford: Oxford University Press, 2012.

Rhode, Deborah L. *The Beauty Bias*. New York: Oxford University Press, 2010.

Scanlon, T. M. *Moral Dimensions: Permissibility, Meaning, Blame*. Cambridge, MA: Harvard University Press, 2008.

Sher, George. *In Praise of Blame*. Oxford: Oxford University Press, 2006.

Smith, Angela. "Moral Blame and Moral Protest." In *Blame: Its Nature and Norms*, edited by D. Justin Coates and Neal A. Tognazzini, 27–48. Oxford: Oxford University Press, 2013.

Smuts, Aaron. "Normative Reasons for Love, Part 1." *Philosophy Compass* 9, no. 8 (August 2014): 507–17.

Steele, Claude M. *Whistling Vivaldi: How Stereotypes Affect Us and What We Can Do*. New York: W. W. Norton and Company, 2010.

Stoljar, Natalie. "Autonomy and Adaptive Preference Formation." In *Autonomy, Oppression, and Gender*, edited by Andrea Veltman and Mark Piper, 227–52. Oxford: Oxford University Press, 2014.

Tolbert, David. "Japan's Apology to South Korea Shows What Public Apologies Should (Not) Do." *Huffington Post*, December 6, 2017. https://www.huffingtonpost.com/david-tolbert/japans-apology-to-south-k_b_9111566.html.

Twambley, P. "Mercy and Forgiveness." *Analysis* 36, no. 2 (January 1976): 84–90.

Vargas, Manuel. *Building Better Beings*. Oxford: Oxford University Press, 2013.

Walker, Margaret Urban. *Moral Repair: Reconstructing Moral Relations after Wrongdoing*. Cambridge: Cambridge University Press, 2006.

———. "Third Parties and the Social Scaffolding of Forgiveness." *Journal of Religious Ethics* 41, no. 3 (September 2013): 495–512.

Wallace, R. Jay. "Dispassionate Opprobrium." In *Reasons and Recognition: Essays on the Philosophy of T. M. Scanlon*, edited by R. Jay Wallace, Rahul Kumar, and Samuel Freeman, 348–69. Oxford: Oxford University Press, 2011.

———. *Responsibility and the Moral Sentiments*. Cambridge, MA: Harvard University Press, 1994.

Warmke, Brandon. "The Economic Model of Forgiveness." *Pacific Philosophical Quarterly* 97, no. 4 (December 2016): 570–89.

———. "The Normative Significance of Forgiveness." *Australasian Journal of Philosophy* 94, no. 4 (2016): 687–703.

Wolf, Susan. "Blame, Italian Style." In *Reasons and Recognition: Essays on the Philosophy of T. M. Scanlon*, edited by R. Jay Wallace, Rahul Kumar, and Samuel Freeman, 332–47. Oxford: Oxford University Press, 2011.

———. "The Importance of Free Will." *Mind* 90, no. 359 (July 1981): 386–405.

Young, Iris Marion. "Five Faces of Oppression." In *Geographic Thought: A Praxis Perspective*, 55–71. New York: Routledge, 2009.

Zaibert, Leo. "The Paradox of Forgiveness." *Journal of Moral Philosophy* 6, no. 3 (2009): 365–93.