

University of Zurich
Department of Philosophy
Term Paper in *Philosophy of the 19th and 20th Century*
Supervisor: Peter Schulte, PD Dr.

**Moorean Metaethics and Evolutionary Metaethics:
Moore's Principia Ethica facing Street's Darwinian Dilemma**

Autumn Term 2022
Submission Date: 1 December 2022
Number of Characters (with Spaces): 44355

Eleos Arete Citrini

Abstract G. E. Moore’s arguments for moral non-naturalism, moral intuitionism, and their pairing in his 1903 book *Principia Ethica* made the latter a revolutionary treatise in ethics with a perhaps unmatched legacy in metaethics, inadvertently spurring the search for metaethical alternatives to moral realism. The recent rise of evolutionary debunking arguments such as the famous one in Sharon Street’s 2006 paper *A Darwinian Dilemma for Realist Theories of Value* can be seen as an outgrowth of that search. I discuss a few objections to Street’s argument by two defenders of moral non-naturalism, finding that opting for either of the dilemma’s two horns offers potentially promising avenues for moral realists, however I tentatively conclude that Street’s argument poses a very intriguing and enduring challenge for moral realists and even more so for Moorean metaethicists.

Keywords metaethics · Moore’s *Principia Ethica* · Street’s *Darwinian Dilemma* · moral non-naturalism · moral epistemology · evolutionary debunking arguments

Table of Contents

Introduction	3
Part 1: The Metaethics and Legacy of Moore’s <i>Principia Ethica</i>	4
Part 1.1: Moore’s Metaethics in <i>Principia Ethica</i>	4
Part 1.2: Metaethics after <i>Principia Ethica</i>	6
Part 2: Street’s <i>Darwinian Dilemma</i>	8
Part 2.1: Preliminary Remarks on Street’s <i>Darwinian Dilemma</i>	8
Part 2.2: The Two Horns of Street’s <i>Darwinian Dilemma</i>	10
Part 3: Moore’s <i>Principia Ethica</i> facing Street’s <i>Darwinian Dilemma</i>	11
Part 3.1: Opting for the First Horn	11
Part 3.2: Opting for the Second Horn	14
Conclusion	16
Bibliography	18

Introduction

G. E. Moore's revolutionary 1903 book *Principia Ethica* (henceforth: *PE*) has proved to be among the most influential 20th century treatises on ethics. He devotes its first four chapters to metaethics, developing his *non-naturalism* in moral semantics and in moral metaphysics and, less comprehensively, his *intuitionism* in moral epistemology. – Part 1.1.

He hoped his discussion of the *non-analyticity of good*, the *autonomy-of-ethics thesis*, the *naturalistic fallacy*, and the *open question argument* would impel many philosophers to abandon naturalist moral realism in favour of non-naturalist moral realism. However, Moore also faced the *inadvertent* surge of metaethical alternatives developed by *non-cognitivists* who took *PE* to undermine a position so obvious to Moore he did not even argue for it in *PE: moral realism*. *Evolutionary debunkers* arguing that the genealogy of our moral beliefs undermines them are likewise part of *PE's* legacy. – Part 1.2.

Sharon Street's seminal 2006 paper *A Darwinian Dilemma for Realist Theories of Value* constitutes one of the most influential *evolutionary debunking arguments*. Her *Darwinian Dilemma* (henceforth: *DD*) arguably poses a major challenge for all forms of moral realism. To get her *DD* going, some preliminary remarks are in order. – Part 2.1.

Street's *DD* consists in moral realists having to opt for either of what she terms the first / second horn of her *DD*, both options seemingly being too epistemologically unattractive: Opting for the first horn, one *denies* any relation between a) evolutionary influences on our evaluative attitudes and b) the attitude-independent evaluative truths posited by moral realism. The problem here is that this seems to lead to moral scepticism. Opting for the second horn, one *affirms* a relation between a) and b). The problem here is that the *tracking account* moral realists would have to give seems epistemically inferior to an *adaptive link account* making no recourse to attitude-independent evaluative truths favoured by Street. – Part 2.2.

Constituting a central sub-question of the compatibility of Moorean and evolutionary metaethics, what I explore in this paper is how *PE's* metaethics could be defended against Street's *DD*. This being merely a term paper, plus *PE*, *DD* and – considering this is a *history of philosophy* term paper – their relation requiring quite a bit of unpacking, I limit my discussion to a few arguments by two defenders of moral non-naturalism.

Tomas Bogardus identifies in *DD's* first horn an *epistemic principle of nonaccidentality*. Of the three inferences in his reconstruction of *DD's* first horn, he argues the first to be limited in scope and the other two to be invalid. – Part 3.1.

William FitzPatrick objects to Street's claim that the adaptive link account is more parsimonious than the tracking account. He is also more sympathetic than Street is to *the byproduct hypothesis* and to a related *companions in guilt argument*. – Part 3.2.

I summarise rationalism in moral psychology and the byproduct hypothesis to be the two avenues for replying to Street discussed here I deem most promising but conjecture that Moorean metaethicists seem to face more challenges than other moral realists. – Conclusion.

Part 1: The Metaethics and Legacy of Moore's *Principia Ethica*

Part 1.1: Moore's Metaethics in *Principia Ethica*

G. E. Moore's moral non-naturalism is a combination of i) *moral realism* and ii) his *autonomy-of-ethics thesis*. He believed, like most ethicists of his time, i) that there are attitude-independent moral facts, even implicitly taking i) for granted in *PE*, focusing instead on ii): He was among the first to argue that the discipline of (normative) ethics is *sui generis*, i.e. independent of all other disciplines, such that issues in (normative) ethics cannot be settled with recourse to e.g. physics or metaphysics (cf. Hurka 2021, sec. 1).

It is Moore's conviction that how *good* is to be defined constitutes the most fundamental ethical question (cf. *PE*¹, 57). He contends (cf. *PE*, 61f.) that attempts to analyse this adjective (*good*) in other terms rest on confusion about the nature of *good* and commit what he terms the *naturalistic fallacy* (*PE*, 62):

It may be true that all things which are good are also something else [...]. And it is a fact, that Ethics aims at discovering what are those other properties belonging to all things which are good. But far too many philosophers have thought that when they named those other properties they were actually defining good; that these properties, in fact, were simply not 'other,' but absolutely and entirely the same with goodness.

PE's core argument for naturalists being confused is Moore's *open question argument* (henceforth: OQA), here illustrated by hedonism, "the doctrine that pleasure *alone* is good as an end" (*PE*, 114). My reconstruction, partly inspired by Nuccetelli and Seay 2022, 97:

¹ References to "*PE*" refer to the revised 1993 edition of Moore's *Principia Ethica* edited by Thomas Baldwin.

- (P1) To be good means to be [X=] pleasurable.
 (P2) "P is [X=] pleasurable but is P good?" is an open question.
 (P3) (P1) implies ¬(P2).
 (P4) The set of (P1), (P2), and (P3) is inconsistent and (P1) is the least plausible premise.
 (C1) ¬(P1)
 (P5) For any X, we can run the analogue of P1 through P4.
 (CX) ¬(PX)

(PX) states the analyticity of *good*, i.e. that a definition of the form "To be good means to be X." can be given. This is the premise he rejects but argues moral naturalists are committed to.

(P1) comprises (PX) plus hedonism as defined above.

(P2)'s question is *open* in the sense that the answer to it is not to be found in the concepts alone. "[...] whatever definition be offered, it may be always asked, with significance, of the complex so defined, whether it is itself good." (PE, 67). Contrast this with the definition "Brother means male sibling.", where, due to that definition's analyticity, it is a *closed* question whether a person's brother is that person's male sibling.

(P3): If (P1), then "P is pleasurable but is P good?" would be a closed question in the same sense and for the same reason as "B is your brother but is B your male sibling?".

(P4): Implicitly taken for granted by Moore.

(C1): Being good is not analysable in terms of being pleasurable. To Moore's knowledge, Sidgwick is the only ethical writer to precede Moore in realising that *good* is unanalysable (cf. 69). But "[w]hen Sidgwick noticed Bentham or Spencer equating goodness with a natural property such as pleasure, he thought it a minor slip that ought in charity to be ignored; Moore deemed it a decisive error that vitiated the philosopher's entire system." (Hurka 2021, sec. 5).

(P5) PE's chapters 2–4 are dedicated to examples of analogues of P1 through P4.

Moore's *moral intuitionism* is his proposal for how non-naturalists can have moral knowledge, here again illustrated by hedonism: Moore emphasises that, while he sees the *naturalist* hedonism defended by Jeremy Bentham and John Stuart Mill as defeated by his OQA, Henry Sidgwick's *non-naturalist* hedonism is immune to the OQA. He prizes Sidgwick (cf. PE, 111f.) for having identified the need for recourse to *intuition*: hedonism is to be defended as *self-evident* or not at all. In the preface to the first edition (PE, 34), Moore summarises that when confronted with questions about what is of intrinsic value

[...] no relevant evidence whatever can be adduced: from no other truth, except themselves alone, can it be inferred that they are either true or false. We can guard against error only by taking care, that, when we try to answer a question of this kind, we have before our minds that question only, and not some other or others; but that there is great danger of such errors of confusion I have tried to shew, and also what are the chief precautions by the use of which we may guard against them.

While Sidgwick's intuitions favour hedonism, Moore argues for his intuitions to the contrary using what he deems the only safe method here: *the method of isolation*. Moore invites us (cf. *PE*, 142–147) to contemplate a world where whatever quantity and quality of pleasure (or, subsequently, of consciousness of pleasure), “stripped of all its usual accompaniments” (*PE*, 142), were the only thing to exist, judging it self-evident that this is not the best possible world, hence rejecting hedonism.²

Sidgwick and Moore eschewed grounds for allegations of pertinent dogmatism by holding that their moral intuitionism is modest due to moral intuition being fallible (cf. Hurka 2021, sec. 2). Moore agrees with Sidgwick that no *proof* for our moral intuitions can be given (or sensibly demanded) but reckons our dissatisfaction about this stems chiefly from not the unprovability itself but our sometimes disagreeing with each other about even our reflected moral intuitions (cf. *PE*, 126f.).

Part 1.2: Metaethics after *Principia Ethica*

PE is undoubtedly among the very most influential books covering metaethics ever and Moore is probably the historical figure most often associated with moral non-naturalism and moral intuitionism (cf. Ridge 2019, sec. 0). “[H]e was influential in the switch of the general focus to metaethics and the consequent neglect of practical ethics that characterized the discipline until at least the 1970s.” (Nuccetelli 2022, 252). Thomas Baldwin concludes his introduction of *PE* thus: “No serious thinker could now agree with the totality of Moore’s position; but that is not the point – the ‘Preface to the Second Edition’^[3] shows how critical Moore himself had become of many aspects of his earlier position. It is, rather, Moore’s achievement to have provided the definitive starting-point for twentieth-century ethical theory.” (*PE*, xxxvii).

The legacy of *PE*, I figure, may stem in great part from *PE*’s iconoclasm: Moore laments the naturalistic fallacy “is to be met with in almost every book on Ethics; and yet it is not recognised” (*PE*, 65f.), arguing virtually all ethicists before him have been fundamentally in error due to virtually all of them being moral naturalists overlooking moral naturalism’s untenability. Of all arguments ever directed against moral naturalism, Moore’s OQA is “[b]y

² What matters here, i.e. in metaethics, contra normative ethics, is not the result but the method itself.

³ Unpublished until 1993, this is the preface Moore had written for a planned but never completed 2nd edition.

far the most famous and influential” (Lutz and Lenman 2021, sec. 2.1), though “his own enthusiasm about the OQA was short-lived” (Nuccetelli and Seay 2022, 93, cf. also 45).⁴

Non-naturalists’ typically intuitionist moral epistemology was likewise both very influential and met with considerable criticism. While naturalists can arguably rely on their general epistemological theory to explain how knowledge about natural (whether non-moral or moral) properties is possible (cf. Ridge 2019, sec. 3), non-naturalists face the *causal objection* that non-natural properties are causally inert, precluding knowledge about them (cf. Nuccetelli 2022, 148–150). Relatedly, non-naturalists typically resort to intuitionism, “[t]his pairing [being] part of Moore’s legacy” (Bedke 2009, 189). However, Bedke 2009 argues that this combination faces an epistemic defeater such that it would be a *cosmic coincidence* if our moral intuitions reliably coincided with the moral facts, giving us reason to abandon moral intuitionism or, more plausibly, he thinks, moral non-naturalism. Paralleling Bedke’s defence of moral intuitionism, Nuccetelli and Seay (2007, 13) find “a conviction about the proper methodology of ethics that a growing number of contemporary moral philosophers share with Moore: namely, that appeal to intuition at some level is after all unavoidable in moral reasoning, and is to that extent a legitimate move in ethics.”.

Yet *PE*’s moral epistemology might also be judged to be problematic due to *PE* containing arguably no satisfying account of how we could reliably resolve moral disagreements. Hence, Moore’s reliance on intuition for moral knowledge was taken by some who agreed with his autonomy-of-ethics thesis to undermine not only moral naturalism but moral realism and moral cognitivism in general: “it is not surprising that Moore’s ethical theory was regarded as undermining the cognitive status of morality, and thus that it led directly to the development of ethical non-cognitivism by those who were influenced by Moore, such as A. J. Ayer and C. L. Stevenson.” (Baldwin 2010, sec. 3).

Moore, with his OQA, was among those who harshly criticised Herbert Spencer’s Social Darwinism on theoretical grounds (cf. *PE*, 97–110). *PE* contributed to philosophers overwhelmingly abandoning recourse to evolution in ethics for much of the 20th century, indirectly partly paving the way for evolutionary ethics’ re-emergence in the form of evolutionary *metaethics* in the early 21st century (cf. Allhoff 2003, 109 and Nuccetelli 2022, 151–154 and 253).

Evolutionary debunking arguments (henceforth: EDAs) in metaethics are a family of moral scepticist lines of argument rising to prominence a century after *PE*’s publication. While EDAs differ widely in their scope and in the epistemic principles as well as empirical and metaethical assumptions they rely on⁵, I think there is a common denominator to be summarised as

⁴ For discussions of objections to Moore’s OQA(s) and to the sweeping conclusions he drew therefrom, cf. e.g. Ridge 2019, sec. 2, Hurka 2021, sec. 1, and Nuccetelli and Seay 2022, chs. 5+6.

⁵ For reviews of EDAs, cf. e.g. Wielenberg 2010, Vavova 2015, and FitzPatrick 2021.

follows: Evolution is argued to have sufficiently likely influenced at least some of our beliefs in a sufficiently pervasive, sufficiently epistemically dubious manner such that contemplating the genealogy of these beliefs should make us substantially lower our credence in them.

Some EDAs target not all but only some moral beliefs (e.g. deontological ones) and have been used to argue against theories in normative ethics (e.g. deontology). But what is of interest here are not such *selective EDAs* but only EDAs that have been used to argue against theories in *metaethics* by targeting *all* our moral beliefs (cf. FitzPatrick 2021, secs. 4.1+4.2). The extent to which any such *general EDA* gives us net reason to abandon belief in moral knowledge or at least moral realism in general or non- / naturalist moral realism in particular is a topic of contemporary academic debate.

Though there are dissenters (cf. e.g. Bogardus 2016), the majority view is that, to the extent EDAs cause trouble for naturalist realists, this is at least as true for non-naturalist realists. Naturalists can take this route (cf. Lutz and Lenman 2021, sec. 1.2.1): First, argue that “[t]he fundamental worry is that our moral beliefs are the product of evolutionary facts *rather than* moral facts.”. Second, emphasise naturalism entails that moral facts are natural facts. Third, argue that, considering naturalism might be true, it is a live option that the moral facts are not contrary to but among the evolutionary facts that explain our moral beliefs. It seems (cf. also Campbell 2019, sec. 4.1) non-naturalists have greater trouble accommodating this worry than naturalists: “[P]rominent non-naturalist moral realists regard debunking arguments as their most arduous test” (Klenk 2022, 43) and one could speculate whether among them would also be Moore.

Part 2: Street’s Darwinian Dilemma

Part 2.1: Preliminary Remarks on Street’s Darwinian Dilemma

Sharon Street notes she does not presume to have laid out in maximal detail a fully exhaustive, perfectly accurate account of *The Role* evolution played in the genealogy of all our evaluative attitudes. Most of her paper is an exploration of a conditional, viz. what it would mean in moral epistemological terms *if* her evolutionary picture coincides roughly with the most plausible scientific explanation of *The Role*. The evolutionary picture Street offers, the details of which she is sceptical of, but the outlines of which she is “certain enough to make it well worth exploring the philosophical implications” (*DD*⁶, 113), is what I now turn to.

There are numerous factors that could plausibly be hypothesised to have (had) an influence on our evaluative attitudes, among them social, cultural, and historical forces, rational reflection, and evolutionary forces. Street emphasises that her evolutionary picture is

⁶ References to “*DD*” refer to Sharon Street’s 2006 paper *A Darwinian Dilemma for Realist Theories of Value*.

perfectly compatible with all these factors having played a role in the genealogy of our evaluative attitudes but that “one enormous factor in shaping the content of human values has been the forces of natural selection, such that our system of evaluative judgements is thoroughly saturated with evolutionary influence” (DD, 114).

To show why this is so, she first lists various evaluative attitudes that serve as examples of attitudes where we have a convincing explanation of why they would be *overwhelmingly* (dis)favoured by evolutionary pressures (cf. DD, 114f.). Possibly the easiest case is the evaluative judgement that (1) something promoting one's survival is a reason in favour of it. Holding the opposite of (1) would clearly be overwhelmingly disfavoured by evolutionary pressures. Hence, we should expect (1) to be an extremely widespread evaluative judgement (among the living) – regardless of its truth-value and even its truth-aptness.

The same reasoning applies to the five further evaluative judgements Street listed, though she grants that the explanation of evaluative tendencies in their direction is more complex (cf. DD, 115f.), building upon the evolutionary theory of kin selection in these two cases:

- (2) something promoting the interests of a family member being a reason to do it and
- (3) our having greater obligations to help our own children than to help complete strangers ...and on the biological theory of reciprocal altruism in these three cases:
 - (4) someone having treated one well is a reason to treat that person well in return,
 - (5) someone being altruistic is a reason to admire, praise, and reward him or her, and
 - (6) someone having done one deliberate harm is a reason to shun that person or seek his or her punishment.

Street considers an interesting case of counterfactual evidence (cf. DD, 116f.): If it were the case that the opposites of (1) through (6) were among the most deeply and widely held beliefs of humans across spacetime – despite this so clearly decreasing their reproductive success – this would constitute powerful evidence against Street's evolutionary picture. However, given the pervasiveness of (1) through (6), what we see instead is for the most part exactly what we would expect if evolutionary pressures have played roughly the role in shaping our evaluative attitudes that Street claims they have.

Having considered complications to this rough evolutionary sketch, she concludes that evolutionary pressures have *indirectly* but nevertheless tremendously influenced our evaluative judgements, many of which we can view as conscious, reflective endorsements of what she calls our *basic evaluative tendencies*, the latter being shared with other animals and being what is *directly* influenced (cf. DD, 117–121).

Realist theories of value now face the choice (cf. DD, 109) of either denying or affirming a relation between a) evolutionary influences on our evaluative attitudes and b) the attitude-independent evaluative truths posited by moral realism (henceforth: realism's truths). What

makes Street title this a dilemma is her deeming both these options untenable, seemingly leading either to salvaging moral realism at the price of incurring a thoroughgoing moral scepticism, “a logically coherent position that contains about zero appeal” (Shafer-Landau 2012, 1), or relying on a seemingly scientifically untenable explanation. She refers to the two options of denying / affirming the above relation as the first / second horn of her *DD*.

Part 2.2: The Two Horns of Street’s *Darwinian Dilemma*

The first horn: If one denies that there is any relation between a) and b), one faces the challenge of explaining how, except by sheer coincidence, evolutionary pressures would just so have happened to influence our evaluative attitudes such that they perfectly or even just mostly align with realism’s truths. Given how astronomical the option-space of conceivable evaluative attitudes is, and how teeny a share of that would, given realism, be constituted by (mostly) true evaluative attitudes, it would seem astronomically more likely that evolutionary pressures arbitrarily happened to push us in realism’s true evaluative directions in only a teeny subset of instances. Hence, we should expect to have ended up, in virtually all instances, with evaluative attitudes not at all perfectly aligning with or even just closely resembling realism’s truths but to have instead been “evolving towards affirming whatever evaluative content tends to promote reproductive success” (*DD*, 122).

The second horn: If one affirms that there is a relation between a) and b), one faces the challenge of explaining what kind of relation this is. One could give a *tracking account* (cf. *DD*, 125f.) according to which it has been evolutionarily advantageous to arrive at realism’s truths, such that the evolutionary causes track the truth. Street argues (cf. *DD*, 134f.) that such a tracking account constitutes the only option for realists here (i.e. for the second horn) because they are committed to either accept that evolutionary pressures pushed us towards believing in realism’s truths (leaving one with a tracking account) or deny this. But denial would merely leave open whether evolutionary pressures pushed us away from realism’s truths, constituting an unappealing *distortionary account* “or pushed us in ways that *bear no relation to*” (*DD*, 135) realism’s truths, bringing us back to the first horn.

Street thinks both that the second horn is the “more plausible route for the realist to take” (*DD*, 125) but that the problem for realists is that there is another explanation, an *adaptive link account*, which she argues to be superior to the tracking account in at least three respects (cf. *DD*, 129–134). Street on the adaptive link account: “tendencies to make certain kinds of evaluative judgements rather than others contributed to our ancestors’ reproductive success [...] because they forged adaptive links between our ancestors’ circumstances and their responses to those circumstances, getting them to act, feel, and believe in ways that turned out to be reproductively advantageous.” (*DD*, 127).

First, the adaptive link account is *more parsimonious*: Whatever the tracking account explains with recourse to precisely what is at question, viz. the existence of realism's truths, the adaptive link account explains no less well without that (cf. *DD*, 129). Conversely, the adaptive link account requires no controversial assumptions not shared by the tracking account. Hence, it seems the tracking account is less parsimonious and moral realism explanatorily superfluous.

Second, the adaptive link account is *much clearer*: While the causal story is clear re how being able to grasp non-evaluative truths about fires and predators could have proved evolutionarily adaptive, it should prove a challenge to give a satisfactory answer to this when evaluative truths are at issue, especially for those who posit moral truths as causally inert, as non-naturalists seem committed to (cf. Ridge 2019, sec. 3). This leaves (especially non-naturalist) realists with explanations more obscure than the adaptive link account (cf. *DD*, 129–132).

Third, the adaptive link account is *more informative*: E.g., one of the questions that the tracking account fails to answer (cf. *DD*, 132f.) is why it is that the evaluative judgements (1) – (6) are deeply and widely held beliefs of humans across spacetime, while their opposites strike virtually all of us as immoral and evaluative judgements like “the fact that something is purple is a reason to scream at it” (*DD*, 133) as weirdly amoral. The adaptive link account explains that some links between circumstances and responses fostered by various evaluative tendencies prove evolutionarily advantageous for the individual, while other links prove disadvantageous or useless. In contrast, the tracking account can merely insist that some evaluative judgements are simply attitude-independently truer than others but is unable to give us a comparably informative explanation of why we have the ones we do (cf. *DD*, 133).

Part 3: Moore's *Principia Ethica* facing Street's Darwinian Dilemma

How can the Moorean metaethicist reply to Street's evolutionary debunking argument?

Part 3.1: Opting for the First Horn

One option consists in opting for the first horn of Street's *DD* and critically engaging with the claim that *it would undermine our claim to knowledge of realism's truths if it were a sheer coincidence that evolutionary pressures just so happened to influence our evaluative attitudes such that they perfectly or even just mostly align with realism's truths.*

Tomas Bogardus argues that an *epistemic principle of nonaccidentality* is driving Street's argument here (I agree and think she also would), the following reconstruction of the first horn being based on Bogardus 2016, 651–653:

(B1) Our moral faculty was naturally selected to produce adaptive moral beliefs, and not naturally selected to produce true moral beliefs.

Hence (B2): Our moral beliefs would not be true if our species had formed them using the same method we actually used but had evolved under circumstances where other moral beliefs proved (mal)adaptive than in ours.⁷

Hence (B3): It is accidental that we are right about our moral beliefs – or we are not right.⁸

Hence (B4): Our moral beliefs do not count as knowledge.

First, Bogardus intriguingly argues that moral psychology is to determine whether (B1) entails (B2) (cf. Bogardus 2016, 640–643). “If the process of moral judgment [always necessarily] relies on intermediary mental states (sentiments, gut reactions, ‘affect-laden intuitions,’ etc.) [*representationalism*], and the faculty that produces these intermediaries was selected for adaptiveness but not selected for truth” (Bogardus 2016, 647), (B1) entails (B2). But he sees *representationalism* as but one of three widespread views, the other two, *rationalism* and *divine revelation*, allowing for moral beliefs to be (also) formable *directly*, without relying on any intermediaries that evolution could have influenced. On these two accounts, the inference from (B1) to (B2) seems “restricted to moral beliefs based, fully or partly, on those mental intermediaries” (Bogardus 2016, 642), limiting the scope of Street’s argument.

I am very impressed by the creativity of this line of argument, but not *quite* as enthusiastic about it as he is. Even if granted that representationalism is to be rejected, my intuition⁹ is that most of our moral beliefs would still be within the target of this restricted scope due to relying at least partly on such intermediaries. However, I reckon Moorean metaethicists could try to wed their intuitionism in moral epistemology with rationalism in moral psychology in hope of allowing moral non-naturalists to mitigate or even dissolve Street’s *DD*.

Second, Bogardus argues (B2) does not entail (B3), relying on Peter Unger’s distinction between types of accidentality that are in- / compatible with knowledge. Unger and Bogardus argue accidentality is compatible with knowledge if the accidentality pertains merely to the existence of evidence or/and to the existence of the individual considering that evidence and incompatible only if, *given that individual considering that evidence*, the formation of a true belief is an accident (cf. Bogardus 2016, 651)¹⁰.

⁷ I take this to be relevantly synonymous with but clearer than Bogardus’ phrasing: “Therefore, it is false that: *had our species evolved elsewhere, elsewhere and we later formed moral beliefs using the same method we actually used, our beliefs would be true.*” (Bogardus 2016, 653).

⁸ For present purposes, I think it matters little whether to interpret (B3)’s first proposition as presupposing that all / most / some of us are always / often / sometimes right, so I will not make such distinctions here.

⁹ being far from an expert in moral psychology

¹⁰ referencing p. 159 of Unger, Peter. 1968. “An Analysis of Factual Knowledge.” in *Journal of Philosophy*. Vol. 65, No. 6. pp. 157–170

To illustrate, Bogardus invites us to imagine we are, in world A, the first person ever who Copernicus has shared his knowledge of heliocentrism with, such that we arrive at knowledge about heliocentrism via the method of expert testimony. In world B, we never met Copernicus but used the same method, such that another expert of that time made us not rightly believe but falsely disbelieve in heliocentrism (cf. Bogardus 2016, 654). Bogardus argues that, while it is an accident that we got Copernicus' testimony (instead of being led astray by another expert's testimony), our formation of a true belief about heliocentrism *based on his testimony* is not an accident. Transferring the analogy back: Even if we accept (B2), we can avoid accepting (B3) since it is conceivable that there are analogues of world A and that we have evolved in one of them, i.e. have evolved in a world where our evolutionary circumstances have been such that it is not an accident that the moral beliefs we have formed are true.

Here is why I find this lacking as a reply to Street: In world A, we have evidence (viz. for our belief in heliocentrism being true) that we lack in world B. Regarding evidence for some of our moral beliefs being attitude-independently true, however, it is unclear whether we would have it in *any* world. But to disarm Street's argument, it is not enough to claim that we *might be* in an analogue of world A but leave open how we could ever know that we *actually are* in such a world (if in fact we are in such a world) because the evolutionary debunkers of moral realism are open to agree with this much. What they could not agree with is that *if we are in such a world, we can know about our being in such a world*, but Bogardus' argument leaves open how if at all we could attain that knowledge.

Third, Bogardus argues (B3) does not entail (B4), referring to an argument based on a hypothetical device always reporting the time reliably but its' doing so being an accident: "John Hawthorne's swampwatch, a particle-for-particle duplicate of the world's most accurate wristwatch 'created by a fortuitous coming together of molecules.' Swampwatch reports the time but, given its birth from chaos, its reports are not aimed at the truth or anything else." (Bogardus 2016, 655)¹¹. Bogardus shares Hawthorne's intuition that believing this device's reports constitutes knowing, admitting even accidentality of the kind Unger argues to be incompatible with knowledge may after all be compensated by sufficiently high reliability (cf. Bogardus 2016, 655), concluding: "This may well go for our evolved moral faculty; for all the argument shows, our moral faculty is regularly, reliably successful, like a swampwatch." (Bogardus 2016, 655).

Contrary to Bogardus, I lean towards thinking (B3) does entail (B4). (B3) trivially entails (B4) if (B3)'s second proposition is true: if our moral beliefs are not true, they do not count as knowledge. Hence, if the inference from (B3) to (B4) is to be rejected, (B3)'s first proposition must be true. Bogardus' argument for (B3) not to entail (B4) rests on our moral faculty being regularly, reliably successful, *and thereby compensating* the accidentality such as to allow for

¹¹ referencing p. 252 of Hawthorne, John. 2002. "Deeply Contingent A Priori Knowledge." in *Philosophy and Phenomenological Research*. Vol. 65, No.2. pp. 247–269

at least some of our moral beliefs to count as knowledge. But I am afraid I fail to see where Bogardus established that *our moral faculty is regularly, reliably successful*, the truth-value of this ostensible fact naturally being a key issue in evolutionary metaethics. Hence, due to relying on a reliability claim that Street would be sceptical of, this does not constitute what to me seems like a conclusive argument against her position.

Part 3.2: Opting for the Second Horn

Another option consists in opting for the second horn of Street's *DD* and critically engaging with the claim that *the best explanation of the relation between a) and b) – evolutionary influences on our evaluative attitudes and realism's truths – moral realists can offer is worse than the adaptive link account favoured by Street*.

William J. FitzPatrick develops an interesting objection to Street's claim that the adaptive link account is more parsimonious than the tracking account. He argues that if we ask what the more parsimonious explanation is while implicitly already rejecting realism, we fail to see one crucial alternative: if realism is true, it is a live option that realism's truths figure into the best explanation for why we have (some of) the moral beliefs we have (cf. FitzPatrick 2015, 893).

On the one hand, I agree with FitzPatrick that if there is such an explanation, the adaptive link account might be better thought of as a misrepresentation than as more parsimonious, so I think that Street's parsimony verdict is premature, given metaethical uncertainty. On the other hand, I think at least moderate scepticism of there being such a better explanation, even conditional on some form of realism being true, is *prima facie* warranted, especially in the case of non-naturalism(,) which posits moral truths to be causally inert (cf. Ridge 2019, sec. 3). This connects back to Street's arguments that the adaptive link account is not just more parsimonious but also clearer and more informative than the tracking account. Hence, even if I might not grant her point re parsimony but only the other two, the adaptive link account seems to me the most plausible starting point against which alternative explanations should but maybe cannot measure up to.

We can try to make sense of the genealogy of our ability to do astrophysics, which we may assume we have not been directly selected for, on the grounds that this ability plausibly constitutes a combination of byproducts of capacities that were directly selected for. What if, so the *byproduct hypothesis*, (cf. *DD*, 142–144) an analogous explanation can be given for the ability to grasp realism's truths? Street argues defenders of the byproduct hypothesis face a dilemma analogous to her original *DD* of either denying or affirming the relation at issue, here viz. the relation between the ability to grasp realism's truths and *capacity C*, the latter being what the former is a byproduct of (cf. *DD*, 142).

Denying any such relation yields the verdict of an analogous sheer coincidence. Though, as Street rightly points out, if we correctly believe that sheer coincidence sometimes happens, the fact that an explanation involves sheer coincidence does not force us to reject that explanation. However, a combination of three assumptions makes her deem explanations involving *this particular* sheer coincidence implausible. First, “the more complicated and uniquely specialized a faculty is, the less plausible it is to hypothesize that it could have arisen as a sheer fluke, as the purely incidental byproduct of some unrelated capacity that was selected for on other grounds entirely” (DD, 143). Second, we should reject as too implausible the hypothesis that the human eye in its present form developed in such a way. Third, the ability to grasp realism’s truths constitutes a faculty at least as complicated and uniquely specialised as the human eye in its present form. I suppose the last assumption is the least robust, due to “uniquely specialised”, returning to this shortly.

Affirming such a relation poses the analogous challenge of explaining what kind of relation this is. Capacity C could have as byproduct the ability to grasp realism’s truths while directly facing selection pressure. This assumes capacity C is something related to but more rudimentary than that ability, “of which our present-day ability to grasp evaluative truths is a refined extension, in much the same way that our present-day ability to do astrophysics is presumably a refined extension of more basic abilities to discover and model the physical features of the world around us” (DD, 143f.). According to Street, denying the preceding about capacity C would lead us back to denying any such relation, so for her the question is how we got this more rudimentary ability. The problem she sees is that this again seems to require the realist to choose between sheer coincidence (albeit of a lesser magnitude) and the tracking account.

Street, concluding that the byproduct hypothesis, as an extension of the tracking account, is ultimately vulnerable to the same weaknesses, still favours the adaptive link account. FitzPatrick is more sympathetic to the byproduct hypothesis than Street is, arguing that our ability to arrive at truths in other philosophical and scientific pursuits can be explained as the byproduct of selection pressure on our general cognitive capacities (cf. FitzPatrick 2015, 886f.).¹² If the same extends to ethics, and we can develop our general cognitive capacities such that we can “mak[e] our cognitive dispositions in the relevant domain non-accidentally reliably truth-tracking” (FitzPatrick 2015, 886), then, contra Street, capacity C does not have to be a domain-specific ability. The crux naturally is whether the same does extend to ethics.

FitzPatrick affirms this, asserting two parallels: “At a formal level, we employ the same logical and analytic abilities in moral reasoning as in other forms of reasoning. And in terms of conceptual content, moral reflection and reasoning is continuous with broader evaluative and normative thinking that our cognitive capacities were plausibly designed to do accurately.”

¹² I cannot thoroughly assess the merits of this partly empirical claim here, but it seems plausible to me.

(FitzPatrick 2015, 888). He argues it was adaptive for our ancestors to make accurate evaluative and deontic judgements about e.g. what is good/bad to eat and who one ought (not) to cooperate with. Admitting that how exactly we got from those true judgements to knowledge of realism's truths, especially if non-naturalism is true, is in need of an explanation, he thinks "this falls far short of *debunking realism*" (FitzPatrick 2015, 889).

This line of argument might be the one I find most promising. What could be adduced here is a *companions in guilt argument* (henceforth: CGA) to the effect that if we reject the possibility of ethical knowledge because we think this would have to be *synthetic a priori knowledge*, it seems, we would, *unacceptably*, have to give up some deeply held beliefs in e.g. maths and metaphysics, too (cf. also Shafer-Landau 2012, 25–32). Yet I think this CGA works against evolutionary debunking of moral realism at the cost (for Moorean metaethicists) of parallelly working against *non-naturalist* realism, viz. I fear there is a tension in finding convincing both Moore's autonomy-of-ethics thesis and this CGA: if one holds normative ethics to be relevantly similar enough to maths and metaphysics to rely on this CGA, I deem it questionable to also hold normative ethics to be relevantly different enough for holding it to be *sui generis*.

Conclusion

I explored how the metaethics of G. E. Moore's revolutionary 1903 book *Principia Ethica* fare against the moral antirealist argument in Sharon Street's seminal 2006 paper *A Darwinian Dilemma for Realist Theories of Value*. To this end, I discussed a few objections to Street's evolutionary debunking argument by two contemporary philosophers I judged to subscribe to relevantly similar metaethics as Moore did. Among the avenues for escaping Street's moral antirealist conclusion I discussed, the most promising I deem these two:

First, with Tomas Bogardus, one could argue as follows: Even if we grant that our moral faculty was naturally selected to produce adaptive moral beliefs, and not naturally selected to produce true moral beliefs, this does not force us to accept the following inference Street makes: Our moral beliefs would not be true if our species had formed them using the same method we actually used but had evolved under circumstances where other moral beliefs proved (mal)adaptive than in ours. Given moral realism and a rationalist view in moral psychology, it is possible to form attitude-independently true moral beliefs directly, without relying on any mental intermediaries that evolution could have influenced. This allows for such directly formed moral beliefs to be attitude-independently true even in worlds where our species evolved under circumstances where other moral beliefs proved (mal)adaptive than in ours. Hence, that we are right about (some of) such directly formed moral beliefs is not accidental.

However, I think that uncertainty about the correct view in moral psychology should keep us from being very confident that all or even just most of our moral beliefs have been formed in such a direct manner. Hence, I fear if Moorean metaethicists want to not only mitigate but dissolve the first horn of Street's dilemma, they would have to argue that a pairing of their intuitionism in moral epistemology with rationalism in moral psychology can explain not for only some but for all of our moral beliefs a genealogy devoid of epistemically dubious evolutionary influences (on top of defending these two metaethical positions).

Second, with William J. FitzPatrick, one could argue as follows: We need not grant Street's claim that the ability to grasp attitude-independent moral truths would constitute a faculty at least as uniquely specialised as the human eye in its present form. Rather, that ability, can be understood, much like our ability to do highly sophisticated maths and metaphysics, to be a byproduct of selection pressure on our general cognitive capacities. Hence, moral realists need not grant that the best explanation for the relation between evolutionary influences on our evaluative attitudes and the attitude-independent evaluative truths posited by moral realism is the adaptive link account making no recourse to attitude-independent evaluative truths favoured by Street. Enhancing the tracking account realists are committed to with such a byproduct hypothesis allows for dissolving the second horn of Street's dilemma.

However, while I feel this avenue might be very promising for moral realists in general, I fear the Moorean metaethicists among them face an additional challenge. Assessing the merits of the byproduct hypothesis, the crux of the issue is, in my view, whether we accept that companions in guilt argument can sensibly be adduced here. If maths and metaphysics are relevantly similar to ethics, including re the formation of (true) beliefs in these disciplines, then our presumed unwillingness to abandon belief in some of our beliefs in maths and metaphysics being true secures the analogous belief re ethics. The challenge for Moorean metaethicists is how to explain away what I think is a tension resulting from holding ethics to be relevantly different enough to maths and metaphysics to subscribe to the autonomy-of-ethics thesis but holding ethics to be relevantly similar enough to maths and metaphysics to rely on such a companions in guilt argument.

Lest I run the risk of being accused of doing so, let me emphasise I by no means feel warranted in dismissing Moorean metaethics as clearly incompatible with evolutionary metaethics. Given the limits related to this being merely a term paper, I could only consider a small share of all evolutionary debunking arguments and replies to them. However, even among those anti-debunking arguments I studied but did not discuss here, none I found as convincing as Street's antirealist evolutionary debunking argument. My tentative conclusion is that Street's argument poses a very intriguing and enduring challenge for moral realists and even more so for Moorean metaethicists.¹³

¹³ I would like to thank Peter Schulte and Alice Soru for giving helpful feedback on an outline and a first draft.

Bibliography

- [Allhoff, Fritz. 2003. Evolutionary Ethics from Darwin to Moore. in *History and Philosophy of the Life Sciences*. Vol. 25, No. 1. pp. 51–79.](#)
- [Baldwin, Tom. 2010. “George Edward Moore”. in Edward N. Zalta, principal editor. *The Stanford Encyclopedia of Philosophy* \(Summer 2010 Edition\).](#)
- [Bedke, Mathew. 2009. Intuitive non-naturalism meets cosmic coincidence. in *Pacific Philosophical Quarterly*. Vol. 90, No. 2. pp. 188–209.](#)
- [Bogardus, Tomas. 2016. Only All Naturalists Should Worry About Only One Evolutionary Debunking Argument. in *Ethics*. Vol. 126, No. 3. pp. 636–661.](#)
- [Campbell, Richmond. 2019. “Moral Epistemology”. in Edward N. Zalta, principal editor. *The Stanford Encyclopedia of Philosophy* \(Winter 2019 Edition\).](#)
- [FitzPatrick, William. 2015. Debunking evolutionary debunking of ethical realism. in *Philosophical Studies*. Vol. 172, No. 4. pp. 883–904.](#)
- [FitzPatrick, William. 2021. “Morality and Evolutionary Biology”. in Edward N. Zalta, principal editor. *The Stanford Encyclopedia of Philosophy* \(Spring 2021 Edition\).](#)
- [Hawthorne, John. 2002. Deeply Contingent A Priori Knowledge. in *Philosophy and Phenomenological Research*. Vol. 65, No. 2. pp. 247–269.](#)
- [Hurka, Thomas. 2021. “Moore’s Moral Philosophy”. in Edward N. Zalta, principal editor. *The Stanford Encyclopedia of Philosophy* \(Summer 2021 Edition\).](#)
- [Klenk, Michael. 2022. Debunking, Epistemic Achievement, and Undermining Defeat. in *American Philosophical Quarterly*. Vol. 59, No. 1. pp. 43–60.](#)
- [Lutz, Matthew and James Lenman. 2021. “Moral Naturalism”. in Edward N. Zalta, principal editor. *The Stanford Encyclopedia of Philosophy* \(Spring 2021 Edition\).](#)
- Moore, G. E.. 1903. *Principia Ethica*. Cambridge: Cambridge University Press. revised edition with the preface to the second edition and other papers, edited and with an introduction by Thomas Baldwin. 1993. Cambridge: Cambridge University Press. [page references are to the revised 1993 edition]
- Nuccetelli, Susana and Gary Seay. 2007. “Introduction”. in Susana Nuccetelli and Gary Seay, editors. *Themes from G. E. Moore: New Essays in Epistemology and Ethics*. Oxford: Oxford University Press. pp. 1–21.
- Nuccetelli, Susana with Gary Seay. 2022. “*Principia Ethica* in its context”. ch. 2 in Susana Nuccetelli. *The Routledge Guidebook to Moore’s Principia Ethica*. New York and Oxfordshire: Routledge. pp. 25–51.
- Nuccetelli, Susana with Gary Seay. 2022. “*The Open Question Argument*”. ch. 5 in Susana Nuccetelli. *The Routledge Guidebook to Moore’s Principia Ethica*. New York and Oxfordshire: Routledge. pp. 91–115.

- Nuccetelli, Susana with Gary Seay. 2022. "The Naturalistic Fallacy". ch. 6 in Susana Nuccetelli. *The Routledge Guidebook to Moore's Principia Ethica*. New York and Oxfordshire: Routledge. pp. 116–136.
- Nuccetelli, Susana. 2022. "Moral knowledge". ch. 7 in Susana Nuccetelli. *The Routledge Guidebook to Moore's Principia Ethica*. New York and Oxfordshire: Routledge. pp. 137–159.
- Nuccetelli, Susana. 2022. "Epilogue: The legacy of G. E. Moore and *Principia Ethica*". in Susana Nuccetelli. *The Routledge Guidebook to Moore's Principia Ethica*. New York and Oxfordshire: Routledge. pp. 251–255.
- [Ridge, Michael. 2019. "Moral Non-Naturalism". in Edward N. Zalta, principal editor. *The Stanford Encyclopedia of Philosophy* \(Fall 2019 Edition\).](#)
- [Shafer-Landau, Russ. 2012. Evolutionary Debunking, Moral Realism and Moral Knowledge. in *Journal of Ethics and Social Philosophy*. Vol. 7, No. 1. pp. 1–38.](#)
- [Street, Sharon. 2006. A Darwinian Dilemma for Realist Theories of Value. in *Philosophical Studies*. Vol. 127, No. 1. pp. 109–166.](#)
- [Unger, Peter. 1968. An Analysis of Factual Knowledge. in *Journal of Philosophy*. Vol. 65, No. 6. pp. 157–170.](#)
- [Vavova, Katia. 2015. Evolutionary Debunking of Moral Realism. in *Philosophy Compass*. Vol. 10, No. 2. pp. 104–116.](#)
- [Wielenberg, Erik. 2010. On the Evolutionary Debunking of Morality. in *Ethics*. Vol. 120, No. 3. pp. 441–464.](#)