Evolutionary Debunking of (Arguments for) Moral Realism

Arnon Levy & Itamar Weinshtock-Sa'adon

The final version of this paper is forthcoming in Forthcoming in Synthese, please cite published version.

Abstract

Moral realism is often taken to have common sense and initial appearances on its side. Indeed, by some lights, common sense and initial appearances underlie all the central positive arguments for moral realism. We offer a kind of debunking argument, taking aim at realism's common sense standing. Our argument differs from familiar debunking moves both in its empirical assumptions and in how it targets the realist position. We argue that if natural selection explains the objective phenomenology of moral deliberation and judgement, then this undermines arguments from that phenomenology. This results in a simpler, and in some ways more direct, challenge to realism. It is also less vulnerable to the main objections that have been leveled against such arguments. If accepted, our conclusion should make a real difference to the dialectic in this area. It means that neither realism nor its denial is the default, to-be-refuted, position.

1. Introduction

Moral realists hold that there are genuine, objective moral truths. Early on in his Stanford Encyclopedia of Philosophy entry on realism, Geoff Sayer-McCord notes that this view "can fairly claim to have common sense and initial appearances on its side." (2015). Indeed, by some lights, "common sense and initial appearances" underlie all the central positive arguments for moral realism.

Here we offer a novel argument against moral realism—more precisely, a counter to existing arguments for realism—taking aim at its common sense standing.[1] Our argument is a kind of evolutionary debunking argument, although it differs from familiar arguments of this sort both in

---

[1] We know of a few partial, scattered, precedents: Ruse (2009) sometimes seems to make an argument akin to ours, although he can be interpreted in different ways. Jeroen Hopster mentions a closely related argument, but only devotes a single, relatively brief paragraph to it (2019, 848-9). Loeb (2007, 476) and Björnsson (2012, 376) attribute such an argument to Joyce (2001, Ch. 6; 2006, Ch. 6). But as we explain in section 5.1, there are significant differences between our argument and Joyce's.

its empirical assumptions and in the manner in which it targets the realist position. In a nutshell, we suggest that if natural selection explains *the objective phenomenology* of moral deliberation and judgement, then arguments from that phenomenology are found wanting. This presents a simpler, and in some ways more direct, challenge to realism, relative to previous debunking arguments. Previous attempts at debunking allege that combining realism with evolutionary explanations of our moral psychology leads to first-order moral skepticism. Our line of reasoning does not pass through a skeptical way station.

Some arguments for realism we will discuss embody a fairly simple move from the objective phenomenology of morality to its metaphysical nature. Others take such phenomenology to play a subtler, motivating role. We shall suggest that evolutionary accounts of moral phenomenology, if true, pull the rug out from under both types of argument.

The debunking move we offer is directed at a key assumption of arguments for realism. This allows us to stay relatively non-committal on the exact construal of the realist position. In particular, we take the core of moral realism to be a metaphysical claim, as stated above, namely that there are objective (i.e., mind- or stance-independent) moral truths. It is this claim that arguments from objective phenomenology seek to establish and that our argument targets. Further, except for a minor wrinkle (see footnote 5), we remain neutral between naturalist and non-naturalist versions of realism. So long as the realist relies on our objective moral phenomenology, our argument kicks in. Realists typically also assume cognitivism, i.e., that our moral judgements and discourse are truth-apt. We are unsure whether cognitivism is strictly necessary for our argument. But in order not to beg any questions, and because this figures, to an extent, in the evolutionary work we rely on, we will assume that moral judgements express propositions and that they can be straightforwardly true or false.

Following Kahane (2011), one can distinguish between the empirical and the epistemic elements of a debunking argument. Empirically, such arguments rest on an explanation of some facet(s) of our moral psychology. We rely on evolutionary explanations of morality's objective phenomenology—as we elaborate in the next section. Epistemically, debunking arguments can be filled-in in different ways—by appeal to truth-tracking (Kahane 2011), to sensitivity and/or safety (Clark-Doane 2016; Handfield 2016), to accidentality (Street 2006; Bedke 2009), or by appealing to whether reasons are (or are not) needed to rationally maintain one's "natural" opinions (Vavova 2014). In presenting our argument (sections 3 and 4), we remain relatively coy on this score, as

different ways of understanding the epistemic element all dovetail, so far as we can see, with our argument. Section 5 provides a comparison of our proposal with existing attempts at evolutionary debunking—explaining how and why it is novel. Section 6 discusses the overall argument's implications, given the central role that appeals to phenomenology play in debates over realism.

2. *Explanations of morality's objective phenomenology*

Some recent work in empirical moral psychology has looked at "folk meta-ethics", i.e., attitudes *about* first-order ethical attitudes. A central issue, in this context, is whether and under what circumstances morality is experienced as 'objective' (other terms include 'real' and 'factual'). In philosophical discussions, the objectivity of morality is typically understood in terms of *mind- (or stance-)independence*—roughly, the idea that moral statements or propositions are true or false irrespective of the beliefs, feelings or other states of minds of the agents making or considering them.

Psychologists have investigated several features of folk attitudes, some more closely related to objectivity in the philosophical sense just noted, others less so. Goodwin and Darley (2008, 2012) are widely regarded as having probed folk moral realism, in a sense that closely matches the one that dominates philosophical discussion. These authors tested for objectivity by looking at attitudes towards disagreement, and by testing whether subjects saw moral statements as truth apt. On the first score, they found that some moral statements—especially those with which subjects evinced strong agreement—were such that subjects found it hard to accept that both sides to a disagreement could be correct (and tended to think their view was the correct one). On the second score, they found that many subjects tended to view moral statements as capable of being true or false, and not as expressions of "opinion or attitude". Goodwin & Draley's conclusion was stark: "ethical beliefs were treated almost as objectively as scientific or factual beliefs, and decidedly more objectively than social conventions or tastes. Individuals seem to identify a strong objective component to their core ethical beliefs, and thus treat them as categorically different from social conventions." (*ibid*, 1359). It should be noted, however, that we are not arguing that "the folk" are, either in general or in a strong way, objectivist. Moreover, it has been suggested that Goodwin and Darley's results, as well as some related work, is best seen as supporting "folk metaethical pluralism", i.e. as regarding some moral judgements as objective and others less so (Pölzler & Wright 2019). Following Stanford (2018) we therefore think that the phenomenon should be seen

not so much as a blanket folk realism concerning morality, but as a tendency to view *some* moral norms, especially strongly held ones, as objective—in the sense that that if two parties to differ with respect to their truth or falsity then one party is typically seen as being in error.

Corresponding to the empirical findings just discussed are evolutionary explanations of our tendency to experience morality as objective[2], cast in terms of natural selection. In the existing literature, two directions can be discerned. The first concerns personal moral motivation whereas the second centers on social coordination. Richard Joyce (2006, Ch. 4) advocates a hypothesis of the first kind, according to which the seeming objectivity of morality was selected for because it allowed humans to overcome temptations that would, if succumbed to, destroy our ability to cooperate and function in society. This is especially so, thinks Joyce, inasmuch as morality is seen as "external" to the agent, or as imposed on her, which he understands as very close to its seeming mind-independent. In this manner, "moral judgments can act as effective personal commitments better than mere inhibitions, providing a kind of motivational bulwark" (*Ibid*, 121).[3]

Joyce also advances a further hypothesis, according to which moral judgements, when publicly expressed, can serve as intrapersonal commitment devices (*Ibid*, subsection 4.3). Here, the idea is that in voicing a moral opinion one displays one's firm determination to act a certain way. Such a public display affects others' decisions, promoting joint cooperative action. This further aspect of Joyce's theory brings it closer to the second genre of evolutionary explanations, focusing on social coordination. Before describing this latter option, let us note that Joyce's suggestions about the origins of morality form the basis of an overall case for anti-realism, which is perhaps better known to philosophers than his specifically evolutionary hypotheses. Part of the overall case takes the form of an evolutionary debunking argument. Here we confine ourselves to his explanatory proposals. We will come back to Joyce's debunking argument and its relation to our view in subsection 5.1.

---

[2] Some advocates of evolutionary explanations have in mind the explanandum as we have just described it. Others are less clear on this score. As we have followed Stanford's (2018) characterization of the explanandum, we also follow him in assessing to what extent explanations on offer can account for said explanandum.

[3] A related suggestion is found in Dennett (1995). His idea is that morality serves as a "conversation stopper", contributing to efficient social decision making. Dennett alludes to the objective phenomenology of moral thinking, but his primary emphasis is on we he describes as its "seriousness" – roughly speaking, that moral injunctions have overriding force.

The social coordination hypothesis is motivated by a central insight of work on the evolution of cooperation over the last half century, namely that for cooperation to have a selective advantage, interactions among cooperators must be positively correlated (Skyrms 2004). A number of theorists have suggested that experiencing morality as objective is fitness enhancing insofar as it directs an individual to choose partners that are likely to share her cooperative behavioral tendencies. In this vein, DeScioli and Kurzban (2013) suggest that moral judgement (in particular, moral condemnation) evolved as a signal that allows bystanders to a conflict to choose sides without excess costs.

More recently, Kyle Stanford (2018) has suggested that, evolutionarily speaking, humans experience morality as objective—primarily in the mind-independence sense described above, although Stanford also seems to think this goes along with wide scope—because this guaranteed that humans could, on the one hand, cooperate in a flexible and open-ended way while, on the other hand, minimizing the potential for exploitation of cooperators. In Stanford's view, this mechanism allowed humans to engage in "hypercooperative" social interactions, thereby facilitating the spread of humans into diverse environments and the development of an ever-more-complex social life style. Note that for Stanford's hypothesis to work, it is not necessary for all moral norms to seem objective—and this is in good agreement with psychological findings, as described above.

Stanford (*Ibid*, section 3) surveys a number of previous attempts to account for morality's objective phenomenology, including those we've discussed, and raises compelling objections to them. We find both his positive proposals and the criticism of rival theories to be quite plausible, and we cautiously accept both. What we wish to highlight is that neither according to Stanford nor on the other proposals is it the case that morality's objective phenomenology is explained by its genuine objectivity. The fitness advantages that an objective phenomenology confers stem from its effects on the agent's motivations, or on the efficiency of social cooperation, or on the avoidance of exploitation. These effects come about regardless of whether there are objective moral truths.

Let us be as clear as we can on how we view the theoretical-explanatory situation: We think there is a phenomenon to be explained, namely the tendency humans have to experience some moral norms as mind-independently true, and in this sense as objective. We will refer to this below as morality's "objective phenomenology". We also tentatively endorse Stanford's proposed explanation for this phenomenon. However, we are well aware that theoretical hypotheses in this

area are very hard to settle, empirically. We also suspect that existing hypotheses are likely to be augmented and refined, as work continues. So our endorsement is cautious, and our overall argument can, in this sense, be seen as conditional in form: it addresses the metaethical upshots of the truth of Stanford's explanation, or some related evolutionary hypothesis explaining morality's objective phenomenology. After we have laid out our argument, we will revisit some of these empirical questions, in subsection 5.2.

3. Argument(s) from moral phenomenology

We now discuss arguments for realism, beginning with a simple but important argument, which moves from the observation that we experience morality as objective to the claim that morality is, in fact, objective. It takes somewhat different forms (Loeb 2007), of which we will tackle two. *The face-value argument* (FVA) says, in essence, that the fact that morality seems objective gives rise to a presumption in favor of it being objective. Meanwhile, *the explanationist argument* (EA) suggests that objective moral facts are the best explanation for the objective phenomenology of morality.

Arguments from phenomenology take as their point of departure a set of observations about our moral practice, namely that it "looks and feels" like a practice directed at external, mind-independent facts. These arguments do not, typically, motivate such observations by reference to empirical findings (most of them predate the empirical literature) but we think they are aiming at a closely related phenomenon. Proponents often say that when we judge, for instance, that stealing or breaking promises is wrong, we experience ourselves as *responding to the objective fact* that such actions are wrong (e.g., Brink 1989, 25-6; Dancy 1986, 172). Likewise, they depict moral deliberation as an attempt to *discover* the right course of action. Jonathan Dancy, for instance, speaks in this context of "the struggle to find, not any answer that we can bring ourselves to accept, nor any answer that we can accept in consistency with previous answers, but the right answer. We present our search to ourselves as one governed by a criterion which does not lie in ourselves" (1986, 172). Even more clearly, the experience of moral disagreement is depicted as akin to factual disagreements, in which there are objectively right or wrong answers (Brink 1989, 29).

So the first step of the argument is a phenomenological description, a description of how moral thought and talk feels (which corresponds fairly well to empirical findings). To move from this to a metaphysical conclusion regarding moral reality, a bridging principle is required. The two

6

versions we will discuss differ on this score. The first appeals to *Phenomenal Conservatism*. The second is an inference to the best explanation.

3.1. The face value argument

Jonathan Dancy argues as follows:

> We take moral value to be part of the fabric of the world; taking our experience at face value, we judge it to be the experience of the moral properties of actions and agents in the world. And if we are to work with the presumption that the world is the way our experience represents it to us as being, we should take it in the absence of contrary considerations that actions and agents do have the sorts of moral properties we experience in them. This is an argument about the nature of moral experience, which moves from that nature to the probable nature of the world (Dancy 1986, 172).

Dancy is suggesting that we take our moral experience at face value, relying on the idea that, at least in a probable and presumptive manner, "the world is the way our experience represents it to us as being." Michael Huemer formulates this more carefully as a principle of *phenomenal conservatism*: "If it seems to *S* that *p*, then, in the absence of defeaters, *S* thereby has at least some degree of justification for believing that *p*" (2007, 30). Here, 'it seems' is taken to include either perceptual or intellectual seemings. Thus, this version of the argument moves from the claim that morality *seems to us* objective, via the principle of phenomenal conservatism, to the conclusion that there is some justification to believe that morality *is objective*, i.e., to believe in moral realism.[4]

Note the cautious character of Huemer's formulation. The principle of phenomenal conservatism kicks into action only if there are no defeaters and, even then, it supplies "some degree of justification." Given statements by Huemer, Dancy, and others, we take this to mean that the FVA, if successful, establishes a *prima facie* or *presumptive case* for realism in that it shifts

---

[4] As [name redacted for blind review] has pointed out to us, one should distinguish the claim that moral practice "feels" objective from the claim that we have a propositional seeming to the effect that morality is objective. For Huemer's principle to apply, the latter needs to hold. We assume that Dancy and others understand the situation, perhaps treating the objective "feeling" as grounding, or as evidence for, the propositional state.

the burden of proof onto the anti-realists' shoulders (e.g., Dancy 1986; Brink 1989, 36; 1988, 40; Pölzler 2018b, 43-4).

The FVA, then, can be stated as follows:

(P1) Morality seems to us as a mind-independent, objective domain.

(P2) *Phenomenal Conservatism*: If it seems to *S* that *p*, then, in the absence of defeaters, *S* thereby has at least some degree of justification for believing that *p*.

(P3) There are no defeaters to morality's seeming objectivity.

(C) Therefore, we have at least some degree of justification for believing that morality is objective.

For the sake of this discussion, we grant (P1) (though see the discussion in section 5.2.). We also grant, at least provisionally, the truth of phenomenal conservatism. However, we argue that (P3) is false: evolutionary explanations of the sort discussed in section 2 provide defeaters for morality's seeming objectivity.[5] More specifically, the evolutionary explanations serve as *undercutting* defeaters, i.e., they supply a reason to think that one's basis for believing is unreliable or somehow defective as a source of evidence (Pollock & Cruz 1999, 196). To clarify, suppose one learns that an object that seems red is in fact illuminated by a red light. Although this doesn't, by itself, provide evidence that the object isn't red, the new information functions as an undercutting defeater for the object's redness, indicating that its apparent redness doesn't supply a reason to think it is actually red.

We suggest that the evolutionary hypotheses presented in section 2 are akin to learning that what seems to be a red object is illuminated by a red strobe. They serve as an undercutting defeater of morality's objective phenomenology, and thereby of the FVA. For, given that we have such an evolutionary explanation, and given that this explanation is indifferent to whether there are mind-independent moral facts, we have no reason, *prima facie* or otherwise, to believe that moral realism is true on the basis of its seeming so. Recall that on Stanford's (2018) hypothesis, we experience

---

[5] An alternative formulation of the argument would involve abandoning premise (P3) and instead weakening the conclusion to:

(C*) Therefore, in the absence of defeaters, we have at least some degree of justification for believing that morality is an objective domain

Under this understanding, we would accept (C*), but argue that it doesn't permit one to infer (C)—since, as we argue, there are defeaters. We suspect that many philosophers who favored the FVA have implicitly assumed something like (P3), and thus take themselves to argue for (C) (and not merely for (C*)).

8

morality as objective because this enhanced the likelihood that our ancestors could cooperate in a flexible and open-ended manner, while minimizing the potential for exploitation. Clearly, such a fitness advantage accrued to ancestral humans whether morality is an objective realm of facts or not. Thus, if a Stanford-style explanation is correct, it shows that the relevant seeming is not a reason to form conclusions about moral reality, not even in a *prima facie*, presumptive sense.

As noted, the version of the FVA discussed here utilizes Huemer's phenomenal conservatism, which simply builds into the theory that defeaters can affect justification. One might wonder, however, if the FVA can be stated without appealing to such a theory. Perhaps other, stronger versions of epistemic conservatism, for example, can be used in order to secure the inference from (P1) to (C) even in the presence of defeaters. Or perhaps adopting a coherentist framework on which justification depends on whether one's current beliefs and experiences form a coherent package—while setting aside considerations about the causal origin of one's beliefs or experiences—can secure the FVA despite said evolutionary explanations.

We doubt that such roundabout ways to avoid our argument succeed. For one thing, versions of epistemic conservatism (or 'dogmatism') that cannot account for the relevance of defeaters for justification are implausible—indeed, they are implausible precisely *because* they don't allow for defeaters (cf. McCain 2018). And the same goes, we take it, for coherentist views that dismiss the relevance of information about the causal origin of one's belief and/or seeming for justification. In general, we think that any theory of justification that is committed to saying, for example, that even after one learns that an object that seemed red is illuminated by a red light, one is still justified—given one's current evidence—to continue believing that the object is red will face severe difficulties.

Second, perhaps some of these alternative epistemological frameworks can allow for some defeaters, but not others, to play a role in undercutting justification. Thus, perhaps such a framework can vindicate a distinction between the defeater which is provided by the information about the red light, and the defeater which is provided by the evolutionary explanations of the seemingness objectivity of morality, suggesting that only the former is a genuine defeater. (Perhaps, for example, only the former, but not the latter, introduces sufficient incoherence to render sticking one's guns unjustified.) We would need further details in order to assess this possibility. In both cases, we argue, the relevant information shows that the relevant seeming— that the object is red, or that morality is objective—is an off-track process; it's indifference to the

truth. The information about the red light suggests that the object would have seemed red whether or not it's in fact red; the information about the evolutionary explanations of morality's seeming objectivity suggests that morality would seem to us as objective whether or not it is, in fact, objective.

Finally, a person may insist that the difference between the cases has to do with the degree of justification that we have to believe in the relevant piece of (debunking) information. In the red light case, it is stipulated that the object was illuminated by a red light and that the subject is justified—even knows—that it was so illuminated. By contrast, our argument rests on a scientific hypothesis, which enjoys some evidential support but—as we acknowledged above—remains quite a ways from conclusive conformation.[6] We do not contest the idea that, all else equal, the stronger one's basis for believing a defeater the greater the degree to which it undermines the relevant belief. But we think that in this respect our argument is on par with, and perhaps (as we discuss in section 5) even on firmer grounds than, many other Evolutionary Debunking Arguments (EDAs) that have been discussed in recent metaethics. All such arguments can be strengthened by further exploring, and more thoroughly confirming, the underlying evolutionary explanations. We can only state that we think they are sufficiently plausible to merit interest in debunking arguments based on them, and hope that further work will supply more empirical evidence.

3.2. The explanationist argument

David Brink provides a succinct statement of the explanationist version of the argument from phenomenology:

> Realism, and realism alone, provides a natural explanation or justification of the way in which we do and can conduct ourselves in moral thought and inquiry. Of course, even so, moral realism could still be false; moral inquiry might be confused or misguided in some fundamental way. But if this claim about the realist nature of moral inquiry is right, we have reason to accept moral realism that can be overturned only if there are powerful objections to moral realism (1989, 24).

---

[6] As we noted in section 2, there are several rival evolutionary explanations of this phenomenon. Though we think that, among all of these explanations, Stanford's explanation if the most plausible one, *none* of these explanations assume the morality is experienced as objective *because* it's objective. So the truth of *any* of these theories would serve our proposes equally well.

The EA proceeds from much the same set of observations about how we experience morality as the FVA. Further, in most versions it too aims for a presumption in favor of realism (Brink 1989, 36.) It differs from the FVA in taking the form of an inference to the best explanation: given that we experience morality as objective, and given that realism is the best explanation for this, we should plump for moral realism.

It is worth clarifying that this argument is distinct from another kind of explanation-based defense of realism, according to which objective moral facts explain our moral judgements (Sturgeon 2006). Such an argument operates with a different explanandum—it asserts that moral facts explain *the first-order moral judgments* we make. It too is subject to significant, albeit rather different, concerns (Sinclair 2011). We set it aside here.

Now, the EA can be critiqued in several ways. One might wonder whether realism can explain moral phenomenology and how. Or one may suspect that what is at stake is *justifying* our moral practice, rather than explaining it (Brink speaks of an "explanation *or justification*"). But for present purposes we grant that realism provides *an* explanation of morality's objective phenomenology. Our claim is that evolutionary accounts provide a *much better* explanation of the same phenomena, in several respects. First, those accounts rely on a well-understood and well-confirmed mechanism, namely evolution by natural selection. We know how this process works and we have excellent evidence that it explains a range of biological phenomena. Second, explaining moral phenomenology by means of evolution unifies it with a wide range of other phenomena, both within the domain of human cognition and well beyond it. Third, the account is ontologically more economical: it does not posit facts beyond those that are already part of any reasonable explanation of human evolution—social structure, the fitness advantages of cooperation etc. In all these respects the evolutionary account outperforms the explanation from moral realism.[7] And we can see no significant respect in which the converse holds, i.e., no way in which the moral realist account is superior to the evolutionary one.

Thus, we conclude that advocates of the explainationist argument are simply wrong to suppose that "[r]ealism, and realism alone, provides a natural explanation […] of the way in which we do

---

[7] A caveat: if one holds a naturalistic moral realism, then one's explanation of the objective phenomenology of morality may be as economic as the evolutionary explanation. Assessing this point would require a detailed statement of the explanation on the part of the (naturalist) realist, which hasn't been given as far as we know.

and can conduct ourselves in moral thought and inquiry" (Brink 1989, 24). Not only is realism not alone in this explanatory arena, an evolutionary explanation appears superior.

4. Enoch's argument from impartiality

We turn to an argument that involves a more subtle reliance on morality's objective phenomenology—David Enoch's "argument from the moral implications of objectivity" (2011, Ch. 2). In a nutshell, it says that realism sits better—relative to metaethical subjectivism —with a first-order principle regarding moral disagreement. This is because such disagreements are, in relevant respects, more like disputes over questions of objective fact than like differences of taste or preference. We contend that this argument, while not explicitly invoking moral phenomenology, in fact relies on it to buttress a key premise.

Enoch's overall argument has two parts. First, he offers a *reductio ad absurdum* of Caricaturized Subjectivism (CS), a simplistic version of metaethical subjectivism. He then generalizes the argument, putting forward an explanatory challenge for more sophisticated subjectivist views. We believe that both parts are vulnerable to a debunking move, but we'll focus on the first. For if the first part falls then so does the second, as subjectivists face no explanatory challenge to begin with.

Suppose *A* and *B* are planning to spend the afternoon together. *A* prefers to watch a movie, *B* would rather play tennis, and they don't have time for both. Enoch suggests that *A* and *B* ought to, or at least have significant reason to, resolve the dispute impartially, i.e., neither of them should give precedence to his or her (mere) preference. In contrast, imagine *A* and *B* are trying to neutralize a bomb. *A* thinks the blue wire should be cut, while *B* is convinced that it's the red wire. Assume that *A* is right and that she's rational in believing as she does. Here, Enoch says, it seems that *A* "should act on what [*A*] (rightly, and rationally) take[s] to be the truth of the matter" (*Ibid*, 22). In other words, this factual dispute should not be resolved impartially.

Enoch's thinks that moral disagreements are more like the bomb case than the how-to-spend-the-afternoon case; they should not be resolved impartially. Giving an example of two persons who disagree over the permissibility of causing serious pain to a dog, he suggests that in that kind of case one is permitted, and maybe even morally required, to stand one's ground. In a moral disagreement like that "[g]oing for an impartial solution will be—unless it can be justified by other factors—morally wrong" (*Ibid*, 24-5).

These observations lead to a *reductio* against CS, the view that "moral judgments report simple preferences, ones that are on a par with a preference for playing tennis or for catching a movie" (*Ibid*, 25). Reformulated for brevity, it runs as follows:

(1) If CS is true, then interpersonal conflicts due to moral disagreements are due to differences in mere preferences.

(2) IMPARTIALITY: When an interpersonal conflict (of the relevant kind) is a matter of mere preferences, then an impartial solution is called for.

(3) In cases of interpersonal conflict (of the relevant kind) due to moral disagreement, often an impartial solution is *not* called for.

(4) Therefore, CS is false.

As noted, Enoch's argument doesn't rely *explicitly* on the objective phenomenology of morality. Premise (3) is a first-order moral premise, concerning proper conduct in situations of moral disagreement. How, then, can it be debunked by explanations of the sort discussed in section 2? The answer has to do with Enoch's *motivation* for distinguishing moral conflicts from disagreements over mere preferences: he all but admits that it stems from the fact that moral disagreement *seems* much more like disagreement about factual matters: "it seems to me that (when other things are equal) the right way to proceed in cases of interpersonal conflicts due to moral disagreement is analogous to the right way to proceed in cases of interpersonal conflicts due to factual disagreement" (*Ibid,* 24). Indeed, in the introduction to the book in which he presents this argument—and as a way of explaining what he's trying to achieve with arguments of this sort—Enoch says: "I suspect that as a psychological matter, I hold the metaethical and metanormative views I hold not because of highly abstract arguments in [philosophy, but because] like many other realists (I suspect), I pre-theoretically feel that nothing short of a fairly strong metaethical realism will vindicate our taking morality seriously" (*Ibid*, 8).[8]

Thus, we suspect—in a truly debunking spirit—that what is driving the intuition behind premise (3) is, in fact, the psychological tendency to experience morality as objective, the exact tendency explained by the evolutionary explanations presented earlier. This, if true, should lead

---

[8] In his 2017 Enoch defends a version of the argument from moral phenomenology. There he makes clear that he accepts the description of moral phenomenology on which it rests.

one to doubt (3), and with it Enoch's entire argument. Note that, as we discussed in section 2, the thought that moral disputes aren't faultless disagreements—that, like factual disagreements, at least one of the disagreeing parties must be mistaken—is taken as one of the earmarks of objectivity in the psychological literature. Indeed, this is the very same phenomenon that explanations such as Stanford's are aiming to explain. As is apparent from his statements, it's clear that Enoch takes the resemblance—here, perhaps, *normative* resemblance—between factual and moral disagreement as supporting his claim that an impartial solution is *not* called for in cases of moral disagreements. It is this reliance that makes him vulnerable to our debunking argument.

Now, one might detect an *ad hominem* scent in the claim we've just made. After all, Enoch does not explicitly justify premise (3) by appealing to phenomenology. And even if his, or anyone else's, acceptance of (3) can be explained in the psychological manner we have described, that doesn't mean that the explanation is part of the argument or that the argument is beholden to it. However, we do think that once evolutionary accounts like those surveyed above are brought to the fore, the need to justify (3) becomes pressing. And it is unclear to us that one can justify such a principle without presupposing that moral disputes *appear* to be concerned with objective, mind-independent matters of fact. The thought that IMPARTIALITY is inapplicable in cases of moral disagreement holds whatever appeal it does because of an evolutionarily-based psychological tendency to experience morality as objective, should worry Enoch and those who find his argument compelling.

To further clarify the point, note that we are not claiming that Enoch's argument begs the question against the subjectivist. Premise (3) is a first-order moral premise, and as such, it allows for different metaethical interpretations—including subjectivist ones. Rather, our claim is that what underlies the intuition behind premise (3) is precisely the tendency to experience morality as objective, thus making it vulnerable to our argument. But the existence of such a psychological tendency is something that even a subjectivist can and usually does accept while denying its evidential force or while trying to accommodate it in some other ways. In this sense, our argument can serve the subjectivist in responding to Enoch's argument, by claiming that the intuitive appeal of premise (3)—an intuitive appeal that even the subjectivist can acknowledge—relies on said psychological tendency, and so that its evidential force is debunked.

5. *Yet another debunking argument?*

Evolutionary debunking arguments (EDAs) have generated extensive discussion in recent metaethics. One might therefore reasonably wonder whether and how the arguments we have presented are novel. This is the topic of the present section. While our argument clearly shares the general structure of existing EDAs, we believe it differs in important ways. To this end, we will first discuss the two best-known EDAs, advocated by Street and Joyce, showing how these arguments target different phenomena, relative to our discussion. Then, we look at the empirical bases of Street's and Joyce's claims, highlighting further differences from our argument, and (cautiously) arguing that the empirical assumptions we rely on are more plausible. Lastly, we suggest that our argument is well-equipped to cope with key objections to standard EDAs.

5.1. ***Comparison to Street and Joyce.*** The current debate about EDAs was instigated, in large part, by two texts published in 2006: Sharon Street's paper "A Darwinian dilemma for realist theories of value" and Richard Joyce's book "The Evolution of Morality" (especially chapters 4 and 6.) Street's and Joyce's arguments differ somewhat; we begin with the former.

Street's overall aim is to argue against realism about moral value and for a form of constructivist antirealism (understood broadly as a view on which "evaluative facts or truths [are] a function of our evaluative attitudes." 2006, 152). Our focus is the first, anti-realist part of the argument. Street's strategy is to argue that if evolution explains the content of our evaluative beliefs, then it is unlikely that such beliefs reflect the evaluative truths. Hence the realist faces a dilemma—either our beliefs do not reflect the moral facts, or she must reject the relevant evolutionary explanations, which Street takes to be scientific explanations in good standing. The key point for our purposes is that Street's argument is directed at the (justification for the) contents of our *first-order moral beliefs and attitudes*. Her argument does not concern moral phenomenology, nor does it target the face-value argument or other arguments relying on morality's objective phenomenology. Moreover, because her focus is first-order normative attitudes, Street's argument rests on empirical assumptions concerning such attitudes. As she states early on: "[T]he opening premise of the Darwinian Dilemma argument is this: the forces of natural selection have had a tremendous influence on the content of human evaluative judgements." (*Ibid*, 113).

Street's argument differs from ours in several important respects. Most directly, it alleges that evolutionary considerations impact the epistemic standing of first-order moral attitudes and affects realism somewhat indirectly—via the first horn of the Darwinian Dilemma. Moreover, as we will explain in the following two subsections, the first-order character of the argument means that it relies on a different empirical basis, and it affects its vulnerability to objections. Before getting to that, however, let us discuss Joyce's version of evolutionary debunking.

Joyce's discussion is couched in slightly different terms and is somewhat harder to interpret. In his 2006 Joyce argues that his account of the evolution of "the moral sense" leads to a form of (first-order) moral skepticism.[9] The argument has two main parts: first, Joyce argues that an evolutionary explanation of the moral sense operates as a kind of defeater, in that learning of the explanation is much like learning that one's beliefs about Napoleon are the product of a "Napoleon pill" (2006, 181). But Joyce holds that this argument only works against non-naturalist versions of realism.[10] He therefore spends considerable time arguing against naturalist realism. Both parts of this fairly complex argument are relevant to our discussion, albeit in different ways. Let us explain.

The first part has, like Street's argument, a first-order character: it aims to undermine the justification for our moral beliefs, on the assumption that they can be explained by natural selection. This is the point of the "Napoleon pill" analogy. As Joyce clarifies, the argument is that "once we become aware of this [evolutionary] genealogy of morals we should (epistemically) … cultivate agnosticism regarding all positive beliefs involving these [moral] concepts…" (*Ibid*, 181). To be clear, Joyce's *evolutionary* assumptions differ slightly from Street's. He holds that evolution shaped the contents of our moral beliefs, by *leading to our possession of moral concepts*. Nonetheless, the evolutionary debunking concerns first-order moral beliefs, not our meta-ethical attitudes.

---

[9] In his 2006 (Ch. 7) Joyce confusingly refers to this skeptical conclusion as an "error theory". As he clarifies in later work, he does not think that an EDA can establish anything beyond a form of first-order moral skepticism which is, in principle, compatible with realism (2016, 144n3). As noted above, Joyce's evolutionary hypothesis can serve our argument as well. But Joyce himself sees it as casting doubt on first-order moral beliefs, and not as undermining arguments for metaethical views (2001, Ch. 6; 2006, Ch. 6). This marks an important difference between Joyce's position and ours.

[10] Here Joyce differs from Street, who holds that the Darwinian Dilemma applies to naturalist realisms too—see especially section 7 of her 2006 article.

Now, as noted, Joyce's argument also aims to block a naturalist reduction of morality. This is where his second evolutionary hypothesis, which we discussed in section 2 above, enters. It concerns the evolutionary origins of what he terms the "inescapable practical authority" of morality (*Ibid*, 190). In essence, Joyce thinks the naturalist cannot provide a satisfying account of this aspect of our moral thinking. We will not assess this argument here; we only note that Joyce does not, certainly not in his 2006 book, connect this evolutionary hypothesis, and phenomenology-based arguments for realism that we discussed above. Moreover, Joyce's argument against naturalism doesn't in any significant sense rely on the aforementioned evolutionary hypothesis. For he argues that a naturalist reduction of morality cannot account for the phenomenology itself, irrespective of its origin. In other words, Joyce aims to debunk first-order beliefs, by means of a claim about the impact of evolution on our moral concepts. His appeal to moral phenomenology is part of separate, non-debunking argument.

A final comment before we move on. In later work, Joyce has given various (condensed) reformulations of his original argument, mostly in response to criticism. Some of these reformulations make it sound akin to the arguments we have given above.[11] But nowhere, so far as we know, has he made a clear and explicit second-order argument of the sort we have suggested. And, certainly, he nowhere defends this sort of argument in detail or shows how it connects to existing arguments for realism, as we have done. That said, to the extent that our argument is reminiscent of these later reformulations by Joyce, we are happy to acknowledge our

---

[11] Perhaps most relevantly, consider the following excerpt from Joyce's Stanford Encyclopedia entry on moral anti-realism (the only point where evolutionary debunking is mentioned). Considering the counter-intuitiveness of moral anti-realism he comments:

"One noteworthy type of strategy here is the "debunking argument," which seeks to undermine moral intuitions by showing that they are the product of processes that we have no grounds for thinking are reliable indicators of truth. (See Street 2006; O'Neill 2015; Joyce 2013, 2016.) To the extent that the anti-realist can provide a plausible explanation for why humans would tend to think of morality as objective, even if it is not objective, then any counter-intuitiveness in the anti-realist's failure to accommodate objectivity can no longer be raised as an ongoing consideration against moral anti-realism."

Note that Joyce refers here to Street's argument (and to O'Neill, another paper addressing "causes of beliefs"), and that he regards the role of evolutionary debunking as explaining away the counter-intuitiveness of anti-realist positions. So it is unclear whether he has a first- or second-order form of debunking in mind. Still, as we say in the main text, we readily acknowledge that such formulations bear a kinship with our argument.

indebtedness to them and to see ourselves as fully articulating and defending a line of reasoning present in his work.

5.2. Different empirical bases. As noted, the argument we offer differs from previous forms of evolutionary debunking not only in its target, but also in its empirical underpinnings. To be clear, we do not wish to suggest that the evolutionary explanations we rely on are on firm ground, as opposed to the explanations that Street and Joyce appeal to. Most theoretical work in this area is speculative. Nonetheless, the fact that the empirical basis of our argument is distinct means that (empirical) critiques of existing EDAs are largely irrelevant to our argument. Furthermore, we think that, at least to some extent, our argument is better-supported. This is so in several respects.

First, to the extent that an hypothesis concerns the idea that evolutionary forces have shaped the content of our first-order moral attitudes, an important source of evidence would be the distribution of normative attitudes among human groups and individuals. In the extreme, one would expect to find certain normative universals, although that is not especially likely. More plausibly, if moral attitudes are evolved adaptations, then one would expect to find significant thematic clustering – moral norms would, across different cultures and periods, be *about* the same issues (Sripada 2008). A related possibility is that moral norms exhibit shared structural features, perhaps by analogy with language (Mikhail 2011; Hauser, 2006). In contrast, hypotheses concerning the evolution of moral phenomenology imply far less universality of content, across cultures and historical contexts. As Stanford (2018) notes, a significant part of the motivation for his view is the finding that moral systems are, in terms of content, very diverse. It is hard to locate true moral universals, and those that enjoy some plausibility can be explained by non-evolutionary means (Machery and Mallon 2010; Levy and Levy 2020). There are also fairly convincing critiques of the linguistic analogy (Prinz 2008; Sterelny 2010). In this regard, we think hypotheses pertaining to second-order moral phenomenology, and by extension our reliance on them, have better standing.

Second, the most relevant evidence for the kind of evolutionary explanations our argument assumes concerns the ways and extents to which humans moralize and objectify norms: to which norms do they attach special authority? Do they experience relevant norms as externally imposed, or as trumping other norms? Etc. Such evidence is highlighted by Stanford, and reviewed more extensively by Pölzler (2018b). This body of work is fairly recent and it is hard, at present, to glean

a clear picture from it. Some studies suggest that laypeople tend toward moral realism but others don't. As Pölzler (*Ibid*, Chapters 3 and 7; 2018a) argues, this is partly due to divergent findings, but also because of uncertainty as regards the validity of methods and results.

We suggest that work in this area can fortify our argument whether or not it ends up vindicating the idea that people experience morality as a domain of objective facts. For if it turns out that humans "objectify" moral norms, then that lends support to evolutionary hypotheses of the sort our argument presupposes. But if the contrary holds, then this tells against arguments for realism that we aim to undermine. To clarify, suppose, as a fairly extreme possibility, that we learn that a very small proportion of humans experience morality as objective—perhaps primarily metaethicists. In that case, the evolutionary explanations we've discussed would lose much of their plausibility; they would be directed at a non-existent explanandum. But, by the same token, this would affect the plausibility of *arguments for realism* that are grounded in morality's objective phenomenology. [12] To be sure, an advocate of such arguments is likely to be a person for which morality seems objective and therefore she may hold that they work *for her*. This seems to us an unattractive move, but we will not enter into the complex epistemological considerations necessary to evaluate it. We only note that to the extent that moral psychology shows that objective phenomenology isn't prevalent this is likely to cut both ways—against debunking explanations, but also, independently, against realism itself. This is not the case, it should be clear, for the EDAs offered by Street and Joyce.

A related empirical matter is whether the moral forms a psychologically distinct category—including whether moral claims are taken to have different correctness conditions relative to other norms and whether moral thinking has a distinctive developmental signature. Here, an important line of thought stems from the work of Erik Turiel (1983), who argued that there exists a fairly sharp division, emerging in relatively young children across a range of cultures, between moral and conventional norms. Such work has been subjected to criticism, and there are doubts about its empirical underpinnings.[13] But these doubts probably affect both our argument and previous EDAs. For the idea that the moral forms a distinctive psychological category is likely implicit in Street and (very likely) in Joyce as well (Machery and Mallon 2010; Levy and Levy 2020).

---

[12] Pölzler & Wright (2020) suggests a view fairly similar to this latter option.

[13] Several of the commentators on Stanford's original BBS paper raise such doubts. See especially Davis & Kelly; Patel and Machery; Stich; Theriault & Young. And see Stanford's response, especially R2. See also Stich (2018)

Finally, a substantial amount of work on the evolution of morality has looked to findings from animal behavior, especially to shore up claims about the importance of reciprocity (de Waal 1996; 2006). These findings, if relevant to humans, pertain primarily to questions concerning *the content* of moral attitudes. And while fascinating, such work is also, quite obviously, highly susceptible to interpretive misgivings. We therefore tend to doubt its ability to play a significant evidential role. Against this background is perhaps an advantage for the evolutionary theorizing that we have been appealing to that it can be tested in more direct and telling ways in humans—where evidence for moral phenomenology can be more reliably detected.

To summarize: none of the extant hypotheses concerning the evolution of moral cognition can claim to have an airtight evidential basis. But we cautiously suggest that the explanations we appealed to are better supported, and that the most telling potential evidence against them— concerning whether and how humans objectify norms—is likely to support our argument (or at least its conclusion), however it turns out.

5.3 Responding to (existing) objections. Having highlighted how our argument differs, both in its target and its empirical underpinnings, from existing EDAs, we now want to discuss its vulnerability to objections. In particular, we suggest that it is by-and-large immune to the most significant issues that have been raised against first-order EDAs.[14]

One such objection is that while evolution by natural selection may explain some of our moral beliefs there are beliefs that it is very hard to find an evolutionary "rationale" for (Copp 2008; FitzPatrick 2014; Parfit 2011, 534-8; Shafer-Landau 2012). Shafer-Landau, for instance, questions whether evolution explains such beliefs as "those that counsel impartial benevolence, compassion for vulnerable strangers, kindness to small animals, concern for distant peoples and future generations, and speaking truth to power" (2012, 7). Whether Shafer-Landau is correct or not, it should be clear that our argument is not open to such a criticism in the first place. For it does not rely on an explanation of first-order beliefs, or of the mechanisms generating them. Instead, we assume that evolution explains how moral thought and talk "feels." To be sure, the explanation we

---

[14] We omit the so-called overgeneralization objection, which alleges that skepticism with respect to morality may run rampant, leading to implausible skeptical conclusions about our knowledge of the external world (Vavova 2014, 82-3, Shafer-Landau 2012, 22) or of other *a priori* domains (Bedke 2009, section 3; Enoch 2011, 175-6). This objection doesn't seem to us very compelling to begin with, and even if it were, it doesn't appear to be adaptable to our argument.

consider makes some assumptions about the contents of moral beliefs—typically, that they pertain to social cooperation—but these amount to fairly weak constraints. Moreover, such an explanation is consistent with the existence of moral beliefs with other contents, including those Shafer-Landau mentions. It is consistent with an expansion, as it were, of the moral domain, so long as such an expansion does not undermine the adaptive value of social cooperation. Thus, even if some significant portion of our first-order moral beliefs aren't explained by natural selection, this does not threaten our argument.

Another objection to standard EDAs rests on the possibility that we can justify the debunked beliefs independently of the influence of natural selection. Such a line of thinking can be developed in several ways. Shafer-Landau (2012), for instance, suggests that moral beliefs may be the product of a faculty for forming *a priori* beliefs, a faculty which can be shown to be reliable irrespective of claims about the evolution of moral cognition. Similarly, David Enoch has suggested to us (personal communication) that if one can make the case that we have a general ability to distinguish appearances from reality—perhaps a kind of meta-sensory capacity—then we can rely on it to distinguish reality from mere appearances in the moral domain, too.

As can be seen, this response works by assuming that general epistemic capacities get locally applied in the case of moral or metaethical knowledge. In this sense the response does not distinguish 'standard' EDAs from our meta-ethical debunking argument. Correspondingly, it suffers from the same problem in both cases: the fact that we have a capacity to attain knowledge of some general sort does not mean that we cannot suffer from biases or deficiencies in gaining such knowledge in some particular context. For instance, we have a general capacity to attain mathematical knowledge, but we also have well-known biases in thinking about probability. Similarly, we have a generally reliable sense organs, but we are vulnerable to various perceptual illusions. If evolutionary explanations of our moral cognition are correct—either those invoked by 'regular' debunkers, or those we rely on—then this may be our situation as regards morality, too. Shafer-Landau might be right that we have a capacity for *a priori* knowledge and yet wrong that it is reliably applied in the moral domain; Enoch may be right that we have a capacity to tell apart reality from mere appearance, and yet wrong to suggest that it is reliable vis-à-vis moral metaphysics.

Moreover (and focusing on Enoch's suggestion), we seem to have good reasons to think that our general capacity to tell apart reality from mere appearance doesn't extend to the moral (or

indeed, the metaethical) domain. Such a general capacity clearly has an evolutionary advantage. But when it comes to the metaethical domain, it's not clear whether and why such a capacity would be fitness enhancing. Indeed, perhaps *lacking* such a capacity in the metaethical domain would carry an evolutionary advantage, if there is no objective moral reality. Again, the evolutionary explanation we rely on suggests that experiencing morality as objective would be advantageous not because morality is indeed objective, but due to other fitness-increasing effects, ones that would occur whether or not objective moral reality exists. So to posit that in addition to the evolutionary mechanisms responsible to our objective moral phenomenology, evolution has generated another, distinctive capacity to tell apart *moral* reality from illusion, seems somewhat mysterious and explanatorily superfluous. If the explanations we rely on are correct, we can see no evolutionary rationale for extending our general capacity to tell apart reality from illusion to the metaethical domain.

The final reply we will tackle is the so-called 'third factor' response. Here, the key idea is that the realist may grant that our moral beliefs were *selected for* because of their effects on social cooperation, or more generally in a manner that is indifferent to whether there are mind-independent moral truths. And yet she may suggest that these beliefs are non-causally correlated with moral truths, so that there was *selection of* a truth-tracking moral sensibility. In general, such responses start by assuming the truth of some moral principle, and then show that while selection may have acted on the corresponding beliefs irrespective of their truth, this is compatible with the beliefs being true.

Wielenberg, for instance, suggests that it is plausible to think that humans evolved "to view themselves as surrounded by a kind of moral barrier that it is wrong, unjust, evil, or somehow morally inappropriate or illegitimate for others to cross" (2010, 444-5). He goes on to argue that *if* we have such rights, then there is a plausible explanation of how we could have knowledge of these rights: the capacity to view oneself as "surrounded by a kind of moral barrier" is essentially identical to the capacity for moral knowledge of the relevant sort; if I view myself as possessing rights, I have rights. Here, the relevant 'third-factor'—what generates a correlation between our beliefs about rights and our possessing of rights—are certain cognitive faculties that are "responsible for the presence of moral rights in that the presence of the relevant faculties entails the presence of rights" (*Ibid.*, 450). Structurally similar stories are given by Brosnan (2011), Enoch (2011, Ch. 7) and Skarsaune (2011).

Generally speaking, then, third-factor responses work by assuming the truth of some first-order moral principle. Wielenberg's account assumes that persons have rights, whereas Skarsuane's account, for example, assume that pleasure is usually (morally) good and pain is usually (morally) bad. As some have suggested, this carries the risk of begging the question (e.g., Shafer-Landau 2012, 33-4; Vavova 2015, 81). But note that advocates of third-factor responses to 'standard' EDAs are (potentially) begging the question by assuming a first-order moral principle. In many cases, such a principle will be acceptable to non-realists and at any rate there is some "distance" between the assumed principle and the sought-after (realist) conclusion. In our case, however, what is being explained via selection is *the self-same phenomenon* on which realist arguments rely—our tendency to view morality as objective. This makes the threat of circularity more acute, as we now explain.

A third-factor response to our argument would start, presumably, by assuming that there exist mind-independent moral truths. If there are such truths, the response continues, then the fact that its effects, namely efficient social coordination and avoidance of exploitation, are fitness-enhancing explains the correlation between our beliefs about the existence of such truths and their genuine existence.

Now, perhaps it is okay to assume that there are mind-independent moral truths in response to 'standard' EDAs. After all, standard debunkers themselves usually assume that there are such truths, arguing that *even if this is the case*, evolutionary explanations undermine knowledge of (or justified beliefs about) their content (Wielenberg 2010, 447; Enoch 2011, 156; cf. Copp 2019, 237). But a parallel third-factor move would seem even weaker in relation to our argument. That's because we aim to debunk *arguments for the existence of mind-independent moral truths*. So to assume that there are such truths in the present context amounts to begging the question *simpliciter*. To avoid this, proponents of the FVA, for instance, would have to settle for the following conditional conclusion: *if* there exist mind-independent moral truths, then we have a *prima facie* reason to believe that they exist. Similarly, to avoid begging the question, Enoch's conclusion should say: *if* there are mind-independent moral truths, then since we are justified in believing that moral conflicts shouldn't be resolved impartially, CS is false. We think this weakening of the conclusion makes third-factor response far less significant in relation to our argument.

Overall, then, it appears that the main objections to familiar EDAs, even if they work in their original context, have considerably less force against the argument we have made.

6. Conclusion: What's at stake?

The arguments in sections 3 and 4 are simple and straightforward, we think. But if successful, they undermine realist appeals to moral phenomenology as well as Enoch's argument from IMPARTIALITY. This, we think, makes a real difference for the dialectic over moral realism. In particular, if the argument from moral phenomenology is defeated, then realism ceases to be the default position. If Enoch's argument is defeated, the upshot is that subjectivist views, even in a caricaturized version, must be taken more seriously. Let us expand on each of these points in turn.

The argument from moral phenomenology has played a central dialectical role in the debate over moral realism. Many metaethicists, including prominent advocates of antirealism (Blackburn 2006, 153; Mackie 1977, 35) endorse the argument's conclusion, agreeing that antirealists bear the burden of proof. This explains why the debate consists almost exclusively in objections to realism – like concerns about how to reconcile it with naturalism, or over the possibility of moral knowledge – and attempts to rebut them. As Dancy puts it, the FVA is "perhaps the only argument for realism, remaining thoughts being used for defense/offence" (1986, 175). Or as Nagel acknowledges: "[I]t is very difficult to argue for such a possibility [as moral realism] except by refuting certain arguments against it" (1986, 143).

If correct, then, rejecting the argument from moral phenomenology alters the dialectical situation. Realism ceases to be 'the view to beat,' a view that we should accept absent compelling objections. Rather, it is 'a view to defend;' a view we should accept (only) under the pressure of argument, as it were, much like any other metaethical position. Moreover, its defense cannot rely on the objective phenomenology of morality, since if the evolutionary explanations we have surveyed are correct, it cannot serve as evidence for realism.

What about the upshots of rejecting Enoch's argument? Recall that we have argued that the evolutionary hypotheses presented in section 2, if correct, undermine not only Enoch's case against subjectivist views in general, but even against CS. Now, as Enoch notes (2011, 27) no contemporary metaethicist is a Caricaturized Subjectivist. As the label is intended to convey, CS is a caricature. Indeed, CS is exactly the kind of view that is usually ruled out early on in introductory texts and courses in moral philosophy (e.g., van Roojen 2015, 106-11). Thus, if our argument succeeds then even CS, a caricature in the sense that it's a view so simple as to be a non-starter, cannot be ruled out by appeal to our objective moral phenomenology.

Overall, we believe that the argument we've presented, while simple and straightforward, represents more than a local intervention in the literature on realism. If sound, it should seriously worry contemporary realists inasmuch as it alters the structure of debates in this area. Realism cannot be taken as the default view, and more effort should be put into developing positive arguments for it. Moreover, if the appeal to phenomenology is blocked, then even simple forms of subjectivism may pose a real threat to realism.

To be sure, our argument's significance depends in part on the cogency of the underlying evolutionary explanations. But if they are cogent and if our conclusions are accepted, this may spur realists to formulate additional arguments for their views. If they succeed, this would result in strengthening the case for realism, inasmuch as it will rest on positive considerations, rather than on mere burden-shifting. However, unless and until that is done, the playing field would appear to be levelled, with neither realism nor its denial serving as the default, to-be-refuted, position.

References

Bedke, Matthew S. (2009). Intuitive non-naturalism meets cosmic coincidence. *Pacific Philosophical Quarterly*, 90(2): 188-209.

Björnsson, Gunnar. (2012). Do 'objectivist' features of moral discourse and thinking support moral objectivism? *The Journal of Ethics,* 16(4): 367-393.

Blackburn, Simon. (2006). Antirealist expressivism and quasi-realism. In *The Oxford Handbook of Ethical Theory*, ed. David Copp. Oxford University Press.

Brink, David O. (1989). *Moral Realism and The Foundations of Ethics*. Cambridge University Press.

Brosnan, Kevin. (2011). Do the evolutionary origins of our moral beliefs undermine moral knowledge? *Biology & Philosophy*, 26(1): 51-64.

Clarke-Doane, Justin. (2016). Debunking and dispensability. In *Explanation in Ethics and Mathematics: Debunking and Dispensability*, eds. Uri D. Leibowitz and Neil Sinclair. Oxford University Press.

Copp, David. (2008). Darwinian skepticism about moral realism. *Philosophical Issues,* 18:186-206.

Copp, David. (2019). How to avoid begging the question against evolutionary debunking arguments. *Ratio,* 32(4): 231-245.

Dancy, Jonathan. (1986). Two conceptions of moral realism. *Proceedings of the Aristotelian Society, Supplementary Volumes,* 60: 167-205.

Dennett, Daniel. (1995). *Darwin's Dangerous Idea*. Simon and Schuster.

DeScioli, Peter, and Kurzban, Robert. (2013). A solution to the mysteries of morality. *Psychological Bulletin* 139(2): 477–496.

Enoch, David. (2011). *Taking Morality Seriously: A Defense of Robust Realism*. Oxford University Press.

Enoch, David. (2017). Non-naturalistic realism in metaethics. In *The Routledge Handbook of Metaethics*, eds. Tristram McPherson and David Plunkett. Routledge.

Fitzpatrick, William J. (2014). Why there is no darwinian dilemma for ethical realism. In *Challenges to Moral and Religious Belief*, eds. Michael Bergmann and Patrick Kain. Oxford University Press.

Goodwin, Geoffrey P., and Darley, John M. (2008). The psychology of meta-ethics: Exploring objectivism. *Cognition* 106:1339–66.

Goodwin, Geoffrey P., and Darley, John M. (2012). Why are some moral beliefs perceived to be more objective than others? *Journal of Experimental Social Psychology* 48(1):250–56.

Handfield, Toby. (2016). Genealogical explanations of chance and morals. In *Explanation in Ethics and Mathematics: Debunking and Dispensability*, eds. Uri D. Leibowitz and Neil Sinclair. Oxford University Press.

Hauser, Marc D. (2006). The liver and the moral organ. *Social Cognitive and Affective Neuroscience*, 1(3): 214-220.

Hopster, Jeroen. (2019). The meta-ethical significance of experiments about folk moral objectivism. *Philosophical Psychology*, 32(6): 831-852.

Huemer, Michael. (2007). Compassionate phenomenal conservatism. *Philosophy and Phenomenological Research* 74(1): 30–55.

Joyce, Richard. (2001). *The Myth of Morality*. Cambridge University Press

Joyce, Richard. (2006). *The Evolution of Morality*. MIT press.

Joyce, Richard. (2016). Evolution, truth-tracking, and moral skepticism. In his *Essays in Moral Skepticism*. Oxford University Press.

Kahane, Guy. (2011). Evolutionary debunking arguments. *Noûs*, 45(1): 103-125.

Levy, Arnon, and Levy, Yair. (2020). Evolutionary debunking arguments meet evolutionary science. *Philosophy and Phenomenological Research*, 100(3): 491-509.

Loeb, Don. (2007). The argument from moral experience. *Ethical Theory and Moral Practice,* 10(5): 469-484.

Machery, Edouard, and Mallon, Ron. (2010). Evolution of morality. In *The Moral Psychology Handbook*, eds. John M Doris and the Moral Psychology Research Group. Oxford University Press.

Mackie, John L. (1977). *Ethics: Inventing Right and Wrong*. Penguin.

McCain, Kevin. (2018). Explanationist aid for phenomenal conservatism. *Synthese*, 195: 3035-3050.

Mikhail, John. (2011). *Elements of Moral Cognition: Rawls' Linguistic Analogy and the Cognitive Science of Moral and Legal Judgment.* Cambridge University Press.

Nagel, Thomas. (1986). *The View from Nowhere*. Oxford University Press.

O'Neill, Elizabeth. (2015). Which causes of moral beliefs matter?. *Philosophy of Science*, 82(5): 1070-1080.

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

Pollock, John, and Cruz, Joseph. (1999). *Contemporary Theories of Knowledge*. Rowman and Littlefield.

Pölzler, Thomas. (2018a). How to measure moral realism. *Review Of Philosophy and Psychology,* 9: 647–670.

Pölzler, Thomas. (2018b). *Moral Reality and The Empirical Sciences*. Routledge.

Pölzler, Thomas, and Wright, Jennifer Cole. (2019). Empirical research on folk moral objectivism, *Philosophy Compass.* 14(5).

Pölzler, Thomas, and Wright, Jennifer Cole. (2020). An Empirical Argument against Moral Non-Cognitivism. *Inquiry: An Interdisciplinary Journal of Philosophy*.

Prinz, Jesse. (2008). Resisting the linguistic analogy: A commentary on Hauser, Young, and Cushman. In *Moral psychology, Vol. 2: The Cognitive Science of Morality: Intuition and Diversity,* ed. Walter Sinnott-Armstrong. MIT Press.

Ruse, Michael. (1986). *Taking Darwin Seriously*. Basil-Blackwell.

Ruse, Michael. (2006). Is darwinian metaethics possible (and if it is, is it well taken)?. In *Evolutionary Ethics and Contemporary Biology,* eds. Giovanni Boniolo and Gabriele De Anna. Cambridge University Press.

Ruse, Michael. (2009). Evolution and ethics: The sociobiological approach. In his *Philosophy After Darwin*. Princeton University Press.

Sayre-McCord, Geoffrey. (2015). "Moral Realism." In *The Stanford Encyclopedia of Philosophy*, edited by Edward N Zalta. http://plato.stanford.edu/entries/moral-realism/

Shafer-Landau, Russ. (2012). Evolutionary debunking, moral realism and moral knowledge. *Journal of Ethics and Social Philosophy,* 7: 1–37.

Sinclair, Neil. (2011). The explanationist argument for moral realism. *Canadian Journal of Philosophy*, 41(1): 1-24.

Skarsaune, Knut. O. (2011). Darwin and moral realism: survival of the iffiest. *Philosophical Studies*, 152(2): 229-243.

Skyrms, Brian. (2004). *The Stag Hunt and the Evolution of Social Structure*. Cambridge University Press.

Sripada, Chandra S. (2008). Nativism and moral psychology: Three models of the innate structure that shapes the contents of moral norms. In *Moral psychology, Vol. 1: The Evolution of Morality: Adaptations and Innateness,* ed. Walter Sinnott-Armstrong. MIT Press.

Stanford, Kyle P. (2018). The difference between ice cream and Nazis: Moral externalization and the evolution of human cooperation. *Behavioral and Brain Sciences,* 41: 1-13.

Stich, Stephen. (2018). The quest for the boundaries of morality. In *The Routledge Handbook of Moral Epistemology*, eds. Aaron Zimmerman, Karen Jones, and Mark Timmons. Routledge.

Sterelny, Kim. (2010). Moral nativism: A sceptical response. *Mind and Language* 25(3): 279-297.

Street, Sharon. (2006). A Darwinian dilemma for realist theories of value. *Philosophical studies*, 127(1): 109-166.

Sturgeon, Nicholas. (2006). Moral Explanations Defended. In *Contemporary Debates in Moral Theory*, ed. James Dreier. Blackwell.

Turiel, Elliot. (1983). *The Development of Social Knowledge: Morality and Convention*. Cambridge University Press.

Van Roojen, Mark. (2015). *Metaethics: A Contemporary Introduction*. Routledge.

Vavova, Katia. (2014). Debunking evolutionary debunking. In *Oxford Studies in Metaethics, Vol 9*, ed. Russ Shafer-Landau. Oxford University Press.

Vavova, Katia. (2015). Evolutionary debunking of moral realism. *Philosophy Compass*, 10(2): 104-116.

Wielenberg, Erik. J. (2010). On the evolutionary debunking of morality. *Ethics*, 120(3): 441-464.