Routledge
Taylor & Francis Group

🔓 OPEN ACCESS Check for updates

Counterfactual genealogy and metaethics in Pettit's *The Birth of Ethics*

Tristram McPherson[a] and David Plunkett[b]

[a]Philosophy, The Ohio State University, Columbus, OH, USA; [b]Philosophy, Dartmouth College, Hanover, NH, USA

ABSTRACT
One of the primary goals of Pettit's *The Birth of Ethics* is to offer a novel defense of a form of naturalistic realism in metaethics, drawing on a kind of "counterfactual genealogy" for ethical thought and talk, in a community he dubs "Erewhon". We argue that Pettit's argument faces a deep dilemma. The dilemma begins by noting the reasonable controversy about which metaethical view is true of our ethical thought and talk. We then ask: is the thought and talk in Pettit's Erewhon apt for the same reasonable controversy? If so, this raises doubts about Pettit's case for naturalistic realism about Erewhonian "ethical" thought and talk. If not, this disanalogy between Erewhonian "ethical" thought and talk and our ethical thought and talk renders it difficult to argue smoothly from Erewhonian premises to conclusions about our own ethical thought and talk. We then consider an alternative use that someone might make of Pettit's discussion of Erewhon: as part of a conceptual ethics argument that we should use "ethical" concepts that are relevantly similarly to the ones described in Erewhon. We conclude by reflecting on the broader methodological significance of the sort of dilemma that we have posed.

Introduction

In *The Birth of Ethics*, Philip Pettit introduces a fictional human community he dubs "Erewhon", whose inhabitants start out as largely self-interested and amoral.[1] Pettit then puts forward a philosophically rich narrative that describes how these humans develop capacities for commitment, holding others responsible, and (seemingly) ethical thought and talk. Pettit draws on this narrative to advance views on a wide range of philosophical issues.

However, Pettit's central ambitions for his discussion concern the foundations of ethics. In particular, one of his primary goals is to offer a novel defense of a form of "naturalistic realism" in metaethics, according to which (put roughly) our ethical thought and talk refers to naturalistic properties, facts, and relations that are instantiated in the actual world.[2]

In this paper, we argue that Pettit's use of Erewhon to defend naturalistic metaethical realism faces a deep dilemma. In short, the dilemma begins when one observes that there is a great deal of reasonable controversy about which metaethical view is true of our ethical thought and talk. We can ask: is the relevant Erewhonian thought and talk apt for the same sort of reasonable controversy? If so, it is not clear that Pettit has a compelling case for naturalistic realism about Erewhonian "ethical" thought and talk. (We use "ethical" in scare quotes, when talking about the relevant parts of Erewhonian thought and talk, due to the fact that its status as *ethical* is a central locus of controversy.) However, if Erewhonian "ethical" thought and talk is not apt for the same sort of reasonable controversy that marks our ethical thought and talk, this looks like a striking disanalogy between Erewhonian thought and talk and ours. And this disanalogy makes it difficult to argue smoothly from Erewhonian premises to conclusions about our own ethical thought and talk (§2). After laying all of this out, we then consider an alternative way that someone might use Pettit's discussion of Erewhon: as part of a *conceptual ethics* argument that we should use "ethical" concepts that are relevantly similar to the ones described in Erewhon (§3). We conclude by reflecting on the broader methodological significance of the sort of dilemma that we have posed.

It is worth stating up front that we approach this discussion with significant sympathy for naturalistic realism in metaethics (even if we are far from certain it is correct). Our aim in this paper, however, isn't to assess the general merits of this kind of metaethical view. Rather, it is to assess the innovative kind of argument that Pettit gives in its favor in *The Birth of Ethics*, and in particular, the role that his appeal to Erewhon is meant to play in that argument.

1. Metaethics, counterfactual genealogy, and Erewhon: an overview

In order to set the stage for our argument in this paper, we need to have a working understanding of Pettit's argument in *The Birth of Ethics* that we

---

[2]In this paper, we use single quotation marks (e.g. 'bicycle') to mention linguistic items. We use double quotation marks (e.g. "bicycle") for a variety of tasks including quoting others' words, scare quotes, and mixes of use and mention. We use small caps (e.g. BICYCLE) to pick out concepts.

are focused on. Our aim in this section is to provide an overview of that argument. We begin by introducing Pettit's methodology in his book. We then locate and explain the metaethical view he aims to defend: namely, *naturalistic realism*. Finally, we briefly reconstruct the core strategy of Pettit's distinctive argument for that metaethical view in the book.

1.1. Methodological orientation

To get a sense of Pettit's methodology for defending naturalistic metaethical realism in *The Birth of Ethics*, consider the following quote. Pettit writes that his account "starts … from a naturalistic story about how recognizably ethical terms and concepts could have emerged among creatures of our ilk and could have played a referential, yet pre-scriptive role in registering bona fide properties of the world. And then it argues on that basis for a naturalistic realism about desirability and responsibility."[3] As this quote suggests, in *The Birth of Ethics*, Pettit does not seek to provide an account of how ethical thought and practice *in fact* developed among humans. Rather, the heart of his book is an account of how something *a lot like* our ethical thought and talk *would* have developed, in a clearly formulated set of possible circumstances (namely, those of Erewhon). This is an exercise in what Pettit aptly dubs "counterfactual genealogy".[4] As the quote further suggests, the aim of this counterfactual genealogy is not merely to illustrate philosophical theses, which are then to be defended on entirely independent grounds. Rather, Pettit argues that his counterfactual genealogy provides a novel basis for defending philosophical theses about our actual thought and talk, such as naturalistic realism in metaethics.

We can better appreciate the ambition of Pettit's argument by con-trasting it with Allan Gibbard's aims in his book *Thinking How to Live*.[5] A core part of that book consists of Gibbard providing what we might also think of as a kind of counterfactual reconstruction of "ethical" thought and talk. However, Gibbard does not move directly from this reconstruction to the conclusion that expressivism is true of our ethical thought and talk. Rather, Gibbard treats his reconstruction as providing what he calls a "possibility proof" of the coherence and intelligibility of the brand of metaethical expressivism he favors. If this "possibility

[3](Pettit 2018, 22).
[4](Pettit 2018, 5).
[5](Gibbard 2003).

proof" is on track, it may help to undercut various criticisms of expressivism, especially those tied to its (purported) impossibility or incoherence.

Contrast Gibbard's aims with Pettit's. For Pettit, the facts about Erewhon are not meant to merely be part of a kind of "possibility proof" for naturalistic metaethical realism, which could in turn help provide indirect support for the view (e.g. by undercutting certain criticisms of it). Rather, for him, those facts are meant to enable a direct argument in favor of the view.

The use of Erewhon for these metaethical ends is tied to Pettit's general advocacy, announced at the start of The Birth of Ethics, for the methodological importance of counterfactual genealogy. In the first chapter of The Birth of Ethics, Pettit claims that the "reconstructive" form of argument he deploys in his book (based on counterfactual genealogy) has important methodological advantages over other well-known defenses of philosophical views (such as naturalistic realism in metaethics). Pettit argues that his counterfactual genealogy approach allows him both to sidestep certain burdens that other kinds of well-known defenses of naturalistic realism might take on, and to better illuminate the relationship between our ethical concepts and our ethical practices.[6]

Pettit rightly notes that the use of certain forms of counterfactual genealogy runs throughout the history of philosophy, including, to take an example that Pettit gives pride of place, in H.L.A. Hart's account of the nature of law in The Concept of Law.[7] Yet, as the brief sketch above indicates, Pettit aims to use counterfactual genealogy in a way that is more ambitious than most uses of it in recent philosophy (including not only Gibbard's, but Hart's as well). Moreover, Pettit gives the approach a much more central role in his explicit discussions of philosophical methodology than others who have (more briefly) drawn on it. The Birth of Ethics thus provides an important and methodologically novel contribution to contemporary metaethical inquiry.

1.2. Naturalistic metaethical realism

Pettit's aim in the book is to defend naturalistic realism in metaethics. In this section, we briefly explain how we understand that view.

On the approach that we favor (which we have developed at length in previous work), "metaethics" can be understood in terms of an

---

[6](Pettit 2018, 24–28).
[7](Hart [1961] 2012). See (Pettit 2018, 5–6 and 52–54).

overarching explanatory project. That project is to explain how actual ethical thought and talk – and what (if anything) that thought and talk is distinctively about – fits into reality.[8] As we understand him, Pettit is not merely interested in defending naturalistic realism as a claim about the *Erewhonians'* (eventual) "ethical" thought and talk. Rather he aims to defend it as a claim about *our actual* ethical thought and talk, and the (purportedly naturalistic) properties this thought and talk is about.[9] That is, he aims to defend a metaethical conclusion, in our sense of "metaethics".

We take naturalistic metaethical realism, which Pettit aims to defend, to be centrally characterized by four metaethical commitments:

Cognitivism	Actual ethical thoughts, such as the thought that *it is wrong to eat meat*, consist in ordinary beliefs about ethical matters of fact.
Descriptivism	Actual grammatically indicative ethical sentences, like "it is wrong to eat meat", purport to state ethical facts.
Realism	There are ethical facts about the actual world.
Naturalism	The actual ethical facts are purely "naturalistic". That is, put roughly, they are "of a kind" with the sorts of facts revealed by the natural sciences.[10]

Pettit's particular, rich form of naturalistic metaethical realism involves further, more specific commitments. These include, for example, a commitment to a kind of anti-relativism or invariantism, according to which the truth-conditions of ethical sentences do not vary with the perspective of the speaker or evaluator.[11] While these further elements of Pettit's positive view are well worth reflecting on, in this paper we focus on the general form of naturalistic metaethical realism characterized by the labeled commitments above.

---

[8](McPherson and Plunkett 2017). See also (Plunkett and Shapiro 2017).
[9]In *The Birth of Ethics*, Pettit states that he will use the terms 'ethical' and 'moral' interchangeably. See (Pettit 2018, 13). Many philosophers want to draw a distinction between ethics and morality; e.g. by thinking of ethics as encompassing a wide range of questions about how to live and act, and morality as dealing with only a subset of those questions, such as what we owe to other persons, or those tied to the aptness of certain kinds of moral emotions. (See (Darwall 2017) for an overview of a range of views here). If one wanted to separate "ethics" from "morality" in some way, an interesting question would be whether Pettit's naturalistic realism would cover his views in both metaethics and metamoral inquiry. We take it that it would, at least on many of the dominant ways of drawing that distinction.
[10]Cf. (Pettit 2018, 17–18). For a helpful overview of "naturalistic realism" in metaethics in general, see (Railton 2017). For further discussion about how to best understand the commitments of this kind of metaethical view, see (McPherson 2015).
[11](Pettit 2018, 157 and 314–315).

1.3. From Erewhon to naturalistic metaethical realism: an overview

In this section, we briefly sketch the outlines of Pettit's counterfactual genealogy. We then clarify how we understand the general structure of Pettit's argument from that genealogy to the conclusion that naturalistic metaethical realism is true.

At the outset of Pettit's genealogy, Erewhonians are competent instrumental reasoners. They are largely motivated by self-interest, but are capable of relying on each other. This means, for example, that they do not face the problem of David Hume's imagined pair of farmers, whose crops spoil because they are incapable of cooperating with each other.[12] Further, they are capable of joint action, and of being able to use language to build on all of these capacities.[13] It is important for Pettit that these initial conditions do *not* involve the Erewhonians engaging in anything resembling ethical thought and talk.

Pettit then offers what he takes to be a naturalistically acceptable story of the development of Erewhon, which illuminates how the conditions outlined above "would almost certainly have given rise to ethical ways of thinking and acting".[14] That is, his aim is to show that – at least in the imagined circumstances – the "birth" of ethics was a near-inevitability. Pettit pursues this aim by, in short, ably tracing a compelling account whereby Erewhonians respond to clear incentives by incrementally developing increasingly sophisticated discursive and conceptual tools, culminating in practices organized around concepts of "ethical" desirability and responsibility. On Pettit's account, these concepts form the foundations of "ethical" thought and practice more generally among the Erewhonians.

The culminating chapter of Pettit's book seeks to draw metaethical conclusions from this counterfactual genealogy. In brief, Pettit first suggests that his narrative invites us to see the desirability and responsibility concepts that arise in Erewhon as "of a kind" with our own ethical concepts.[15] He then seeks to defend a form of naturalistic realism as a plausible account of the "ethical" thought and practice in Erewhon. Given the assumption that the Erewhonians' "ethical" concepts are of a kind with our own, this suggests a novel argument for naturalistic realism about *our* ethical concepts.

---

[12](Hume [1739] 2000, sections 3.2 and 5.8).
[13](Pettit 2018, 33).
[14](Pettit 2018, 31).
[15](Pettit 2018, 241).

In order to argue for naturalistic metaethical realism, one needs to explain why it is a more plausible hypothesis than its main competitors, which include expressivism, error theory, and non-naturalism in metaethics. One might thus expect Pettit to offer distinctive arguments against (say) metaethical expressivism in his book. In light of this, the first striking fact about the book's argument is that its explicit account of expressivism's alleged shortcomings does not appeal to Erewhon *at all*.[16] Instead, it adverts to familiar challenges to expressivism, including a broad form of the Frege-Geach argument, and a distinctive argument against expressivism developed previously by Pettit and Frank Jackson.[17]

As we understand Pettit's argument, Erewhon's role in rejecting expressivism (and other competitors to naturalistic realism) is encapsulated in the following passage:

> The theory supported by our reconstruction of morality makes it unnecessary to resort either to expressivism or to error theory. For on that theory the properties that become salient to people from within practices of avowing desire – in effect, robustly attractive desiderata – are excellent candidates for properties that judgments of desirability predicate.[18]

The idea here, we take it, is that because a successful realist semantics is so appealing as a theory of the Erewhonians' thought and talk, there is simply no motive for taking expressivism or error theory seriously as interpretations.

We can sum this idea up in a key thesis, which we will call Erewhonian Naturalism:

Erewhonian Naturalism   Naturalistic realism is highly plausible as an account of Erewhonian "ethical" concepts.

To get metaethical conclusions, we need to join Erewhonian Naturalism with what we will call:

The Linking Thesis   Erewhonian "ethical" concepts are relevantly similar to our ethical concepts.

As we read Pettit, he takes The Linking Thesis to be supported by the fact that his narrative enables us to see the Erewhonian "ethical" concepts as "of a kind" with our ethical concepts.[19] Notice that, to get metaethical

---

[16](Pettit 2018, 253).
[17]See (Jackson and Pettit 1998). For an overview of the Frege-Geach problem, see (Woods 2017).
[18](Pettit 2018, 254).
[19](Pettit 2018, 241). We interpret Pettit as using this consideration to support the Linking Thesis because we take it that he needs this thesis if he is going to get a metaethical conclusion from his strategy.

conclusions out of Erewhonian premises, we need to read "relevantly similar" here in a strong way. For example, relevant similarity must include similarity with respect to Cognitivism and with respect to the naturalistic character of the properties that figure in these concepts' extensions. Only this will ensure that, when we conjoin The Linking Thesis with Erewhonian Naturalism, we can conclude that naturalistic metaethical realism is true of our concepts. If Pettit could establish The Linking Thesis, then he would not need to develop novel arguments specifically targeting (for example) metaethical expressivism in order for Erewhon to have metaethical purport. Rather, his arguments for Erewhonian Naturalism and The Linking Thesis could together entail the plausibility of naturalistic metaethical realism, and hence the implausibility of metaethical expressivism.

2. A dilemma for Pettit's argument

In this section, we argue that Pettit's argumentative strategy faces a powerful dilemma. We lay out this dilemma in general terms, and then illustrate it with three examples that draw on influential arguments for competitors to naturalistic metaethical realism. We then argue that some of Pettit's central claims in his book would be strengthened by appealing to an explicit theory of content determination. But such an appeal would also threaten to turn the book's distinctive argument into an idle wheel, as we could apply that theory of content determination directly to our ethical concepts to help evaluate whether naturalistic realism is true of them.

2.1. A general statement of the dilemma

We begin by returning to contemporary metaethics. Philosophers do not accept competitors to naturalistic realism in a vacuum. Rather, competing metaethical views are each characteristically motivated by a collection of familiar, and prima facie powerful, arguments. In what follows, we argue that Pettit's appeal to Erewhon faces an important dilemma in light of these arguments. The basic recipe is this. First, consider the collected arguments for the competitors to naturalistic realism in metaethics. We can then ask: how compelling are precisely parallel arguments against Erewhonian Naturalism?

On the first horn of the dilemma, these arguments are just as prima facie potent for Erewhonian "ethical" concepts as they are for our ethical concepts. In this case, these arguments cast doubt on Erewhonian

Naturalism, because they support competitors to naturalistic realism about Erewhonian "ethical" concepts. On this horn, it is hard to see how the appeal to Erewhon makes any dialectical progress in favor of naturalistic metaethical realism.

On the second horn of the dilemma, we suppose that the relevant arguments are significantly less prima facie potent for the Erewhonian "ethical" concepts than they are for our actual ethical concepts. But, if that is true, it casts doubt on The Linking Thesis: it becomes much less clear that Erewhonians in fact share our ethical concepts. After all, if Erewhonian concepts are much less susceptible to such arguments than our ethical concepts are, this suggests that they may be quite different concepts. But if The Linking Thesis is false, then it is not clear how we could hope to draw conclusions about our ethical concepts from reflection on the Erewhonian "ethical" concepts.

2.2. Three illustrations of the dilemma

In this subsection, we illustrate the dilemma we have just described. We do this by reviewing three canonical metaethical arguments which have been used, respectively, to defend the most prominent competitors to naturalistic metaethical realism: expressivism, error theory, and non-naturalism.

First, consider expressivism.[20] For our purposes, we will understand metaethical expressivism simply, as rejecting Cognitivism and Descriptivism and replacing them with the following competing theses:

Non-Cognitivism	Ethical thoughts, such as the thought that *it is wrong to eat meat*, consist, at the most basic explanatory level, in desire-like, "non-cognitive" attitudes directed at non-ethical contents (in this case: *eating meat*), rather than belief-like "cognitive" ones.
Expressivism	Grammatically indicative ethical sentences, like "it is wrong to eat meat", purport to *express* the possession of a relevant non-cognitive attitude.

As with naturalistic realism (and the other metaethical views we discuss below), different theorists interested in defending a form of expressivism work with a variety of further commitments. So they develop the above schematic theses in different ways.[21] They also conjoin them with a

---

[20]For a helpful overview of expressivism in metaethics, see (Camp 2017).
[21]For example, for a more sophisticated variant on these simple theses above, see (Björnsson and McPherson 2014).

variety of further claims. For example: some contemporary expressivists advance "quasi-realist" forms of expressivism, which, put roughly, aim to show that expressivists can endorse traditionally "realist" sounding claims, such as that there are ethical facts, beliefs, and truths.[22]

One of the most prominent arguments for expressivism – abstracting from details about the exact form it takes – relies on the idea that our ethical concepts leave "open" core substantive ethical questions, in roughly the way that G. E. Moore highlighted in his (in-)famous "open question argument" (which he used to support non-naturalistic realism).[23] We can gloss this argument for expressivism as follows:

(i) For any naturalistic characterization (N) of the meaning of an ethical term like 'ethically desirable', the question "I see that this action is N but is it ethically desirable?" will seem open to competent speakers: that is, they will treat disagreement about whether N is ethically desirable as a substantive ethical dispute, rather than as evidence of conceptual incompetence.

(ii) This open-ness data is best explained by an expressivist account of our ethical thought and talk.[24]

The idea behind (ii) can be put roughly as follows. If expressivism is correct, then competence in using ethical terms involves knowing (perhaps implicitly) that they are used to express attitudes. So such competence will suggest that ethical disagreement quite generally is compatible with competence. By contrast, one might think that if descriptivism is correct, competence with ethical terms would consist at least partly in a grasp of the correct application conditions of those terms. And this might seem to make the "open question" phenomenon puzzling.

Now consider the Erewhonians. Pettit suggests identifying the property of ethical desirability with what he calls *multilateral desirability*.[25]

---

[22]For example, see (Blackburn 1993) and (Gibbard 2003). The "quasi-realist" program in expressivism raises some important complications for stating and evaluating the view. For discussion, see (Cuneo 2017) and (McPherson 2022).

[23](Moore [1903] 1993). See Gibbard's "what's at issue" argument in (Gibbard 2003), drawing on (Gibbard 1990). See also Terry Horgan's and Mark Timmon's Moral Twin Earth arguments for expressivism, as in (Horgan and Timmons 1993), which draw on this kind of "Moorean" idea of the "openness" of ethical thought and talk.

[24]For brief discussion of the history of how Moore's "open question argument" came to be seen as supporting expressivism, rather than non-naturalistic realism (of the kind Moore thought the argument supported), see (Darwall, Gibbard, and Railton 1997). See also (McPherson 2013).

[25](Pettit 2018, 183–196).

Very roughly, multilaterally desirable goods are ones that are generally desired, where each person has no objection to others also attaining those goods. For example, the desire to have love might be of this kind, since most people would not object to others also finding love. By contrast, if Romeo is jealous of Juliet being romantically involved with another person, he will not treat *Juliet's* love as multilaterally desirable.

Suppose an Erewhonian asks "I see that this action is multilaterally desirable, but is it *ethically* desirable?" We can then ask the second-order question: will an Erewhonian typically see that question as "open", in the relevant sense?

On the first horn of our dilemma, we answer "yes": the Erewhonians, like us, will find this question to be "open". But then, to the extent that the open question argument prima facie favors expressivism about our ethical thought and talk, it seemingly also does so regarding Erewhonian "ethical" thought and talk. On the second horn of our dilemma, we answer "no": the Erewhonians will typically find this question to be "closed". But if the argument for expressivism above (based on the "open question" feel of ethical judgment) is prima facie potent for our ethical concepts, this suggests strong evidence that the Erewhonians' concepts are in fact notably different from ours. This casts strong doubt on The Linking Thesis, and hence on any inferences that we might draw from Erewhonian concepts to our own. On either horn of the dilemma, it is hard to see how the appeal to Erewhon could either help defang the open question argument, or give the naturalistic realist a leg up in the debate with the expressivist about the nature of *our* ethical thought and talk.

Next consider the error theorist. The familiar sort of error theorist in metaethics embraces Cognitivism and Descriptivism. But they deny Realism: on the error theorist's view, our ethical thought and talk purport to be about ethical facts, but there are no such facts to talk about. On a familiar analogy, the error theorist sees ethical thought and talk in the way that a paradigmatic kind of atheist sees theistic thought and talk: as attempting to talk about a "reality" that simply fails to obtain.

Pettit gives even less direct criticism of error theory than he does of expressivism. Again, it appears that the general argumentative structure introduced in the previous section is supposed to suffice for rebutting this view. We can construct an instance of our dilemma for this strategy by focusing on the error theorist's most familiar form of argument. This argument begins by claiming that our ethical concepts enshrine a commitment to the "objective prescriptivity" or "irreducible normativity" of

ethical properties.[26] The error theorist goes on to argue that this commitment is incompatible with a naturalistic worldview.

We can pose our dilemma by asking: does the Erewhonians' central ethical concept, MULTILATERALLY DESIRABLE, encode commitment to "objective prescriptivity" or "irreducible normativity"?[27] On the first horn of our dilemma, we say yes. If so, the case for error theory seems as strong in Erewhon as it does for us, casting doubt on Erewhonian Naturalism. On the second horn, we say no: prima facie, the Erewhonian concept MULTILATERALLY DESIRABLE does not encode commitment to "objective prescriptivity" or "irreducible normativity". This horn suggests a striking asymmetry between the Erewhonians' "ethical" concepts and our ethical concepts, which casts doubt on The Linking Thesis: it suggests that our concepts and the Erewhonians' are relevantly different.

Finally, consider the non-naturalistic metaethical realist. The non-naturalist accepts Cognitivism, Descriptivism, and Realism, but rejects Naturalism concerning ethical facts and properties. Now consider one of the most influential contemporary arguments for non-naturalism. According to David Enoch, and many others, non-naturalism is motivated in significant part by the intuition that ethical facts and properties are "just too different" from prosaic naturalistic facts and properties for naturalism to be true. This intuition yields what we can call the "just too different" argument for non-naturalistic realism.[28]

Now consider the property of multilateral desirability that Pettit describes in his account of Erewhon. Does this property strike us as "just too different" from prosaic natural properties? Again we face a dilemma. On its first horn, we suppose that multilateral desirability does strike us as "just too different" from prosaic natural properties. Then there is a prima facie case for non-naturalism in Erewhon: Erewhonian Naturalism is cast into doubt. On the second horn, suppose (more plausibly) that multilateral desirability does not strike us as "just too different" from prosaic natural properties. Then, if Enoch's intuition is indeed plausible of the *actual* ethical facts and properties that we think about, this is prima facie evidence against The Linking Thesis. For it makes salient the worry that our ethical thought and talk is about very different sorts of properties than what Erewhonian "ethical" thought and talk is about.

---

[26]For example, see (Mackie 1977) (at least on the orthodox reading of him) and (Olson 2014).
[27](Miller 2021, 120) poses a similar question to Pettit.
[28]See (Enoch 2011). See (Paakkunainen 2018) for a survey of other appeals to "just too different" intuitions on behalf of non-naturalistic realism, and critical discussion of this form of argument.

We want to be clear about the following. One might credibly seek to argue that "open questions", "objective prescriptivity", or "just-too-different-ness" are not *really* characteristic properties of our ethical concepts, or what they are about. Or one might seek to argue that these properties are in fact compatible with naturalistic metaethical realism. We think both of these responses deserve to be taken seriously. But if we respond in one of these ways to arguments for expressivism, error theory, or non-naturalism, then Erewhon is doing no work in our argument: everything is being done by an argument that directly concerns *our* concepts.[29] For our purposes here, *that's* the key point. The point is not that any of the above arguments for metaethical views that are key rivals to naturalistic realism are especially good, let alone that those other views are correct. Rather, it is that Pettit's appeal to Erewhon does little to advance the debate in favor of naturalistic realism, given the existence of those arguments, which many of his interlocutors in metaethics take to be quite compelling.

We do not claim that these arguments show that Erewhon can play *no* role in an argument for naturalistic realism in metaethics. However, we think that the dilemma we have sketched shows that Pettit faces a strong burden to clarify exactly how something like The Linking Thesis can be substantiated without much more argument. More generally, we think this burden puts pressure on the idea that (purported) facts about Erewhon provide strong evidence directly in favor of any general metaethical view (e.g. naturalistic realism or non-naturalistic realism). That idea leaves open the possibility that a philosopher might appeal to facts about Erewhon to do something else in metaethical argument (e.g. provide a helpful illustration of a view, or play a role in a kind of "possibility proof" that could be used to undercut certain criticisms of a view). But it casts doubt on the more ambitious methodological goals Pettit has for counterfactual genealogy in metaethics.

2.3. Content determination and the argument from Erewhon

In light of the preceding argument, it is worth more closely examining the claims Pettit makes that can be interpreted as supporting The Linking Thesis. We argue that the force of these claims is undercut because they are not located within a clear theory of content determination,

---

[29]For brief discussion of a related idea, see (Phillips 2019).

which might serve to illuminate how they actually support The Linking Thesis.

We focus on two striking claims that Pettit makes:

> "[T]he concepts of desirability and responsibility that the narrative seeks to explain conform ... to the way we ordinarily understand desirability and responsibility."[30]

> "The concepts of moral desirability and responsibility that members of Erewhon would have come to develop are expressively equivalent to our concepts of moral desirability and responsibility, being subject to similar prompts and serving similar purposes."[31]

The first passage suggests that the Erewhonian concepts Pettit describes reflect our *ordinary understanding* of desirability and responsibility. But the three arguments briefly sketched in the previous subsection all cast doubt on this claim. Each purports to reflect explanatorily deep facts about our actual ethical concepts, or the properties they are about. And these arguments do not proceed via esoteric means: rather, they involve claims that are meant to resonate with our ordinary understanding of ethical concepts or properties. We think, then, that in claiming that Erewhonian concepts conform to our ordinary understandings, Pettit needs to engage with these arguments, in order to provide a clear account of what our ordinary notion of ethical desirability (e.g.) is.[32]

The second passage suggests that the Erewhonian concepts are *functionally* similar to our ethical concepts, in that they are elicited by "similar prompts" and serve "similar purposes". But in order for the second claim to be compelling, we would need a clear account of the purposes served by our *actual* ethical concepts, as used in our actual ethical practices. The problem, of course, is that the question of the purposes served by ethical thought and talk is intensely controversial. Some accounts foreground the alleged role of such thought and talk in enabling coordination that allows the involved parties to reap the benefits of cooperation.[33] Pettit's account of multilateral desirability seems to be motivated by two purposes:

---

[30](Pettit 2018, 241).
[31](Pettit 2018, 241–242).
[32]The same is true of other framing assumptions. For example, Pettit introduces the assumption that contextualism and relativism are not true of the semantics of 'morally desirable'. (See (Pettit 2018, 157)). However, some recent experimental work suggests that there might be considerable folk sympathy for these views. (See, for example, (Pölzler and Wright 2019) and (Beebe 2022)).
[33]See, for example, (Gauthier 1986).

- Providing a standard that allows an agent to adjudicate among the various standards by which they might assess an action as desirable.

- Providing a *common* evaluative perspective that all can coordinate around.[34]

One might worry whether these two alleged purposes are in fact co-achievable. But we set that aside, to focus on two connected questions.

First, how strong is Pettit's case that these two purposes are central to the function of *our* ethical concepts? Pettit provides a page of discussion that suggests that features related to these two purposes are "marks" of ethical desirability.[35] But even if this is so, it falls well short of establishing that these marks are *functionally central* to our ethical thought and talk.

Second, suppose that Erewhonian "ethical" concepts served similar functions as our ethical concepts. The clearest way to move from this supposition to The Linking Thesis would be to defend a theory of content determination according to which certain kinds of functional roles of a fragment of thought and talk help to fix (e.g.) what sorts of properties, if any, that fragment picks out. But Pettit does not appeal to a theory of content determination in his argument.

We further illustrate the dangers of failing to systematically engage with theories of content determination by considering a key move in Pettit's case against non-naturalism. Pettit says that "the fact that the residents of Erewhon could have come to master ["ethical"] terms in the absence of non-naturalistic entities implies that they do not posit or presuppose anything of that kind".[36] We will grant that Pettit provides a fully naturalistic description of the practices in Erewhon. But the inference stated here strikes us as flawed. To see why, consider theistic discourse. A good naturalist will insist that we can provide a wholly naturalistic account of theistic discourse. But this position does not entail naturalism about the gods that theists believe exist. The sensible naturalist should accept that this discourse, despite being explicable in wholly naturalistic terms, has *supernatural purport*. This parallel case shows that we cannot do what Pettit seems to want to do in the quoted passage, which is to infer naturalistic *purport* for a discursive practice from the fact of a naturalistic *description* of that discursive practice.[37]

---

[34](Pettit 2018, 180ff).
[35](Pettit 2018, 151–152).
[36](Pettit 2018, 244).
[37](Cuneo 2020, 478–479) makes a related point in more detail. See also (Phillips 2019) for related discussion.

The example of theistic discourse demonstrates that the fact there is a naturalistic account of the working of a discourse does not, by itself, settle what the metaphysical purport of that discourse is. This means that someone who wishes to defend naturalistic realism needs to be able to distinguish the case of ethical discourse from the case of theistic discourse, in terms of the respective ontological purports of these discourses. This point is not merely pedantic. Many ethical non-naturalists insist that ethical properties are causally inert.[38] As such, these non-naturalists are typically *committed* to there being a fully naturalistic causal explanation of ethical discourse that doesn't involve non-naturalistic ethical properties. So pointing out that such an explanation is available does little to shift the dialectic.

Again, the most principled way of defending a contrast between the naturalistic purport of ethical discourse and the supernatural purport of theistic discourse would involve a clear theory of content determination. In other work, Pettit has defended such a theory of content-determination.[39] But if that theory of content determination is plausible, it can serve to directly support the plausibility of naturalistic realism in metaethics, as Pettit has argued in other work. Given this, once we have such a theory of content determination in place, it again becomes unclear what additional work the appeal to the counterfactual case of Erewhon is doing for advancing Pettit's metaethical views.

2.4. Stepping back

In this section, we have argued that Pettit's appeal to Erewhon to support naturalistic metaethical realism faces a deep dilemma. We do not claim that the dilemma is decisive. Rather, we think it reveals a crucial gap in Pettit's argument. Here, we consider two ways that Pettit might try to address the dilemma.

First, Pettit might reply that the similarity between Erewhonian "ethical" concepts and our own is a matter of *degree*, and that the relevant degree of similarity is enough to get him the metaethical results he wants. Perhaps Pettit can avoid the dilemma by finding the right balance here: a balance on which our ethical concepts are similar enough to those used in Erewhon to support The Linking Thesis, but not so similar that competitor views to naturalistic realism are as prima

---

[38]See, for example, (Moore [1903] 1993) and (Enoch 2011).
[39](Jackson and Pettit 1995).

facie compelling for accounts of Erewhonian concepts as they are for our concepts.

We think that there is potentially much to be learned by thinking about degrees of similarity (along different dimensions) between our concepts and those used by the Erewhonians, and that this line of thinking is worth exploring further. But we are unconvinced that this strategy will do much to save Pettit's core argument for naturalistic metaethical realism in *The Birth of Ethics*. One reason why is that proponents of the three sorts of arguments we discussed in §2.2 tend to think that these arguments reveal crucial and deep facts about our ethical concepts, or the properties they are about. Given this tendency, there is a clear danger that discussion of a concept that is not vulnerable to these arguments may seem like it simply misses the point. Further, we suspect that the "horns" of the dilemma we have posed are gradable. That is, the more similar overall our ethical concepts are to Erewhonian "ethical" concepts, the more Pettit will end up on the first horn of the dilemma. And the more dissimilar overall these concepts are, the more he will up on the second horn.

A second way that Pettit might try to address our dilemma focuses on the second horn. Suppose that we notice that certain popular metaethical arguments (such as those discussed in §2.2) are not forceful when applied to Erewhonian "ethical" concepts. Pettit might then want to suggest that our reasonable confidence in the relevant similarity of our ethical concepts to the Erewhonian ones should lead us to doubt the probative force of the popular metaethical arguments when they are applied to our ethical concepts.

In response, we want to grant that something like this can happen. At times, we can recognize, in an elegant account, the explanatory heart of some part of our own, inevitably messy, practices. But we are unconvinced that this kind of line can work in defense of Pettit's use of counterfactual genealogy for defending naturalistic realism in metaethics.

Our pessimism about this response connects to a collection of concerns about Pettit's methodology that we have thus far left in the background. For example, one question about Pettit's counterfactual genealogy is this: why select these precise assumptions about Erewhon's starting conditions? In other words, why are *these* starting conditions, and not others, fit to illuminate our own ethical practices?[40] One example of

---

[40](Miller 2021, 120) raises a related concern.

why this kind of question is salient is that, in some parts of the social con-tract tradition, the central point of ethical norms is to enable rational crea-tures to solve the sorts of collective action problems that Pettit assumes are already solved at the (pre-"ethical") outset in Erewhon.

Relatedly, we can ask: how sensitive are the outcomes that Pettit traces in his account of Erewhon to the precise details of his assumptions about the starting conditions? If we imagined another, slightly different counter-factual community, would we get convergence on ethical practices like ours, a quite different practice, or no determinate counterfactual truth? The crucial question, where we do get variation between counterfactual communities, is this: what is methodologically special about the starting assumptions in Erewhon in particular, as opposed to those in some other imagined community? (Note that since Erewhon is not supposed to be an approximate *actual* genealogy of human ethics, we cannot always reply: "because the assumptions in Erewhon are more realistic".)

Another related concern is this: given that the "ethical" practices in Erewhon evolve over the course of the story that Pettit sketches (and pre-sumably would continue to evolve after that story ends), which stages of those practices are the ones that are philosophically relevant to drawing metaethical conclusions? For example, what should we say if one stage seemed to evolve forms of "ethical" thought and talk that was well-characterized by an expressivist analysis, but then those forms evolved into a kind of thought and talk that is decidedly not expressivist?

As the responses we have considered in this section indicate, we do not take the dilemma we pose to be decisive. However, we also do not see an easy way for Pettit to respond to it.

3. A Conceptual ethics reading of Pettit's use of counterfactual genealogy?

If our argument thus far is correct, it suggests that Pettit's use of a coun-terfactual genealogy about Erewhon does little to advance the overall case for naturalistic realism in metaethics. We now consider a possible alternative use of the counterfactual genealogy that Pettit (or someone on his behalf) might want to explore in response.

To see the alternative, it will be useful to distinguish two projects in the foundations of ethics: *metaethics* and *the conceptual ethics of ethics*.[41] As

---

[41](McPherson and Plunkett 2021b). For connected discussion (in the case of epistemology), see (McPher-son and Plunkett 2021a).

we have described it above, metaethics is a fundamentally descriptive or interpretive project. By contrast, the *conceptual ethics of ethics* encompasses certain kinds of normative and evaluative inquiry about ethical thought and talk.[42] It includes inquiry into which broadly "ethical" concepts we should use, and why, and parallel questions about words. It also includes questions such as the following: are our current ethical concepts defective in some way? Could they be improved? If so, how? And what kinds of standards should we be using to evaluate these concepts, and why?

The term 'ethics' in "conceptual ethics" should be understood *very* broadly, as picking out normative and evaluative inquiry, including such inquiry that appeals to norms and values that are not standardly used in moral and political philosophy. Some accounts in what we call "conceptual ethics" centrally appeal to moral or political norms.[43] But others instead emphasize broadly "metaphysical" norms, such as "carving reality at its joints", or broadly "epistemological" norms, such as fostering smooth inquiry.[44] And 'conceptual' here is meant to give a nod to the idea that many instances of what we call "conceptual ethics" purport to involve the normative assessment of concepts. But this isn't meant to imply that conceptual ethics can only concern the evaluation of *concepts*. (For example, some working in what we take to be "conceptual ethics" are focused on questions about words and their semantic content, and eschew talk of concepts.)[45]

With the idea of the conceptual ethics of ethics in mind, now return to Pettit's arguments in *The Birth of Ethics*. As we have reconstructed his central ambitions in this book with respect to ethical thought, talk, and reality, he is ultimately concerned with illuminating *actual* ethical thought, talk, and reality. As such, his naturalistic realism, as we have understood it, needs to correctly describe our actual ethical thought and talk, and what (if anything) it is distinctively about. It is this aim that generates the dilemma that we have posed for his argument.

But suppose that Pettit changed his aim with respect to ethical thought and talk, from something descriptive to something

---

[42]See (McPherson and Plunkett 2020) and (McPherson and Plunkett 2021b), drawing on (Burgess and Plunkett 2013a) and (Burgess and Plunkett 2013b).

[43]See, for example, (Haslanger 2000).

[44]For example, see (Sider 2011) for emphasis on broadly "metaphysical" norms, and (Pérez Carballo 2020) for emphasis on broadly "epistemological" ones. For further discussion of the varied kinds of norms and values that work in (or about) conceptual ethics appeals to, see (Burgess and Plunkett 2013b) and (Cappelen and Plunkett 2020).

[45]See (Cappelen 2018).

normative.[46] More specifically, suppose that Pettit instead proposed to use his sketch of Erewhon as part of an argument for characterizing the sort of broadly "ethical" concepts that we *should* be using (whether or not we are actually currently doing so). In other words, suppose that Pettit shifted his project from one fundamentally in metaethics to one in the conceptual ethics of normativity.

This possible way of developing Pettit's argument connects with some of his own work on philosophical methodology, published after *The Birth of Ethics*. In his paper "Analyzing Concepts and Allocating Referents", Pettit puts forward an illuminating reading of his influential defense of a "republican" view in political philosophy.[47] Pettit suggests that the idea of "freedom" that is central to this view – a view on which, put roughly, freedom involves a kind of "non-domination" – should be understood as involving what he calls "philosophical analysis". As he presents it, "philosophical analysis" involves a mixture of descriptive and normative considerations. The rough idea is that a purely descriptive account of key concepts (such as the concept FREEDOM) leaves open important questions about which of a range of referents a given concept actually applies to. Given this, Pettit suggests that given the kinds of interests we have when doing political philosophy, the question of which referent the concept has should be answered in part by normative considerations. The overall resulting theory should be judged on the basis of a kind of overall "reflective equilibrium", which tries to balance the different commitments we have about (for example) what freedom is with the normative work we think it should do in our theories and practices.

As we read it, Pettit's "philosophical analysis" can be understood to partly involve what we take to be "conceptual ethics" issues, about which concepts we should use, and which ones are better or worse. Or, perhaps more accurately, depending on how Pettit might want to individuate concepts here, we might say it involves issues about which *versions* of a concept we should use, and which ones are better or worse. In "Analyzing Concepts and Allocating Referents", Pettit does not develop a reading of his core metaethical views (e.g. his commitment to naturalistic realism) in a way that parallels this reading of his work on republicanism in political philosophy. But his interest in this kind of reading of some of

---

[46]Manuel Vargas, whose own work involves significant amounts of what we call "conceptual ethics", such as in (Vargas 2013), also briefly floats something along these lines as a possible way of developing Pettit's use of counterfactual genealogy. See (Vargas 2023).

[47]See (Pettit 2020), discussing the "republican" view in political philosophy he develops in work such as (Pettit 1999) and (Pettit 2012).

his most well-known work in political philosophy suggests that he might well be open to such a way of developing his metaethical views.[48]

We can now ask: if we consider a "conceptual ethics" variant of Pettit's strategy in The Birth of Ethics, in which the target is ultimately a claim about which "ethical" concepts we should use, how does this consideration affect the arguments we have advanced against Pettit in this paper?

To begin, consider again The Linking Thesis, according to which, recall, Erewhonian "ethical" concepts are relevantly similar to our ethical concepts. If, by hypothesis, Pettit's aim in appealing to Erewhon is not to establish a descriptive thesis about the way our actual ethical concepts are, but rather a normative claim about which ethical (or "ethical-ish") concepts we should use, then he no longer needs to defend The Linking Thesis as we have stated it.[49] Instead, he would need to establish something along the following lines: Erewhonian ethical concepts are good in certain respects, and it would also be good if our ethical concepts were similar in these respects. One way in which they might be good, for example, is by picking out naturalistic properties, rather than non-naturalistic ones (and thereby, perhaps, allowing "ethical" inquiry to be more epistemically tractable).

The dilemma argument that we have proposed does not directly apply to this sort of view, since it precisely targets The Linking Thesis. But one might think that a version of the underlying thought guiding the dilemma still applies: if we are interested in asking normative and evaluative questions about our ethical concepts, what motivates the indirect strategy that centrally involves counterfactual genealogy?

Although this is a good challenge, we think there is room for optimism about an answer here. For the counterfactual genealogy might help us to see how certain sorts of "desirability" and "responsibility" concepts and practices distinctively function, in a way that allows us to see what is valuable about concepts and practices with those functions. And it might also allow us to see that naturalistic realism is (or at least can be) true of the concepts that have these valuable functions. The fact that the genealogy is unrealistic does suggest an important gap: why should we be confident that the attractive function we get in these (in certain ways) unrealistic

---

[48]For connected discussion, see (McPherson and Plunkett 2021b).

[49]By "ethical-ish" concepts, we mean (roughly) concepts that play relevantly similar inferential, communicative, and representational roles as our ethical concepts. Where the line is between properly "ethical" concepts and just "ethical-ish" concepts is a further question, which interacts with general questions about the nature of concepts.

scenarios will also be served by similar concepts in our actual world? But this gap seems to us at least potentially bridgeable. If that is right, then Pettit's characterization of Erewhon might then form the basis for exploration of whether we too could and should use such concepts in our lives in the actual world.

We don't want to endorse this "conceptual ethics" way of developing Pettit's naturalistic realism, or his appeal to Erewhon as part of an argument on its behalf. But we hope to have sketched enough of what this approach might look like to make clear that there are interesting avenues worth exploring at further length, even if they involve a significant transformation of the methodology that appears to guide Pettit's use of his brand of counterfactual genealogy in *The Birth of Ethics*.

Conclusion

In this paper, we've considered Pettit's use of counterfactual genealogy to defend a form of naturalistic realism in metaethics. We've argued that while counterfactual genealogy might be philosophically useful in any number of ways, we should be skeptical that it can have the significant metaethical bite that Pettit thinks it can. We have then suggested that counterfactual genealogy has more promise as a method for defending conceptual ethics conclusions.

Perhaps the central methodological question that *The Birth of Ethics* makes salient is whether, and under what conditions, a counterfactual genealogy can contribute to our understanding of our actual concepts and practices, whether questions about those concepts and practices are being raised in metaethics or elsewhere. One way to approach this question is to address it head-on, in entirely general terms. We think there is much to be gained by doing so.[50] However, it is also possible to draw lessons for how to think about this general methodological question by abstracting from the specific argument of this paper.

The general question is important in part because Pettit draws on his story about Erewhon to advance views on a range of philosophical topics in *The Birth of Ethics*, from ones in moral philosophy to epistemology to philosophy of language. And others might seek to draw inspiration from his pathbreaking work in addressing yet further topics.

---

[50]As we think is evidenced by Vargas's discussion in his piece for this symposium. See (Vargas 2023). For connected discussion, see (Queloz 2021).

We take our argument in this paper to suggest a recipe for challenging the use of counterfactual genealogy to draw conclusions about our actual concepts and practices. First, for any given topic (e.g. the nature of moral responsibility, the nature of assertion, etc.) we expect that an argument from counterfactual genealogy will typically need to defend two sorts of claims:

1. A claim about the nature of relevant concepts or practices in the counterfactual scenario (in our case: Erewhonian Naturalism).

2. A claim about the relevant similarity between us and Erewhon (in our case: The Linking Thesis).

Insofar as claims analogous to those made in (1) are controversial concerning *our* concepts and practices, we will be in a position to construct a dilemma analogous to the one we have developed in this paper.

It might well be that such a dilemma has more traction for certain applications of counterfactual genealogy than others. But, given the general nature of the kinds of considerations that make the dilemma forceful for the metaethics cases, there is also good reason to suspect that it might well have bite for some of the further applications of counterfactual genealogy that Pettit presents in *The Birth of Ethics*.

In closing, we want to emphasize that we do not take our brief assessment of Pettit's use of counterfactual genealogy to be conclusive. Rather, we aim for our contribution to be opening moves in what will hopefully be a longer conversation about counterfactual genealogy, especially in metaethics and the conceptual ethics of ethics. As we have emphasized, Pettit's ambitious use of counterfactual genealogy is a novel move in recent metaethics – and, indeed, in recent philosophy more generally. Whenever philosophers stake out new methodological terrain – especially when done in the sort of bold way that Pettit does in *The Birth of Ethics* – they will inevitably face important challenges and hurdles. Our hope is that our contribution to the evaluation of Pettit's methodological views provides him (and others) with space to refine and further develop the potential utility of counterfactual genealogy for philosophy.

Disclosure statement

No potential conflict of interest was reported by the author(s).

References

Beebe, James R. 2022. "The Empirical Case for Folk Indexical Moral Relativism." In *Oxford Studies in Experimental Philosophy*, Vol. 4, edited by T. Lombrozo, J. Knobe, and S. Nichols, 81–111. Oxford: Oxford University Press.

Björnsson, Gunnar, and Tristram McPherson. 2014. "Moral Attitudes for Non-Cognitivists: Solving the Specification Problem." *Mind* 123 (489): 1–38. https://doi.org/10.1093/mind/fzu031.

Blackburn, Simon. 1993. *Essays in Quasi-Realism*. Oxford: Oxford University Press.

Burgess, Alexis, and David Plunkett. 2013a. "Conceptual Ethics I." *Philosophy Compass* 8 (12): 1091–1101. https://doi.org/10.1111/phc3.12086.

Burgess, Alexis, and David Plunkett. 2013b. "Conceptual Ethics II." *Philosophy Compass* 8 (12): 1102–1110. https://doi.org/10.1111/phc3.12085.

Camp, Elisabeth. 2017. "Metaethical Expressivism." In *The Routledge Handbook of Metaethics*, edited by T. McPherson and D. Plunkett, 87–101. New York: Routledge.

Cappelen, Herman. 2018. *Fixing Language: An Essay on Conceptual Engineering*. Oxford: Oxford University Press.

Cappelen, Herman, and David Plunkett. 2020. "A Guided Tour of Conceptual Engineering and Conceptual Ethics." In *Conceptual Engineering and Conceptual Ethics*, edited by A. Burgess, H. Cappelen, and D. Plunkett, 1–26. Oxford: Oxford University Press.

Cuneo, Terence. 2017. "Quasi-Realism." In *The Routledge Handbook of Metaethics*, edited by T. McPherson and D. Plunkett, 626–642. New York: Routledge.

Cuneo, Terence. 2020. "Morality, Money, and Method: Pettit's The Birth of Ethics." *Analysis* 80 (3): 575–583. https://doi.org/10.1093/analys/anaa029.

Darwall, Stephen. 2017. "Ethics and Morality." In *The Routledge Handbook of Metaethics*, edited by T. McPherson and D. Plunkett, 552–556. New York: Routledge.

Darwall, Stephen L., Allan Gibbard, and Peter Albert Railton. 1997. "Toward Fin de siecle Ethics: Some Trends." In *Moral Discourse and Practice: Some Philosophical Approaches*, edited by S. L. Darwall, A. Gibbard, and P. A. Railton, 3–50. New York: Oxford University Press.

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

Gauthier, David. 1986. *Morals by Agreement*. Oxford: Oxford University Press.

Gibbard, Allan. 1990. *Wise Choices, Apt Feelings: A Theory of Normative Judgment*. Cambridge: Harvard University Press.

Gibbard, Allan. 2003. *Thinking How to Live*. Cambridge: Harvard University Press.

Hart, H. L. A. (1961) 2012. *The Concept of Law*. 3rd ed. Oxford: Oxford University Press.

Haslanger, Sally. 2000. "Gender and Race: (What) Are They? (What) Do We Want Them to Be?" *Noûs* 34 (1): 31–55. https://doi.org/10.1111/0029-4624.00201.

Horgan, Terry, and Mark Timmons. 1993. "New Wave Moral Realism Meets Moral Twin Earth." In *Rationality, Morality, and Self-Interest*, edited by J. Heil, 115–133. Lanham, MD: Rowan and Littlefield Publishers.

Hume, David. (1739) 2000. *A Treatise of Human Nature*. Oxford: Oxford University Press.

Jackson, Frank, and Philip Pettit. 1995. "Moral Functionalism and Moral Motivation." *Philosophical Quarterly* 45 (178): 20–40. https://doi.org/10.2307/2219846.

Jackson, Frank, and Philip Pettit. 1998. "A Problem for Expressivism." *Analysis* 58 (4): 239–251. https://doi.org/10.1093/analys/58.4.239.

Mackie, J. L. 1977. *Ethics: Inventing Right and Wrong*. New York: Penguin.

McPherson, Tristram. 2013. "Semantic Challenges to Normative Realism." *Philosophy Compass* 8 (2): 126–136. https://doi.org/10.1111/phc3.12005.

McPherson, Tristram. 2015. "What is at Stake in Debates among Normative Realists?" *Noûs* 49 (1): 123–146. https://doi.org/10.1111/nous.12055.

McPherson, Tristram. 2022. "Expressivism Without Minimalism." In *Meaning, Decision, and Norms: Themes from the Work of Allan Gibbard*, edited by B. Dunaway and D. Plunkett, 147–170. Ann Arbor, MI: Maize Books, University of Michigan Publishing.

McPherson, Tristram, and David Plunkett. 2017. "The Nature and Explanatory Ambitions of Metaethics." In *The Routledge Handbook of Metaethics*, edited by T. McPherson and D. Plunkett, 1–25. New York: Routledge.

McPherson, Tristram, and David Plunkett. 2020. "Conceptual Ethics and the Methodology of Normative Inquiry." In *Conceptual Engineering and Conceptual Ethics*, edited by A. Burgess, H. Cappelen, and D. Plunkett, 274–303. Oxford: Oxford University Press.

McPherson, Tristram, and David Plunkett. 2021a. "Conceptual Ethics, Metaepistemology, and Normative Epistemology." *Inquiry*. https://doi.org/10.1080/0020174X.2021.1930147.

McPherson, Tristram, and David Plunkett. 2021b. "Metaethics and the Conceptual Ethics of Normativity." *Inquiry*. https://doi.org/10.1080/0020174X.2021.1873177.

Miller, Alexander. 2021. "The Birth of Ethics: Reconstructing the Role and Nature of Morality: By Philip Pettit." *International Journal of Philosophical Studies* 29 (1): 116–121. https://doi.org/10.1080/09672559.2021.1873542.

Moore, G. E. (1903) 1993. *Principia Ethica*. Cambridge: Cambridge University Press.

Olson, Jonas. 2014. *Moral Error Theory: History, Critique, Defence*. Oxford: Oxford University Press.

Paakkunainen, Hille. 2018. "The 'Just Too Different' Objection to Normative Naturalism." *Philosophy Compass* 13 (2): 1–13.

Pérez Carballo, Alejandro. 2020. "Conceptual Evaluation: Epistemic." In *Conceptual Engineering and Conceptual Ethics*, edited by A. Burgess, H. Cappelen, and D. Plunkett, 304–332. Oxford: Oxford University Press.

Pettit, Philip. 1999. *Republicanism: A Theory of Freedom and Government*. Oxford: Oxford University Press.

Pettit, Philip. 2012. *On the People's Terms: A Republican Theory and Model of Democracy*. Cambridge: Cambridge University Press.

Pettit, Philip. 2018. *The Birth of Ethics: Reconstructing the Role and Nature of Morality*. New York: Oxford University Press.

Pettit, Philip. 2020. "Analyzing Concepts and Allocating Referents." In *Conceptual Engineering and Conceptual Ethics*, edited by A. Burgess, H. Cappelen, and D. Plunkett, 333–357. Oxford: Oxford University Press.

Phillips, David. 2019. "Review of The Birth of Ethics: Reconstructing the Role and Nature of Morality by Philip Pettit." *Notre Dame Philosophical Reviews*.

Plunkett, David, and Scott Shapiro. 2017. "Law, Morality, and Everything Else: General Jurisprudence as a Branch of Metanormative Inquiry." *Ethics* 128 (1): 37–68. https://doi.org/10.1086/692941.

Pölzler, Thomas, and Jennifer Cole Wright. 2019. "Empirical Research on Folk Moral Objectivism." *Philosophy Compass* 14 (5): 1–15.

Queloz, Matthieu. 2021. *The Practical Origins of Ideas: Genealogy as Conceptual Reverse-Engineering*. Oxford: Oxford University Press.

Railton, Peter. 2017. "Naturalistic Realism in Metaethics." In *The Routledge Handbook of Metaethics*, edited by T. McPherson and D. Plunkett, 43–57. New York: Routledge.

Sider, Theodore. 2011. *Writing the Book of the World*. Oxford: Oxford University Press.

Vargas, Manuel. 2013. *Building Better Beings: A Theory of Moral Responsibility*. Oxford: Oxford University Press.

Vargas, Manuel. 2023. "Counterfactual Genealogy, Speculative Accuracy, & Predicative Drift." *Inquiry*. https://doi.org/10.1080/0020174X.2023.2225553.

Woods, Jack. 2017. "The Frege-Geach Problem." In *The Routledge Handbook of Metaethics*, edited by T. McPherson and D. Plunkett, 226–242. New York: Routledge.