CrossMark

Old Wine in New Bottles

Evolutionary Debunking Arguments and the Benacerraf-Field Challenge

Michael Klenk[1]

Abstract Evolutionary debunking arguments (EDAs) purport to show that robust moral realism, the metaethical view that there are non-natural and mind-independent moral properties and facts that we can know about, is incompatible with evolutionary explanations of morality. One of the most prominent evolutionary debunking arguments is advanced by Sharon Street, who argues that if moral realism were true, then objective moral knowledge is unlikely because realist moral properties are evolutionary irrelevant and moral beliefs about those properties would not be selected for. However, no evolutionary, causal explanation plays an essential role in reaching the argument's epistemological conclusion. Street's argument depends on the Benacerraf-Field challenge, which is the challenge to explain the reliability of our moral beliefs about causally inert moral properties. The Benacerraf-Field challenge relies on metaphysically necessary facts about realist moral properties, rather than on contingent Darwinian facts about the origin of our moral beliefs. Attempting to include an essential causal empirical premise yet avoiding recourse to the Benacerraf-Field problem yields an argument that is either self-defeating or of limited scope. Ultimately, evolutionary, causal explanations of our moral beliefs and their consequences do not present the strongest case against robust moral realism. Rather, the question is whether knowledge of casually-inert, mind-intendent properties is plausible at all.

✉ Michael Klenk
   m.b.o.t.klenk@uu.nl

[1] Utrecht University, Utrecht, Netherlands

Springer

1 Introduction[1]

1.1 Background

Much current metaethical research focuses on the consequences of the assumption that evolutionary forces shaped our moral cognition. If moral cognition is an adaption, then our capacity to produce moral judgements, and, to some extent, the content of our moral judgements and beliefs are the products of natural selection (Street 2006: 115–121; Joyce 2006: chapter 4).[2] While there are considerable difficulties in establishing this claim, there are also persuasive arguments in its favour (Fraser 2014; Buchanan and Powell 2015; Kitcher 2011: 17–166).

Some proponents of evolutionary debunking arguments[3] purport to show that robust moral realism[4] (the view that there are non-natural and mind-independent moral properties and facts about which we can have knowledge) is inconsistent with the evolutionary influence on our moral cognition, and so should be rejected (Street 2006; Ruse and Wilson 2006).[5] These debunkers reach epistemological conclusions by taking into account *Darwinian considerations* about the origins of our moral beliefs. The grand ambition of debunkers is to use a well-established empirical account of the origins of our moral beliefs to discredit moral realism. The argumentative strategy is intriguing: evolutionary theory seems correct, and it would be an immense cost to any metaethical theory if it could not incorporate it. Sharon Street put forward the most influential evolution-based critique.

Street reaches the sceptical epistemological conclusion that we have sufficient reason to doubt the truth of *all* of our moral beliefs and suggests that Darwinian considerations do important work in the argument (Street 2006: 109) and, eventually, 'settle the [realism vs. anti-realism] debate in favour of the anti-realist view' (Street 2008a: 214). She argues that if robust moral realism were true, then moral knowledge is unlikely because mind-independent moral properties were evolutionary irrelevant and thus beliefs about these properties were not selected for.

An ever increasing number of writers are concerned with the metaethical implications of evolutionary theory, suggesting that much in metaethics depends on one or the other explanation of the causal origins of our moral beliefs (e.g. Ruse and Richards forthcoming; Artiga 2015; Deem 2016; Fraser 2014).

---

[1] For helpful comments on this paper I thank Liam Deane, Hein Duijf, Jonathan Goodman, Wouter Kalf, Herman Philipse, Hanno Sauer, Joeri Witteveen, and two anonymous referees of this journal. Thanks also to audiences at the 4th Annual Conference of the Colombian Philosophy Society, the Objectivity in Ethics conference in Utrecht, the Science vs Common Sense conference in Amsterdam, and the European Epistemology Network Annual Meeting 2016.

[2] The distinction between the *capacity* to make moral judgements and their *content* is crucial for EDAs in metaethics, and it is controversial whether the empirical claim about evolutionary influences on the content of our moral beliefs is well-supported (e.g. FitzPatrick 2015; Mogensen 2016; Buchanan and Powell 2015). I assume it here for the sake of the present discussion.

[3] I focus on Street's *global* EDA that intends to lower the status of *all* moral judgements (Kahane 2011).

[4] Henceforth 'realism' designates robust moral realism. 'Moral properties' entail 'moral facts'. Exemplary proponents are Enoch (2011), Shafer-Landau (2003), and Wielenberg (2014). Street means to include moral naturalism as a target of the EDA, too, and while most see moral naturalism well placed to answer the evolutionary challenge (e.g. Enoch 2010: 422), others have recently called this into question (Barkhausen 2016; Bogardus 2016).

[5] Not all areas of discourse are threatened by evolutionary considerations. For example, the truth of our beliefs about ordinary objects is consistent with their purported evolutionary background (Wilkins and Griffiths 2013; Boudry and Vlerick 2014).

1.2 Argument and Relevance

Street's argument against robust moral realism depends on a version of the Benacerraf-Field challenge, according to which moral realists must show that it is in principle possible to explain how we can have reliable beliefs about the moral properties postulated by moral realism. The Benacerraf-Field challenge does *not* rely on a premise about human evolution, or any alternative empirical explanation of the origins of our moral beliefs. Hence, contrary to the received view, no evolutionary causal explanation plays an essential role in reaching the debunking argument's epistemological conclusion.

But attempts to include an essential empirical premise about the origins of our moral beliefs, while not relying on a version of the Benacerraf-Field challenge, do not succeed: the EDA would be self-defeating if applied to *all* moral beliefs, or it would debunk only some of our moral beliefs, but not all of them. Hence, Street's *evolutionary, causal* debunking argument against robust moral realism fails.

This should alter the locus of the metaethical debunking debate. The crucial point is not what follows from evolutionary theory or alternative causal explanations of our moral beliefs, but whether knowledge of causally-inert, mind-independent, and irreducibly normative properties is possible at all.

1.3 Structure

I proceed as follows. Section 2 introduces Street's EDA and presents a possible escape route for robust realists, which highlights a crucial premise of the EDA. Section 3 shows that this crucial premise is conceptual rather than empirical and that the argument depends on the Benacerraf-Field challenge. Section 4 argues that attempts to reintroduce an essential empirical premise fail.

2 Street's Evolutionary Debunking Argument and an Escape Route for Realists

2.1 Street's EDA

The first premise of Street's argument is the supposition, for the sake of argument, that robust moral realism is true (Street 2006: 109, 121).

REALISM: Robust moral realism is true.

Evolutionary debunking arguments generally rest on claims about our moral beliefs, typically about their origins. The factual premises of the argument – claims about our moral beliefs – provide the grounds for the 'debunking premise,' which states the reason(s) for viewing with suspicion the facts about our moral beliefs that were established by the factual premises (Kahane 2011). Thus, the factual premises need to be combined with a debunking premise to yield a conclusion.

What are Street's 'factual' premises? Almost all discussions of her argument (e.g. Copp 2008; Vavova 2015; Artiga 2015) follow Street in interpreting her first factual premise as stating that natural selection, and other evolutionary factors, have had a 'tremendous,' albeit indirect, influence on the content of human evaluative attitudes (Street 2006: 113). However, a careful

reading shows that Street is, somewhat in tension with the tenor of her piece, not committed exclusively to evolutionary factors: she also mentions that 'other causal influences can shape our evaluative judgements' (Street 2006: 120), and as long as there is 'some sort of causal explanation' (Street 2006: 153) of our evaluative beliefs, 'whether Darwinian or otherwise' (Street 2006: 155), this fact seems sufficient 'for the purposes of the argument' (Street 2006: 158). Hence, Street's challenge is best understood if her first factual premise is construed as making a broad claim about the *causal* origins of our moral beliefs. Let's call this premise INFLUENCE[6]:

INFLUENCE: All our moral beliefs were influenced by causal forces.

Street argues that the causal forces that shaped our moral beliefs are not systematically connected to moral truths and argues the realist must 'take a position on what relation there is, if any, between the selective forces that have influenced the content of our evaluative judgements […] and the independent evaluative truths that realism posits' (Street 2006: 121). One option is to assume that there is no connection between the causal forces that influenced our basic evaluative dispositions and the moral truth. The other option is to affirm that there is such a connection. The latter option is rejected by most realists (FitzPatrick 2014: 241).[7] Hence, the second factual, empirical premise in Street's argument is about the independence of causal influences on our moral beliefs and the moral truth. Let's call this premise MISLEAD:

MISLEAD: The causal forces that influenced our moral beliefs have no systematic relation to the moral truth.

Street argues that REALISM, INFLUENCE, and MISLEAD lead to a sceptical conclusion (Street 2006: 122). We can fill in the required DEBUNKING CONDITIONAL:

DEBUNKING CONDITIONAL[8]:   If REALISM, INFLUENCE, and MISLEAD are true, then we have sufficient reason to doubt the truth of all our moral beliefs.

DOUBT (conclusion)[9]:   So, we have sufficient reason to doubt the truth of all our moral beliefs. (from REALISM, INFLUENCE, MISLEAD, and DEBUNKING CONDITIONAL).

The independence of evaluative truths and the causal origins of our moral beliefs bodes trouble for realism. Intriguingly, the argument seems to have sweeping metaethical

---

[6] Vavova (2014, 2015) uses the same term but refers to evolutionary forces exclusively.

[7] Assuming that there is a connection between evolutionary forces and moral truth confronts realists with an inference to the best explanation (IBE): the content of our moral beliefs can be explained without invoking *moral* properties. Qua parsimony, the evolutionary explanation is better than the realist's explanation. I do not consider the 'EDA as IBE'interpretation here since it would leave much room for the realists to reply. They might claim, for instance, that there are further, prudential reasons to stick with realism (Enoch 2011; Copp 2008: 190).

[8] The epistemic principle behind the DEBUNKING CONDITIONAL is controversial and a crucial point of the debunking debate (cf. Bedke 2014; Bogardus 2016; Vavova 2015). I assess two interpretations in Section 3.3, to support my point about the irrelevance of genealogical claims, but I do not address the debate about the correct *epistemic* principle in this paper.

[9] Street's argument extends beyond DOUBT: she concludes that robust realism ought to be rejected (Street 2006: 135). I am sympathetic to Vavova's interpretation of the argument as the 'beginning of a *reductio* of realism' (Vavova 2015: 108). This makes sense if we make explicit that REALISM entails the possibility of moral knowledge, and an auxiliary premise that states that a metaethical theory ought to be rejected if it both entails that moral knowledge is possible and at the same time gives us reason to doubt the truth of all our moral beliefs. My argument focuses on the steps that lead to DOUBT, so I need not make the additional steps towards the rejection of realism explicit.

implications based on empirical evidence. If debunkers could thereby challenge robust moral realism, they would change our concept of 'man's position in the universe' through reference to solid empirical facts – truly in 'Darwin's spirit' (Mayr 2003: xxi).

2.2 An Escape Route for Realists

However, thus far the argument does not secure a sceptical conclusion as it ignores an easily overlooked complication. Since we assume REALISM, moral properties do exist, and *at least some* of our moral beliefs could, in principle, reliably track them (cf. FitzPatrick 2015; Huemer 2016; Wielenberg 2014). This would be enough to reject the claim that empirical evidence about the origins of our beliefs gives us sufficient reason to doubt the truth of *all* of our moral beliefs. Thus, realists might grant REALISM, INFLUENCE and MISLEAD, but reject the DEBUNKING CONDITIONAL. Realists might argue:

CORRECT:	At least some of our moral beliefs are likely to track realist moral properties.[10]
ANTI-DEBUNKING CONDITIONAL:	If REALISM and CORRECT are true then we do *not* have sufficient reason to doubt the truth of all our moral beliefs.
TRUST (conclusion):	So, we do *not* have sufficient reason to doubt the truth of all our moral beliefs. (from REALISM, CORRECT, and ANTI-DEBUNKING CONDITIONAL).

We can infer TRUST from CORRECT and the ANTI-DEBUNKING CONDITIONAL since CORRECT gives us reason to believe that at least some of our moral beliefs are non-accidentally connected to the moral properties. The DEBUNKING CONDITIONAL and the ANTI-DEBUNKING CONDITIONAL cannot both be true, but which premise realists have to accept depends on whether INFLUENCE and MISLEAD rule out CORRECT.

As mentioned above, the realists this paper is concerned with are committed to a non-naturalistic conception of mind-independent, causally-inert moral properties (e.g. Enoch 2011: 7, 159; Shafer-Landau 2003: 107).[11] This suggests that their claim to CORRECT is unaffected by INFLUENCE and MISLEAD because realists' accounts of how moral believers track moral properties do not rely on causal relations between moral properties and moral beliefs in the first place. Whether we believe that stealing is wrong, for instance, because of some Darwinian force, or because of our upbringing, or because of some other causal factor, realists can hold on to the claim that the belief is non-accidentally true because debunkers have not ruled out realists' non-causal accounts of reliable access to its truth.[12] How do realists defend the claim that our moral beliefs are likely to track realist moral properties? They might claim, for instance, that moral beliefs are reliably formed through rational intuition or direct perception, or that there is a constitutive relation between moral

---

[10] The notion of 'tracking' as I use it in this paper should be broadly understood as capturing any non-accidental, systematic connection between our moral beliefs and realist moral properties.

[11] There are realists who claim that moral properties are causally efficacious. For example, Oddie (2009) and the Cornell realists. It is worth repeating that I am solely concerned with Street's EDA as it applies to *robust* moral realism, which entails that moral properties are causally inert.

[12] Vavova (2015) suggests that the empirical premises alone suffice to rule out CORRECT, if they provide evidence of error. I show that this route fails in Section 4.

properties and moral beliefs, or that divine revelation plays a role in shaping our moral beliefs (cf. Bogardus 2016: 642f; Bengson 2015; Cuneo and Shafer-Landau 2014; Huemer 2016: 1987–1989).[13]

Hence, debunkers have to consider CORRECT as a live option and realists would likely insist that, in an argument that begins with REALISM, debunkers first have to show that the realist's 'escape route' via CORRECT fails.

Debunkers have two options. They can reject CORRECT, which requires showing that none of our moral beliefs is likely to track moral properties. If successful, they could infer DOUBT. Alternatively, as we will see, they may argue that the DEBUNKING CONDITIONAL, and the inference to DOUBT, are valid despite CORRECT.

3 Street's Debunking Argument Relies on the Benacerraf-Field Challenge

3.1 The Benacerraf Field Challenge

Note that debunkers cannot just assume without argument that CORRECT is false because that would beg the question in the argument against realism. As indicated above, their empirical premises do not directly refute CORRECT either. Therefore, debunkers need a different argument to show that CORRECT fails; such an argument depends on the Benacerraf-Field challenge.

The Benacerraf-Field challenge is a problem for knowledge of mind-independent, causally inert entities. The worry originates in Benacerraf's work on the possibility of mathematical knowledge (Benacerraf 1973). He writes:

'[…] I think, that something must be said to bridge the chasm, created by … [a] realistic and platonistic interpretation of mathematical propositions, between the entities that form the subject matter of mathematics and the human knower.' (Benacerraf 1973: 675)

The abstract entities postulated by mathematical Platonism share two important features with moral properties: mind-independence and causal inertness. Therefore, as Peacocke recognises, Benacerraf's problem concerning mathematical Platonism seems to be a problem for moral realism, too:

'What Benacerraf [...] asserts about mathematical truth applies to any subject matter. The concept of truth, as it is explicated for any given subject matter, must fit into an overall account of knowledge in a way that makes it intelligible how we have the knowledge in that domain that we do have. (Peacocke 1999: 1-2)

Benacerraf's worry presupposes a causal theory of knowledge, but Harty Field's development of the challenge makes it independent of this precondition. Thus it also applies to robust moral realists, who commonly reject a causal theory of knowledge. Field challenges realists to explain:

'[…] how our beliefs about [abstract] entities can so well reflect the facts about them [...] [I]f it appears in principle impossible to explain this, then that tends to undermine the

---

[13] I do not address the merits of possible realist replies in this paper. My sole concern here is to show that the debunking argument depends on the Benacerraf-Field challenge. I am sceptical about the ultimate viability of the mentioned realist replies, but my point here is simply that they are live options and not ruled out by the empirical premises of the argument, which is why debunkers rely on the Benacerraf-Field challenge counter these claims.

belief in mathematical entities, despite whatever reason we might have for believing in them ' (Field 1989: 26).

Field's adapted challenge '[…] depends on the idea that we should view with suspicion any claim to know facts about a certain domain if we believe it impossible in principle to explain the reliability of our beliefs about that domain' (Field 1989: 232–233). The Benacerraf-Field challenge raises suspicion about the reliability of beliefs about causally inert and mind-independent properties. This is problematic because it is, other things being equal, a cost of a theory if it treats knowledge about properties presupposed by the theory as merely accidental or altogether inexplicable. We can now see that the Benacerraf-Field arises because of the metaphysical properties of abstract entities and that moral properties are relevantly similar because they are also mind-independent and causally inert. If there is no way in which robust realists can explain how our moral beliefs are likely to track realist moral properties (i.e. give an argument for CORRECT), then moral knowledge would indeed be a startling fact.

3.2 Benacerraf-Field 2 - Darwin 0

The Benacerraf-Field challenge affords a way to argue that it is likely that none of our moral beliefs can reliably track mind-independent and causally inert moral properties (cf. Clarke-Doane 2017). In other words, Benacerraf and Field allege that the nature of moral properties makes it likely that CORRECT is false. If their allegation is correct, then we have reason to accept the DEBUNKING CONDITIONAL and not the ANTI-DEBUNKING CONDITIONAL.[14]

Debunkers rely on the Benacerraf-Field challenge to reject CORRECT and to force realists to accept the DEBUNKING CONDITIONAL. This is because the empirical claims of the debunking argument leave open the possibility of non-causal, truth-tracking determinants of our moral beliefs and so CORRECT is not falsified (cf. Bogardus 2016). Conversely, if realists could vindicate CORRECT, we would have reason to accept the claim that we do *not* have sufficient reason to doubt the truth of all our moral beliefs. Debunkers have to challenge realists, in Benacerraf's and Field's spirit, to explain the reliability of our moral beliefs to get their sceptical argument off the ground.

Importantly, the Benacerraf-Field challenge is a conceptual, normative challenge. It is conceptual because it relies on the epistemological suspicion that knowledge of mind-independent, causally inert moral properties is inexplicable,[15] and it is normative because it demands that proponents of such properties explain how our beliefs about these properties could be reliable. Debunkers rely on it in their argument against robust moral realism. Therefore, empirical, causal considerations, as in INFLUENCE or MISLEAD, are not sufficient to reach a sceptical conclusion against robust moral realism. The empirical debunking argument turns out to rely on a conceptual, normative claim that is based on the metaphysical nature of moral properties.

In addition, it seems that the reliance on the Benacerraf-Field challenge makes the empirical premises of the debunking argument redundant, too. Once the Benacerraf-Field challenge establishes that none of our moral beliefs are likely to track moral properties (hence, that CORRECT is false), the empirical premises INFLUENCE and MISLEAD do not provide any additional sceptical

---

[14] This paper does not assess how the Benacerraf-Field challenge fares against CORRECT. Clarke-Doane (2017) and, in parts, Bogardus (2016) address this issue.
[15] The precise epistemic principles behind the Benacerraf-Field challenge are controversial, cf. Clarke-Doane (2017).

oomph: blocking the realist's escape route via CORRECT would already secure the sceptical conclusion that we have sufficient reason to doubt the truth of all our moral beliefs.

Debunkers might object: even though the empirical premises, INFLUENCE and MISLEAD, are not sufficient to reach the sceptical conclusion, they seem to provide *additional reason* to doubt realism. Let us consider an example to make the objection vivid and then see how it can be answered[16]:

> Case 1: Suppose that Alf ingests an anti-maths drug which makes his mathematical beliefs unreliable; this gives him sufficient reason to doubt the truth of his mathematical beliefs.
> Case 2: Mathematical objects are causally inert, they cannot influence our mathematical beliefs; this, too, gives Alf sufficient reason to doubt the truth of his mathematical beliefs. Although case 1) and 2) lead to the same conclusion, they cannot thereby be reduced to one another. Case 1 gives Alf *additional* reason to doubt the truth of his mathematical beliefs.

The anti-maths drug case resembles the empirical, causal debunking argument: the evidence about Alf ingesting the drug seems roughly analogous to the empirical claim about the causal origins of our beliefs. Clearly, this gives us additional reason to be sceptical about Alf's maths-beliefs.

However, the anti-maths drug case is different from the debunking of moral realism. In the case of the anti-maths drug, we might legitimately stipulate that there is no way for Alf to have at least some reliable true maths beliefs: we just assume that the anti-maths drug rules out this possibility. However, this stipulation cannot be made in the causal debunking argument, since debunkers cannot presuppose that it is impossible for realists to fine-tune their moral beliefs without begging the question. Instead, debunkers need to argue for it, and they rely on the causal inertness and mind-independence of moral properties to fashion an argument to that effect. This, we have seen, is a version of the Benacerraf-Field challenge. Hence, even though the sceptical conclusions of the maths-belief cases above are independent, and the sceptical oomph that they provide is additive, the empirical debunking argument against realism is not independent of the conceptual argument. To the contrary, reaching the conclusion of the empirical argument *depends* on the conceptual argument against CORRECT. So, in contrast to the maths-case, to pose the empirical debunking argument requires a successful conceptual debunking argument – and it is doubtful whether the empirical debunking argument would give us any *additional* reason in any interesting sense since realists would have had to concede defeat already.

More specifically, the sense in which the sceptical conclusion of the empirical causal debunking argument is *additive,* that is, giving us *additional* reason to be sceptical about realism, is additive only in the weak sense in which the following italicised variants of INFLUENCE give us additional reason to doubt the truth of specific realist moral beliefs:

> Case 3: Anton believes that donating to charities is good, not because the belief is reliable and true, but because *he heard his neighbour say it.*
> Case 4: Bob believes that eating animals is wrong, not because the belief is reliable and true, but because *he was moved by the cover of Peter Singer's 'Animal Liberation'.*
> Case 5: Cliff believes that gender equality is just, not because the belief is reliable and true, but because *he has an evolved fairness sense.*

---

[16] Thanks to an anonymous referee for raising this objection.

These particular instances of INFLUENCE surely seem spurious, and the impression created by Singer's book cover, for example, is in itself not a good indicator of the truth about animal ethics. But if we knew that Bob studied the book's content and, as a robust realist sympathetic to Singer's view would claim, thereby formed reliable beliefs about the truth on animal ethics, we should conclude that doubting Bob's belief is unwarranted. Given the realist's commitment to non-causal explanations of moral knowledge, the evidence about the actual causal origins of Anton's, Bob's, and Cliff's beliefs are not, in themselves, troubling. It may raise our suspicion, but it cannot justify our suspicion if we have not ruled out their truth-tracking ability first. What debunkers need is the claim that Anton, Bob, and Cliff are unlikely to believe what they believe *because* it is true. However, once debunkers secure that point, we need not worry about the actual causal origins of our beliefs anymore. Realists would already be in deep trouble if their explanation of non-causal, nonetheless reliable belief-forming methods failed; additional genealogical considerations would not worsen the problem for robust realists.

Hence, there are good reasons to think that the empirical premises, and the DEBUNKING CONDITIONAL, do not suffice to debunk moral realism, and also that *adding* the empirical premises does not *add* problems for robust realism.

3.3 Empirical Premises are not Required

Perhaps, however, debunkers might think that their empirical premises are necessary to reject CORRECT. Two prominent interpretations of the DEBUNKING CONDITIONAL, suggested by Street (2006) and taken up in the literature, suggest that they are not. The mind-independence and causal inertness of moral properties is doing all the work.

First, many have suggested that the debunking premise rests on probabilistic considerations. In the words of Shafer-Landau, the odds of adopting a true moral belief are low because 'our actual moral beliefs represent only a small portion of all possible moral beliefs' (Shafer-Landau 2012). It makes sense to interpret Street's evolutionary debunking argument as an argument that invokes probability, particularly when she writes:

> 'Of course it's possible that as a matter of sheer chance, some large portion of our evaluative judgements ended up true, due to a happy coincidence between the realist's independent evaluative truths and the evaluative directions in which natural selection tended to push us, but this would require a fluke of luck that's not only extremely unlikely, in view of the huge universe of logically possible evaluative judgements and truths, but also astoundingly convenient to the realist.' (Street 2006: 122)[17]

The assumed fact that illegitimate forces had an impact on our moral beliefs is taken to imply that our actual moral belief system has only a slight chance of matching the correct moral belief system. Street concludes that many of our actual moral beliefs are likely to be 'off track' (Street 2008b: 208). On this interpretation, the DEBUNKING CONDITIONAL is motivated by alluding to the conceptual possibility of there being many different possible moral belief systems but no way of discriminating the *correct* moral belief system. Thus, the low probability of adopting just the right moral belief system amongst countless false moral belief systems is sufficient reason to doubt the truth of our moral beliefs. However, for the claim that we are unable to identify the correct moral beliefs system to be convincing, we have to assume that CORRECT has faltered. Hence:

---

[17] See also Street (2008b: 208) and Street (2011: 14).

LOW PROBABILITY:	If REALISM is true and CORRECT is false, then there are many conceptually possible moral belief systems and no way of telling whether the one we currently adopted is the correct one.
LP-DEBUNKING COND.:	If LOW PROBABILITY is true, then we have sufficient reason to doubt the truth of all our moral beliefs.

Others interpret the argument's debunking claim as depending on considerations about counterfactuals. This interpretation is formulated most clearly by Clarke-Doane (2012). He argues that the debunking premise of the argument is substantiated by the following conditional:

'[…] if our moral beliefs were the products of evolutionary forces, then those forces would be 'non-truth-tracking'—that is, [..] if we were selected to have certain moral beliefs at all, then we would not be selected to have true moral beliefs' (Clarke-Doane 2012: 325).

This would create a problem for robust moral realism because 'if we were not selected to have true moral beliefs, then had the moral truths been very different, our moral beliefs would have been the same' (Clarke-Doane 2012: 319). The essence of Clarke-Doane's understanding of the debunking premise is the claim that our moral beliefs are the products of 'non-truth-tracking' forces. However, Clarke-Doane's counterfactual also depends on the claim that non-causal determinants of our moral beliefs cannot play a role; hence that CORRECT is false:

COUNTERFACTUAL:	If REALISM is true and CORRECT is false, then we would still have the same moral beliefs even if the moral truths would be different.
C-DEBUNKING COND.:	If COUNTERFACTUAL is true, then we have sufficient reason to doubt the truth of all our moral beliefs.

Both interpretations of Street's DEBUNKING CONDITIONAL seem plausible. Both could be used to infer a version of DOUBT. However, these interpretations do not require an empirical claim about the origins of our moral beliefs to be convincing. Rather, they depend on the conceptual claim that moral properties are similar to the properties of abstract entities, in that they are causally inert and mind-independent. If it were not for this fact, neither interpretation would have much bite. Consider LOW PROBABILITY. Our inability to tell whether our current moral belief system is the one that matches the facts depends on (the conceptual possibility of) our beliefs being insensitive to the moral properties. That, in turn, depends on the moral properties being mind-independent and causally inert. Consider COUNTERFACTUAL. Our moral beliefs would be insensitive to changes in the moral truth only if the moral properties, which constitute the moral truth, are mind-independent and causally inert.

Therefore, both interpretations of the DEBUNKING CONDITIONAL require the same observation about the nature of moral properties to be convincing, which is, as we have seen, not an observation of any contingent empirical fact, but of necessary metaphysical facts.

Recent discussions of Street's debunking argument frequently miss this point. Bogardus, for instance, looks into the epistemic principle that underlies the argument, which indeed is a key question, but then expects, mistakenly, that it somehow needs to combine with 'the facts of evolution' to create a sceptical conclusion (Bogardus 2016: 636). Vavova claims that the argument rests on a claim about evolutionary influences, but does not emphasise that this is the case only if CORRECT is rejected (Vavova 2015: 108). Referring to Street (2006), Artiga (2015) and Deem (2016) defend elaborate empirical hypotheses about the compatibility of particular

causal explanations of our moral beliefs and moral realism, ignoring that the issue with CORRECT cannot be settled by any causal explanation.

When the crucial nature of moral properties is noted, for instance by Bedke (2014) who refers to the moral beliefs being 'oblivious' to the moral properties or by Crow (2015), then it is too often overlooked that no causal claim of any kind is required to make the argument work. Similarly, Clarke-Doane (2012) observes, in line with Street (2006), that the particular details of the causal history are irrelevant, but he does not argue that no causal claim is needed for the argument (Clarke-Doane 2012: 326). Enoch suggests that the core issue is to explain the correlation between the moral beliefs that we take to be true and realist moral properties, which resembles the Benacerraf-Field challenge (Enoch 2010: 421). But still, his answer to the challenge involves a causal explanation of our moral beliefs, which wrongly suggests that the causal history of our moral beliefs, and competing elucidations of it, are indeed the problem that underpins the debunking argument. Framing the issue like this is misleading as it puts too much emphasis on one or the other interpretation of the causal origins of our moral beliefs, on the soundness of such explanations, or on the seemingly extraordinary fact that we can illuminate the causal background of our moral beliefs. The crucial fact is, as we have seen, that the nature of moral properties puts the realist's claim to CORRECT in jeopardy.

The interpretations of Street's DEBUNKING CONDITIONAL considered here support the conclusion that the debunking argument relies on the Benacerraf-Field challenge and, conversely, on conceptual claims about the nature of realist moral properties. This does not show that *any* debunking argument that leads to DOUBT relies on a conceptual argument against CORRECT rather than on INFLUENCE and MISLEAD. However, the above already gives us strong reasons to suppose that the causal inertness and mind-independence of moral properties (and the resulting conceptual, epistemic problem) lie at the heart of other supposed causal debunking arguments against robust moral realism, too. The alleged Darwinian debunking argument contains the spirit of Benacerraf and Field, but not that of Darwin.

3.4 Intermediate Conclusion

Street's evolutionary debunking argument, an apparently empirically-informed argument against realism, turns out to rest on the conceptual claim that moral properties are mind-independent and causally inert and that this makes reliably true moral beliefs unlikely. The latter observation 'crowds out' empirical considerations from the factual premises of the argument.

Debunkers should be wary in accepting this conclusion. Although the Benacerraf-Field challenge appears to be a powerful epistemological challenge against robust moral realism, it fails to constitute an argument against realism that is based on an empirical premise about the causal origins of our moral beliefs. Debunking arguments would be misleading without an empirical premise because the argument would render empirical premises redundant to making the argument sound.

4 Street's EDA does not Succeed without the Benacerraf-Field Challenge

Debunkers might concede that one strategy against CORRECT relies on the Benacerraf-Field challenge, but argue that another route shows how variants of the empirical premises INFLUENCE and MISLEAD can force realists to accept the DEBUNKING CONDITIONAL without a conceptual

argument against CORRECT. This is how the debunking argument is often understood, and this section will attempt to dispel this myth.

The myth is this: that debunking arguments provide us with evidence that most or even all of our moral beliefs are influenced by such-and-such causal processes; and we see that some of these beliefs are false. Hence, since we know about the pervasive influence of causal forces on our beliefs and the occasional distorting effect, we can infer that we should be sceptical about the truth of *all* our moral beliefs; they do not seem to track moral properties, despite the possible corrective influence of the non-causal relations postulated by robust realists.[18] So, the myth goes, realism is debunked without relying on a version of the Benacerraf-Field challenge.

Why do debunkers need a substantive moral claim? Because to show that causal influences are distorting, without relying on the claim[19] that 'causal influences were the *only* possible determinants of our moral beliefs,' requires evidence that our beliefs are distorted despite possible alternative, non-causal determinants of our beliefs (like those the realists claim exist). The required evidence can only be provided by showing that the causal influences in question lead us to have false moral beliefs, and to determine that we have false moral beliefs debunkers need to commit to substantive moral claims about which moral beliefs are false.

However, assuming the truth of substantive moral claims is problematic in an argument that raises doubt about the truth of *all* our moral beliefs. Such a 'global' debunking argument would call these very assumptions into question.[20] Put another way, an inductive argument supporting the DEBUNKING CONDITIONAL would cast considerable doubt on the truth of all of our moral beliefs. However, in that case, we would also have reason to doubt the truth of the substantive moral claim that debunkers rely on to provide us with evidence of distortion: the argument would be self-defeating.

Debunkers might object that there are cases in which we can legitimately rely on substantive truths about the beliefs that we seek to debunk. Consider the following case[21]:

> Suppose I provide you with evidence that your car's thermometer is distorted by the heat of the car's engine: 'Your car thermometer displayed +7 degrees, but it was -5!' After receiving this information, you ought to doubt all future readings of your car thermometer because it might be distorted at any given moment.

In cases like the thermometer case, where a particular belief-producing faculty is being debunked, we can rely on our knowledge about the truth in question: we have no reason to assume that *our* beliefs about the temperature are unreliable. However, the debunking case is not analogous. The debunkers' 'fact-checking' of the moral truth is called into question by their own argument. They attempt to debunk moral cognition entirely, and thus cannot rely on beliefs that are safeguarded from the potentially debunking information. Hence, debunkers

---

[18] Street's terminology about '*distorting*' causal influences on our moral beliefs may be partly responsible for this myth (Street 2006: 155). Street uses 'distorting'in a weak sense*, in which case there is only a non-truth-tracking but no truth-tracking force present. This is suggested by her Bermuda analogy: setting out by boat for Bermuda and letting the wind and tides determine one's course is not clever because, without using the sail and the rudder, the wind and waves do not push you to your target (Street 2006: 121–122). However, at this point in the dialectic debunkers have to assume that correction is possible and thus understand distortion in a *strong sense*: a non-truth-tracking force distorts even though a truth-tracking force might be present, too. Examples of this kind of distortion are the forces of waves that erode a cliff or the winds at Cape Horn that push a capable sailor off his intended course.
[19] Which would be an instance of the Benacerraf-Field challenge.
[20] 'Global' is also used by Kahane (2011) and Shafer-Landau (2012).
[21] Thanks to an anonymous referee for raising this objection by suggesting this case.

cannot rely on their own moral beliefs in the debunking argument; to avoid self-defeat they must stay agnostic about the moral truth.[22] However, in that case, their global debunking argument does not work because, without evidence of distorted, false moral beliefs (via a substantive moral claim) or evidence of the absence of possible corrective influence (via the Benacerraf-Field challenge) we do not get sufficient reason to doubt the truth of *all* of our moral beliefs.

Second, while this particular problem can be avoided by attempting to debunk only particular moral beliefs, that strategy fails, too, because the inference to DOUBT would fail: we cannot infer sufficient reason to doubt *all* our moral beliefs from the falsity of *one* particular belief. The scope of the debunking argument would be drastically reduced.

Take the example of a xenophobe who believes that all non-group members must be killed. A plausible causal explanation of this belief might be the xenophobe's evolved tendency to distinguish between members of his in-group and members of the out-group (cf. Wielenberg 2014):

XEN-INFLUENCE: The belief that all non-group members must be killed was influenced by evolutionary forces.

To show that the cauasl influence reported in XEN-INFLUENCE is distorting, despite the possibility that moral beliefs non-causally track moral properties, we need to point out that it is not, in fact, the case that all non-group members must be killed.

Now we have evidence of a causal influence on the xenophobic belief and evidence that the belief is false. In the case of the xenophobe, debunkers might claim, CORRECT evidently faltered. But this is no victory for debunkers: they cannot infer DOUBT from the falsity of one particular moral belief. The same holds if debunkers substitute the particular xenophobic belief with *subsets* of moral beliefs, say all deontological beliefs, to increase the impact of their argument (cf. Greene 2008). They could argue:

DEONTOLOGY-INFLUENCE: All deontological moral beliefs are influenced by such-and-such causal forces.

But, again, the argument is limited in scope: it targets only a specific subset of moral beliefs, distinguished by its content or by its origin, but not *all* moral beliefs as in the desired 'global' debunking argument. The best that can be hoped for is an inference to doubt the truth of those beliefs that are part of particular subsets. So, 'going local' does not afford the desired inference to DOUBT and 'going global' puts the required substantive moral assumption into jeopardy.

Moreover, there is an issue with substantive moral assumptions even on the local level. It appears that causal considerations are, yet again, superfluous. We saw, for instance, that premises with substantive moral content, like NOT-XEN, are crucial for inferring sufficient reason to doubt the truth of particular moral beliefs, or subsets of moral beliefs. However, in that case, the substantive moral claim seems to do the debunking, not the factual claim about the genealogy of the beliefs in question.

For instance, suppose that I try to debunk moral beliefs that are moderated by disgust, which would involve an empirical claim about the subset of all moral beliefs moderated by disgust (cf. Sauer 2012). However, the sole fact that these beliefs are moderated by disgust should not bother us in the least. We should be bothered only when we are convinced that moral beliefs moderated by disgust are often false. However, disgust-moderated moral beliefs

---

[22] Shafer-Landau (2012), who distinguishes between *agnostic* and knowledge-based EDAs.