***The Birth of Ethics: Reconstructing the Role and Nature of Morality*, by Philip Pettit, edited by Kinch Hoekstra, Oxford: Oxford University Press, 2018. Pp. 400.**

Oded Na'aman
Hebrew University of Jerusalem, Israel
Oded.naaman@mail.huji.ac.il

This is a penultimate draft. Please cite published version, available here:
https://doi.org/10.1093/mind/fzab013

Though many mistook his intentions as blasphemes, Voltaire meant to defend God's reality when he wrote, in 1768: 'If God didn't exist, we would have to invent him' (Voltaire 1877-1885, Vol. 10, p. 402). The fact that society *needs* God, Voltaire thought, is a reason to accept God's reality, not a reason to dismiss it. Similarly, Philip Pettit aims to vindicate our moral concepts and practices by uncovering their crucial functions. Some might worry that, like Voltaire's proclamation, Pettit's account unwittingly attacks the concepts and practices it is meant to defend; others would find in this book a new, powerful vindication of morality.

To be sure, neither Voltaire nor God makes an appearance in *The Birth of Ethics*. Rather, Pettit takes his cue from a standard genealogy of money in economics (Menger 1892). A (non-historical) barter society in which people trade various goods and services would converge toward a standard medium of exchange to overcome qualitative mismatches between goods offered and goods wanted. If everyone exchanges gold, or cigarettes, then no one has to worry about being able to offer something the other person wants. The commodity that serves as a recognized medium of exchange does for that society what money does for ours. We thus have an explanation of what money is in terms of its function in society. Pettit purports to tell an analogous story about the emergence of moral concepts and the properties they predicate. I spend the first part of this review presenting an all-too-brief summary of Pettit's fascinating reconstructive analysis of morality. Then, in the second part, I consider whether morality has more in common with God or with money.

*

Pettit draws on a venerable philosophical tradition of genealogical explanations. Specifically, Pettit says that he hopes to do for morality in *The Birth of Ethics* what H.L.A Hart did for law in *The*

Concept of Law (Hart 1961). In his 1961 book, Hart argued that we can gain a better understanding of law by examining how its concepts and practices might have come about and that a pre-legal society would have faced pressures sufficient for the emergence of a recognizable legal system. Pettit argues that the same holds for morality. He purports to explain morality by explaining how moral concepts and practices would arise, more or less necessarily, in a naturalistically intelligible, pre-moral society. The book is a significantly expanded version of his 2015 Berkeley Tanner Lectures.

Pettit begins, in chapter 1, by introducing his methodology; the main story unfolds over chapters 2–6; chapter 7 draws out important implications of the view; it is followed by a concluding chapter that summarizes the book, a response by Michael Tomasello—who has argued for a historical account of the development of morality—and, finally, a rejoinder by Pettit. Though I've learned a great deal from each and every chapter, due to limitations of space I focus here on the main account at the heart of the book.

Imagine an isolated, roughly equal society of largely self-interested people who have beliefs and desires, and a competence in natural language. Individuals in this society communicate with each other by making reports about their environment. They do so for the sole purpose of furthering their own ends. What is crucially lacking from this society are prescriptive concepts and norms. To be sure, simple reasoning and deliberation are conducted as individuals form beliefs and pursue their desires accordingly, but Pettit argues that such reasoning requires neither the concept of a reason nor the concepts of ought or should. Nothing is viewed as credible or desirable. Pettit calls this imagined society 'Erewhon', an anagram of 'nowhere', borrowed from a nineteenth century novel by Samuel Butler. The point of the name being that even if such a society never existed, the fact that it would evolve—more or less necessarily, in Pettit's favorite phrase—to acquire concepts that we would recognize as moral, serves to explain and vindicate our own moral concepts.

Why and how should members of Erewhon acquire concepts we would recognize as moral? The main driver of change in Erewhon is a common interest in mutual reliance for individual benefit, which leads to each member's interest in furthering her reputation as a reliable source of information. At the first stage, each person has an incentive against misinforming others due to her interest in seeming reliable and thereby being able to rely on others in the future. This leads, as if

by an 'invisible hand', to a social pattern of truth-telling. Once this pattern is recognized and becomes common knowledge, it turns into a social norm that informs individuals' choices, though it is not yet prescriptive.

At the second stage, members of Erewhon move beyond making mere reports about their environment. They take this next step because even if they are generally reliable as communicators of reports, they can be even more reliable and they have a reputational incentive to be more reliable. A reliable communicator of reports is still liable to misinform in cases where her reports are based on misleading evidence or in cases where the facts reported about no longer obtain (e.g., it really was raining when you were last outside, so your report was truthful, but now the sun is out.) These possibilities cannot be ruled out when members of Erewhon report about their environment . Therefore, such explanations of misinformation are properly seen as valid epistemic excuses. However, in reporting about their own minds, members of Erewhon can commit more strongly to the information they convey.

In reporting about their beliefs, desires, or intentions, members of Erewhon can foreclose future appeal to the excuse of misleading evidence because their self-knowledge is not based on evidence *about their state of mind*, it is based on evidence *about the content of their state of mind*. To determine that she believes that p, an individual at Erewhon does not rely on evidence that she believes that p; rather, she relies on evidence *that p*. She can therefore foreclose appeal to the excuse of misleading evidence when she communicates information about her own attitudes. According to Pettit's terminology, to foreclose future appeal to the excuse of misleading evidence is not merely to report the presence of a belief that something is the case but to avow that state of belief.

The other way in which members of Erewhon can and would go beyond mere reports, is by *pledging*. In pledging that something is the case one forecloses appeal to both kinds of epistemic excuses. That is to say, one denies oneself future access to the excuse that one has been misled by the evidence *and* to the excuse that the facts reported about have changed since the report was made. Pettit argues that members of Erewhon cannot pledge beliefs and desires because they can never be sure that the data that support their beliefs and the desiderata that support their desires will not change. However, members of Erewhon *can* pledge their own intentions, because intentions can be sustained even when their original basis has changed. Indeed, the fact one

pledged an intention can become the basis of that very intention even if the intention was originally based on a different desideratum. Since pledging involves an even greater reputational risk than avowing, it also involves greater reputational benefits. So members of Erewhon would have an incentive and a capacity to reliably avow various attitudes and to reliably pledge intentions.

Next, members of Erewhon move beyond avowing and pledging to co-avowing and co-pledging. At the current stage, a member of Erewhon can speak for herself rather than merely report about her own attitudes. She does so by avowing and pledging her attitudes. When it comes to others' attitudes, however, she can only offer reports on the basis of evidence. But here, again, members of Erewhon *can* go further and, in addition, they have an *incentive* to go further. Co-avowal involves speaking for others in the way one speaks for oneself in avowing one's attitudes. One way in which co-avowal can happen is when a person is authorized by a defined group to speak in its name. Such group authorization requires that each member of the group pledges to live up to the words the speaker utters on the group's behalf. Alternatively, one can speak for an undefined group without prior authorization by presuming to speak for others and then being vindicated when those spoken for do not object. Co-avowal, Pettit argues, is crucial for conversation, which is in turn crucial for furthering the mutual reliance of members of Erewhon. There is therefore a possibility and an incentive for co-avowal of beliefs and desires.

Co-pledging and co-avowing intentions is more demanding, because intentions, even when supported by desiderata, are much more vulnerable than desires and beliefs to disturbances, such as weakness of will. Given this vulnerability, co-pledging and co-avowing intentions is possible only when one has others' prior authorization to speak for them.

Until this point in the story, prescriptive concepts and norms have not entered the scene. Avowing and pledging, as well as co-avowing and co-pledging, are commitments that are entirely intelligible in non-prescriptive terms of reputational costs and benefits. In taking the next step, inhabitants of Erewhon enter prescriptive space. When they avow their beliefs and desires they do so on the basis of data (in the case of belief) and desiderata (in the case of desire). When their actual beliefs and desires diverge from what they avow, they must view these attitudes as failing with respect to the data or desiderata that were the basis of the original avowals. They will therefore view data as that which determines what they and others ought to believe and desiderata as that

which determines what they and others ought to desire. What they ought to believe is what is credible and what they ought to desire is what is desirable.

Now there are prescriptive concepts at Erewhon, but there are not yet moral concepts. The first moral concept—moral desirability—emerges in response to a need to overcome the problem of multiply conflicting prescriptions. The problem does not arise with regard to credibility. There are no multiply conflicting prescriptions of belief because the same body of evidence would support the same belief from different perspective and for different individuals. However, the same desiderata need not support the same desires from different perspectives and for different individuals. What is attractive to one person may not be attractive to another: that the action would help my father or serve my life project is attractive to me but not to you (or it is not as attractive to you as it is to me.) Even within the perspective of a single individual, different sets of facts often support different desires on the same occasion. This means that one might be divided from others as well as from oneself with respect to what to desire. These divisions lead to unpredictability and thus to obstacles for mutual reliance.

The concept of moral desirability addresses the problem at hand. It would be attained by filtering out all agent-relative desiderata, or, more minimally, by filtering out agent-relative desiderata that put people in competition with one another. In other words, the morally desirable is what is desirable at least for some without being undesirable for any. The resulting notion of moral desirability makes possible a common perspective on the prescription of desire which would, in turn, have a special authority in Erewhon due to its role in enabling mutual reliance.

Moral desirability is the first key moral concept Pettit aims to explain. The second is responsibility. Inhabitants of Erewhon would be driven to develop a concept akin to our own concept of responsibility given their interest in, on the one hand, making failures to abide by standards of moral desirability costly, while, on the other hand, not immediately excommunicating those who fail such standards. The idea is that each member would want, for her own sake, that occasional failures to meet one's commitments would be burdensome without putting an end to mutual reliance. There would therefore be a presumption that others have a capacity to pledge fidelity to standards of moral desirability. Repeated failures would have to cross a certain threshold before the presumption is undone.

Moreover, in response to another's failure to meet standards of moral desirability, members of Erewhon would retrospectively exhort this person to have responded as she should. That is to say, they would express the message that prior to her failure it would have been appropriate to exhort her to succeed to meet the relevant standards. Such an injunction expresses their expectations of success and makes the reputational costs of failure salient. Finally, members of Erewhon would reprimand those who fail to respond to considerations of desirability, which is to say that they would register the failure, express a bad opinion of it, and view these reactions as appropriate.

Thus there are three ideas and forms of conduct at play: the idea that (in the absence of excuses and exemptions) those who violate standards of moral desirability have the capacity to meet them; the idea that it would have been appropriate to exhort them to meet these standards and it is appropriate to retrospectively exhort them to have done so; and the idea that in light of their failure it is appropriate to reprimand them. All three ideas can be expressed in the assertion 'you could have done otherwise' and, Pettit argues, they jointly constitute our own concept of responsibility. Once the concept of moral desirability and the concept of responsibility are at hand, they make available the concept of moral obligation: the morally obligatory is the morally most desirable option that it would be wrong or blameworthy for the agent not to take (the qualification excludes morally desirable options that are supererogatory.)

This summary does not do justice to Pettit's rich and careful argumentation, but I hope it encourages those unfamiliar with the book to read it, and read it carefully. Once the exploration of Erewhon's path to morality concludes in chapter 6, Pettit goes on, in the seventh chapter of the book, to explore various implications of the account for moral metaphysics, moral semantics, moral epistemology, moral psychology, and normative ethics. In its discussion of metaphysics, the seventh chapter defends a thesis Pettit considers to be one of the motivating claims of the study, namely, that 'it is in virtue of the commissive practices at the origin of morality that we human beings come to constitute functioning persons' (243). This claim, Pettit maintains, provides an answer to the question 'why be moral?'. The metaphysical thesis, as Pettit calls it, is powerful and intriguing, and it is well worth a lengthy discussion, but instead I now turn to considering Pettit's main account.

There are two sets of questions that can be asked about Pettit's story of Erewhon. The first set of questions asks whether the story is plausible? That is, whether such a society would develop in the way Pettit claims it would? As such, these questions investigate whether the account succeeds by its own lights. The second set of questions is about Pettit's methodological aspirations. Even if Erewhon would develop as Pettit argues, what would that establish? In particular, can the story of Erewhon fully explain and vindicate our moral concepts? In the remainder of this review, I focus on the second, methodological set of questions.

Pettit calls his genealogical methodology *reconstructive analysis*, and in the first chapter of the book he argues that a reconstructive analysis can deliver where other explanations of morality fall short. As Pettit sees it, we want an explanation of morality that accommodates both *realism* and *naturalism*. To accommodate realism an analysis of morality must explain moral desirability and responsibility as bona fide, prescriptive properties. To accommodate naturalism, an analysis of morality must be compatible with the claim that all the properties realized in the actual world are liable to figure in natural science or to be realized by one or another configuration of properties that figure in natural science. The problem is that natural science does not include prescriptive properties and it is not clear how prescriptive properties might be realized by some configuration of the fundamental properties of natural science.

In response to this apparent dilemma, some theories, such as expressivists theories, give up realism while other theories, such as non-naturalist realism, give up naturalism. Pettit wants to hold on to both horns of the dilemma: 'The idea in this book is to resist downgrading ethical discourse ... and, without forsaking naturalism, to try in the spirit of realism to vindicate the assumption that there really are properties like desirability and responsibility in the world and that they have an impact on our actions' (Pettit 2018, 19).

Pettit contrasts his reconstructive solution with naturalistically reductive solutions, which also adopt naturalism and realism. A naturalistically reductive account of morality involves a claim about what non-prescriptive conditions instantiate moral properties and a claim that some configuration of naturalistic properties in the actual world can satisfy these conditions. The first claim is an analysis of moral concepts usually reached by the method of cases; the second claim is an empirical claim about the relevance of the analysis to the actual world.

Pettit's reconstructive analysis is meant as an attractive alternative to naturalist reductivism. The reconstructive analysis, he argues, has more explanatory power than a reductive analysis and fewer commitments. The reconstructive analysis has more explanatory power because, unlike a reductive account, it appeals to an insider's understanding of the development of social practices into moral practices. The reconstructive analysis has fewer commitments because, unlike a reductive account, it is committed neither to a particular conception of the application of moral concepts nor to an account of the natural properties that instantiate moral concepts.

In contrast to the two claims of the reductive account, Pettit's reconstructive account claims:

> First, that insofar as the terms or concepts that emerge in the story [of Erewhon] respond to the same sorts of prompts, and serve the same sort of purposes, as our actual ethical terms, the properties they predicate are good candidates for the properties we ourselves predicate with them. And, second, that since the appearance of those concepts in a predicative role is naturalistically explicable, the properties they ascribe . . . must be naturalistic too; if the concepts ascribed non-natural properties, after all, then those properties would presumably have played a role in explaining how the concepts came into use. (Pettit 2018, p. 20)

Does it follow from the fact that our moral concepts serve the same functions as their analogues in Erewhon that they predicate the same properties? In his *NDPR* review of the book, David Phillips suggests that the answer is 'no'. Phillips argues that two concepts can serve the same functions and yet differ in the properties they predicate. Thus, the first claim in the above passage depends on an additional reductive analysis of our moral concepts. Therefore, the reconstructive account does not constitute a genuine alternative to reductive accounts (Phillips 2019). However, even if we grant that the functional equivalence between Erewhon's concepts and our own moral concepts implies that these concepts predicate the same properties, we might still wonder whether the reconstructive account satisfies the requirement of realism: does it explain moral desirability and responsibility as bona fide, prescriptive moral properties?

Compare Pettit's reconstructive account of morality to a reconstructive account of God and related religious beliefs. Suppose our concept of God predicates the same properties as its naturalistic, functional analogue in Erewhon. This would constitute a naturalist, anti-realist analysis of God. On such a view, our concept of God predicates very different properties than the properties it is normally taken to predicate by those who believe that God exists. A successful

reconstructive account, in this case, would be debunking rather than vindicating, as Voltaire took it to be. Though the account might vindicate our *use* of the concept, it would debunk a theistic understanding of it.

How about a reconstructive account of money, such as the one mentioned at the outset: is it debunking or vindicating? Insofar as such an account implies that the value of money is fully explained by its functional roles, it would seem to undermine an understanding of money as intrinsically valuable. Those who attribute to money intrinsic value would view the reconstructive account as anti-realist and debunking. But for many of us, a view of money as intrinsically valuable seems fetishistic. Thus, for those already disinclined to attribute intrinsic value to money, a reconstructive account of money offers a vindicating, realist explanation—it can explain money as the kind of thing we normally take it to be.

These examples show that what counts as realism about a certain kind of property depends on how we understand a bona fide instance of the property. Pettit's reconstructive analysis of morality raises a similar question: what are bona fide prescriptive, moral properties? It is clear that in our daily life we do not and should not apply moral concepts by considering the functional role of the widespread use of these concepts. It might therefore seem tempting to conclude that, when compared to the above examples, morality stands closer to God than to money: Pettit's account fails to meet the realism constraint on an adequate analysis of morality. However, this conclusion would be too quick. Pettit would be the first to grant that we do not and should not use our moral concepts by appealing to their functions. Rather, he argues that our non-instrumental use of these concepts is explained by the social function of using them so.

There is, therefore, a crucial difference between our use of the concept of money and our use of moral concepts. Our use of the concept of money is entirely transparent to its functional explanation: it is instrumental all the way down, to the level of individual motivation. By contrast, our moral concepts and practices have what Matthieu Queloz calls 'self-effacing functionality' (Queloz 2021, p. 54).

> Where functionality is *self-effacing*, it is a functional requirement on the practice's functionality that participants not be primarily motivated by awareness of that functionality, but when they acquire awareness of it, this awareness is fully compatible with—and may indeed encourage—confident engagement in the practice on any reasonable conception of it (p. 55).

A practice has self-effacing functionality when 'non-instrumental motivations are required to sustain the practice' (p. 58).

If Pettit's account is to meet the realism constraint, it must construe morality as having a functionality that is—unlike money's functionality—self-effacing. On this view, individuals' moral motivations are non-instrumental and not conditioned on awareness of the functionality of moral concepts, and yet moral motivation is compatible with (and perhaps reinforced by) awareness of the functionality of moral concepts. A functional account of morality can therefore be compatible with the idea that bona fide moral properties merit non-instrumental engagement.

Voltaire might have made use of this suggestion to defend his functional account of God. Of course, he might have said, our faith in God is not primarily based on the usefulness of our concept of God, but our faith is compatible with the concept's usefulness and is reinforced by it. However, if God's authority is only reinforced by the social function of the concept of God, then it is not fully explained by it. There remains a need for a non-naturalist explanation. It is, after all, not surprising that religious realism drives us away from naturalism, but the same tension seems to afflict moral realism. While the idea of self-effacing functionality might calm worries about Pettit's realism, it raises worries about naturalism, the other adequacy constraint on an analysis of morality.

If it is essential to our moral concepts and practices that individual motivation is not conditioned on awareness of their functionality, then it might seem that the functionality of our moral concepts cannot provide a full explanation of their non-instrumental authority. A functional analysis of morality that aspires to vindicate morality as we normally understand it is therefore essentially incomplete and depends on a non-functional analysis of our moral commitments. It follows that a naturalist functional story, such as Erewhon's, is either incomplete or debunking. Accommodating the realism constraint seems to make it harder to meet the naturalism constraint, and vice versa.

Notwithstanding the aforementioned doubts, morality, unlike religious faith, is not explicitly committed to a non-naturalist self-understanding. Therefore, a naturalist, reconstructive analysis of morality is not as obviously debunking as an analogous analysis of religious faith. If such an account of morality is forthcoming, this impressive book—innovative in both substance and method—does a great deal to bring it to light.[*]

Oded Na'aman
Hebrew University of Jerusalem, Israel
Oded.naaman@mail.huji.ac.il

*I am grateful to Dan Baras, Philip Pettit, and Matthieu Queloz for comments on earlier versions of this review. I thank Ynon Wygoda for a helpful conversation about Voltaire.

References

Hart, Herbert 1961, *The Concept of Law* (Oxford: Oxford University Press)

Menger, Carl 1892, 'On the Origin of Money', in *Economic Journal* 2: 239–255

Phillips, David 2019, 'Review of Birth of Ethics: Reconstructing the Role and Nature of Morality', in *Notre Dame Philosophical Reviews* <https://ndpr.nd.edu/news/the-birth-of-ethics-reconstructing-the-role-and-nature-of-morality/>

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

Voltaire 1877-1885, 'Epître à l'auteur du livre des Trois imposteurs', in *Œuvres complètes*, edited by Louis Moland, Vol. 10 (Paris: Garnier)