

Thomas Pözlner, *Moral Reality and the Empirical Sciences*. New York, NY: Routledge, 2018. Pp. viii + 237. ISBN 978-1-138-50309-0.

Published in *International Journal for the Study of Skepticism*. ‘Accepted Manuscript’ version.

How should one live? What do we owe to each other? What kind of person should I be, and, while we are at it, what matters? According to moral realism, these questions have objective answers, grounded in observer-independent moral facts. Accordingly, the things that matter, for instance, are not determined by what anyone takes them to be. Moral anti-realists of various stripes, however, do not concur with this story. Sceptics worry that normative questions have no answers at all, or contend that we are hopeless at finding out, while others maintain that normative answers are not objective after all. Both realists and anti-realists increasingly turn to scientific investigations to support their views. However, the debate between realists and anti-realists is ultimately metaphysical. Is there any hope that it can be scientifically evaluated?

Thomas Pözlner’s book offers the first detailed study that focuses explicitly on the promise of science-based arguments for and against moral realism (of both the natural and non-natural kind). His two central claims are that sound arguments bearing on the realism/anti-realism debate are possible, and, yet, that four central attempts to derive metaethical conclusions from science-based arguments uniformly fail. The book then provides several recommendations for future science-based contributions to the realism/anti-realism debate to do better.

The book is a valuable and thought-provoking contribution to experimental metaethics and, in particular, to debates surrounding experimental studies of folk-moral realism, where it

provides an insightful and handy guide to the field. In what follows, I briefly lay out the book's main arguments, focusing on Pölzler's assessment of existing science-based arguments, along with some critical remarks, before providing a broader evaluation.

In chapter 1, Pölzler explains the critical components of moral realism and previews the book. Pölzler's construal of realism is standard, except for the strong association between semantics and metaphysics he assumes. He classifies moral facts as "those facts that we purport to refer to when we speak and think about morality" (5). Accordingly, non-cognitivism implies that there are no moral facts. That assumption is not defended, and appears implausible (cf. Kahane, 2013).

In chapter 2, Pölzler considers a priori arguments against the relevance of science-based arguments in metaethics. Pölzler aims at establishing that sound science-based arguments bearing on moral realism are possible, and he achieves this aim convincingly. Some discussions may seem forced to philosophers familiar with the topic (e.g. of the misguided contention that *any* empirical investigation bearing on moral realism would violate Hume's law), but they might be helpful to newcomers. Assessing the claim that conceptual accounts of moral judgement are logically prior to experimental work leads him to conclude that, first, experiments on intuitions are unaffected by it and, second, that the worry does not exclude *some* relevance of experimental findings in the process of reflective equilibrium (34). I return to this below.

The next four chapters of the book then evaluate four (classes) of science-based arguments directed at moral realism: the presumptive argument, the argument from disagreement, the sentimentalist argument, and evolutionary debunking arguments. Pölzler picks an influential representative of each argument and, in each case, his strategy involves two main steps. First comes clarification of the empirical premise, and then a reality check against

(some of) the available empirical evidence interpreted in light of the clarifications provided before.

Chapter 2 tackles the presumptive argument. Pölzler reconstructs the argument, leaning on McNaughton (1988), as an argument dependent on phenomenal conservatism as follows (45):

1. It seems to ordinary people that morality is a realm of objective truths.
2. If it seems to a person that *p*, then this person has a *prima facie* reason to believe that *p*.
3. So, ordinary people have a *prima facie* reason to believe that morality is a realm of objective truths.

Pölzler's assessment of the presumptive argument is devastating: on closer consideration, he argues, "research on folk moral realism has not been valid at all" (57, 77). He diagnoses problems both with the *external validity* (to wit, whether findings obtained from a sample in an experimental setting apply to the population in general) as well as the *construct validity* (to wit, the degree to which a test measures what it is supposed to measure) of existing studies. Pölzler's worries about construct validity are most interesting (61-76):

- (1) In all studies, test items may, inadvertently, fail to reflect moral issues in the first place, because the realm of the moral is theoretically contested.
- (2) In all studies, recorded answer options may not reflect moral intuitions but instead random or systematic distortions and moral intuitions may not reflect metaethical stances.
- (3) Almost all studies assume that subjects accept a correspondence theory of truth. If that assumption is mistaken, then answer options to the effect of 'this moral statement is true' may not reflect realist intuitions.

(4) In several studies, answer options do not reflect the full breadth of anti-realist options.

Consequently, erroneous inferences may be drawn from some answers.

- a. Asking subjects to describe a moral disagreement as either ‘both disputants can be correct’ or ‘at least one disputant must be mistaken’ fails to capture possible non-cognitivist seemings.
- b. Subjects that chose ‘at least one disputant must be mistaken’ need not be objectivist moral realists but could be subjectivists.
- c. Subjects that classify a moral statement as universal (to wit, as true at all times) need not be realists, nor need subjects rejecting such classification be anti-realists.

(5) In several studies, test items and answer options may illegitimately evoke first-order moral intuitions. Consequently, first-order intuitions might illegitimately be conflated with subjects’ metaethical intuitions.

(6) In at least one study, answer options illegitimately evoke epistemic intuitions when varied in terms of certainty. Consequently, those intuitions may distort findings.

(7) In at least one study, positive instructions may have biased subjects toward anti-realism.

The discussion of the presumptive argument is the most exciting part of the book. The fact that moral realism is taken as the dialectical starting point in metaethics has not received sufficient critical attention thus far. Pözlner is right that moral realists have mostly focused on rebutting anti-realists arguments, assuming that theirs is the position to beat, rather than providing positive reasons in favour of their view. Given that the argument overtly relies on premises that can be empirically investigated, Pözlner’s comprehensive methodological assessment is overdue and highly valuable for scholars attempting to set-up similar experiments in the future.

Nonetheless, the details of Pölzler's rebuttal of the presumptive argument raise too many questions to be convincing. Pölzler sometimes strongly denies that these studies are valid at all (see above), but his weaker and more considered conclusion is that they are at least "low in construct validity" and that they have "failed to (fully or exclusively) measure what they have purported to measure" (59). But even the weaker conclusion is not convincingly supported by the evidence he provides. The problems he finds in all studies (e.g., the problem that test answers *might* not reflect genuine moral intuitions) do not seem damning, and those that might be damning (e.g., that test items evoke epistemic rather than moral intuitions) apply to only some studies. At least some studies, therefore, should have something valid to say about the truth or moral realism.

To illustrate an issue that applies to almost all studies, but fails to be damning, consider point 4a above. Almost all studies ask test subjects to judge a moral disagreement between two people. Subjects have two answer options: declare that both disputants can be right, or that at least one must be mistaken. However, people might be noncognitivism. Noncognitivists, Pölzler points out, will not find either option appealing: according to at least some of them, there is no such thing as 'being right' about moral matters. But their noncognitivism may be of two relevant classes. Their views might resemble emotivism or projectivism, and thus, they should deny that moral claims can be true or false at all. Given only options that presuppose the truth-aptness of moral claims, such subjects would find no answer option fitting – if they answer at all, their answer would be random and invalid indeed. Alternatively, subjects' views might resemble contemporary expressivism, in tandem with deflationary theories of truth. Subjects from this camp might answer that at least one party to the disagreement must be mistaken; their answers will erroneously count as realist, hence invalidating these studies. All but one of the about 20 studies that Pölzler discusses potentially include invalid responses from

noncognitivists. Is it a damning problem? Only if we have reason to believe that test subjects are noncognitivists of either.

Do we have empirical evidence that subjects are noncognitivists from the emotivist or projectivist camp? To the contrary. For instance, Goodwin and Darley (2008) offered an ‘other’ answer-option in their disagreement-item. Subjects aware of their noncognitivist leanings could have chosen it. No participant in Goodwin and Darley’s study chose that option when it came to moral matters. Of course, one study does not tell much. The point is, however, that it would have strengthened Pözlner’s case to explain why we should expect further studies to yield relevantly different results.

Do we have evidence that a significant number of subjects are noncognitivists from the expressivist camp? Contemporary noncognitivists, of course, say ‘yes!’ But to the empirically-minded metaethicists, mere contentions should be suspect, and Pözlner does not review any empirical evidence that might support their claim. On this evidential basis, therefore, it is unwarranted to discredit existing studies based on problem 4a.

To illustrate two further issues that apply to all studies, but fail to be damning, consider points 2 and 3 above (I return to point 1 below). Consider the allegedly suspect inference from test data to moral intuitions to metaethical stances. Pözlner records a perfectly *possible* theoretical issue, which plagues any experimental study. For instance, subjects selecting realist answer options might be unaware of their own non-cognitivism. However, Pözlner fails to provide sufficient evidence to make that issue *probable*. Consequently, though test data *might not* reflect test takers moral intuitions, Pözlner has provided scant reason for thinking that this is the case and that point should not be held against the validity of these studies.

Similarly, though almost all subjects *may* be deflationists about truth, which may invalidate findings from experiments that assume a correspondence theory of truth, Pözlner does not give us sufficient evidence to believe that this is a *probable* scenario. Such scenarios

should be empirically investigated to ascertain whether they hold water. Identifying possibilities in logical space are a starting point for such investigations, but insufficient as evidence for damning judgments about the validity of existing studies.

Of course, there is room for reasonable disagreement about how to best interpret the evidence. Moreover, my complaint depends on the undefended assumption that the most parsimonious explanation of subjects' answers is that they, at least implicitly, hold a correspondence theory of truth and that their answers concerning truth in paradigmatically moral scenarios reflect metaethical intuitions and that, therefore, their answers can, *ceteris paribus*, be taken at face value. Pölzler's discussion of the presumptive argument raises the important question of whether these assumptions are warranted. To scholars working in the field, it also poses the pragmatic challenge of settling on a set of shared, defeasible, and reasonable assumptions about morality to guide the design of future experiments.

Pölzler's discussion of the remaining three arguments proceeds according to the same schema. He takes an exemplary proponent of a science-based anti-realist argument, reconstructs the argument, clarifies the terms of the empirical premise, and then assesses some empirical work bearing on it. I recap these arguments more briefly.

First comes the argument from disagreement in chapter 4. Based mainly on Mackie's (1977) well-known exposition, Pölzler reconstructs the argument as involving an inference to the best explanation, according to which the lack of objective moral truths is the best explanation of wide disagreement about matters of morality (91). He limits his assessment of the empirical record in support of the claim that people disagree widely about morality to studies by Abarbanell and Hauser (2010), Nisbett and Cohen (1996), and an unpublished study by Peng, Doris, Nichols, and Stich, which is discussed by Doris and Plakias (2008). Again, things do not look good for proponents of science-based arguments. According to Pölzler, "eliminative explanations" (102) exist for all relevant disagreements reported in those studies:

the authors rely on illegitimate inferences from observational data to moral judgement, uncover differences in non-moral rather than moral judgements, or, finally, report moral differences too small to be significant. For example, when someone *claims* that sacrificing humans is permissible, or acts as if it is, then, cautions Pölzler, we cannot infer with certainty that that person also *morally judges* that sacrificing humans is permissible (104). After all, that behaviour might be caused by, for example, social pressures rather than moral conviction. Thus, Pölzler claims that “any plausible alternative explanations would have to be ruled out” (104) before concluding that wide-spread moral disagreement exists.

Next up is Jesse Prinz’s (2007) sentimentalist argument in chapter 5. Sentimentalism is a theory about the nature of moral judgement, and Prinz’s variant identifies moral judgements with the disposition to have certain emotions. Pölzler reconstructs the argument for sentimentalism as an inference to the best explanation, according to which the “close empirical association” of moral judgements and emotions is best explained by the fact that (dispositions for) emotions constitute moral judgements. Again, Pölzler’s verdict is damning: Though he acknowledges that the “number of potentially relevant studies” for assessing the empirical basis of sentimentalism is extraordinarily high (143), he claims that “[m]ost of them lack in external or internal validity” (132, emphasis added) and that the evidence “does not allow *any* assessment of sentimentalism’s emotionism-related empirical predictions” (138, emphasis added).

Irrespective of the details of Pölzler’s analysis, it is not clear that sentimentalism is relevant for his project. As he acknowledges, sentimentalism does not bear on the realism/anti-realism debate directly. Still, Pölzler claims that an argument for sentimentalism is naturally interpreted as “an indirect inductive argument against realism” (131). However, since, as Pölzler notes himself, moral realism is perfectly compatible with moral judgments being constituted by (dispositions for) emotions (e.g. Roeser, 2011), it would have been interesting

to see how that inference works out. As it stands, the conclusion of the sentimentalist argument is compatible with both realism and anti-realism, which raises doubts whether the argument is all that relevant for Pölzler's project.

Finally, in chapter 6, Pölzler turns to evolutionary debunking arguments, taking Richard Joyce (2006) as his principal interlocutor. His reconstruction of Joyce's debunking argument starts with the premise that moral judgements are adaptations and, via the claims that adaptations do not track moral truth and that non-truth tracking belief-forming methods defeat, concludes that moral judgements are unjustified (183). Pölzler's interest lies in the question of whether moral judgements are adaptations and he concludes that existing arguments "so far failed to yield valid evidence [to the effect that moral judgements are adaptations], and may not be able to yield strong evidence at all" (212). Again, that is terrible news anti-realists hoping to capitalise on an empirically-sound genealogy of morality.

Pölzler's overwhelmingly negative assessment of all four science-based arguments does not entirely convince me. Partly, this is because of dialectical choices. Pölzler does not confront the most persuasive arguments. For example, evolutionary debunking arguments need no premise about moral judgements being adaptations. They may work with noting mere influences on moral judgements, which lowers the empirical hurdle for such arguments to be sound (cf. Klenk, 2018).

The problem extends to Pölzler's assessment of the empirical evidence, however. His criticism is too often merely suggestive and, therefore, not as strong as it could have been. Pölzler often correctly shows that there is logical space for alternative explanations about, for instance, the processes that drove answers in survey studies. However, the point of such studies is, as Pölzler would surely agree, not to conclusively show that no explanation save a debunking one is possible, but rather that such explanations are most plausible. Pölzler mostly hints at possible avenues without pursuing them. At most, Pölzler's arguments call for restraint in

interpreting the available evidence – but not to disregard it as evidence completely. Pölzler’s strong criticism of existing science-based arguments is much more plausible when one is willing to challenge seemingly innocuous assumptions, such as the claim that answers in paradigmatically moral scenarios defeasibly reflect moral intuitions. As suggested above, the book raises the critical question of which of these assumptions are reasonable, given the practical constraints of setting up experimental studies that cannot do without some such assumptions.

A concluding chapter recaps the arguments and recommendations of the book, zeroing in on recommendations concerning the operationalisation of moral judgement. When running experiments, such as survey studies, he urges experimenters to make sure that they elicit moral judgements rather than judgements about legal or prudential wrongs, or judgements about social acceptability (e.g. 115, 119). Moreover, he demands that the “presence of [moral] judgements” be tested (149), and he cautions that whether science-based arguments are contingent on “conceptual points” about “what one understands by those judgements” (212), which ties back to his discussion of the ‘logical priority objection’ from chapter 2 (33).

Staunch armchair philosophers might yell ‘Gotcha!’ at this point: We do not have an account of settling purely experimentally the presence of moral judgements, or so they might argue. Hence, it seems that, after all, we would first have to settle conceptual questions about what moral judgements are in a non-experimental way before running an experiment to find out more about these judgements.

Even if there is no in-principle hurdle to settle such questions empirically, Pölzler’s recommendations for circumventing the problem of operationalising the nature of moral judgement are suggestive but incomplete. One recommendation is conditionalizing; that is, one assumes that moral judgements are so-and-so, and then conducts the study. However, as Pölzler acknowledges, conditionalizing one’s argument on a specific explicitly articulated

understanding of moral judgement may avert worries of soundness, but only on a superficial level (33). Researchers disagreeing with the conditionally accepted account of moral judgement may dismiss the argument as entirely irrelevant.

Pölzler's more profound proposal is that moral judgements should be operationalised in the process of reflective equilibrium. That process seems to involve both conceptual considerations about whether given findings bear on moral judgements while "experimental findings can bear on our accounts of moral judgements as well" (229). Pölzler's account leaves open how, exactly, experimental findings play a role in reflective equilibrium. He suggests that given an externalist moral semantics, experimental work only needs to presuppose paradigmatic cases of moral judgements, thus rather securely avoiding the charge of presupposing the wrong account of moral judgement (34). However, the pressing methodological question is when experimental findings in tension with one's account of moral judgement disconfirm it and when they do not. Despite his contention to the contrary from chapter 2, Pölzler appears to play into the hands of the staunch armchair philosopher by repeatedly rejecting the relevance of science-based arguments for employing controversial accounts of moral judgement. It does not seem, then, that experimental findings provide input to conceptual accounts of moral judgment. Answering how this would play out would show much more concretely how empirical science can make an impact on moral reality.

Despite these open questions, the book is a thought-provoking and notable addition to the debate about experimental investigations of intuitions bearing on the truth of moral realism. It is written in an accessible style and, given the detailed and careful questions it raises, it will help experimenters improve their methodology in future studies. It can be used both in undergraduate and graduate courses on (experimental) moral psychology, both by philosophers and non-philosophers.¹

¹ Thanks to Thomas Pölzler for helpful comments on a previous version of this review.

Michael Klenk

Delft University of Technology

m.b.o.t.klenk@tudelft.nl

1 References

- Abarbanell, L., & Hauser, M. D. (2010). Mayan morality: An exploration of permissible harms. *Cognition*, *115*(2), 207–224. <https://doi.org/10.1016/j.cognition.2009.12.007>
- Doris, J. M., & Plakias, A. (2008). How to argue about disagreement: Evaluative diversity and moral realism. In W. Sinnott-Armstrong (Ed.), *Moral Psychology: Vol 2: The cognitive science of morality* (pp. 303–331). Cambridge, MA: MIT Press.
- Goodwin, G. P., & Darley, J. M. (2008). The Psychology of Meta-ethics: Exploring Objectivism. *Cognition*, *106*(3), 1339–1366. <https://doi.org/10.1016/j.cognition.2007.06.007>
- Joyce, R. (2006). *The evolution of morality. Life and mind*. Cambridge, MA: MIT Press.
Retrieved from <http://search.ebscohost.com/login.aspx?direct=true&scope=site&db=nlebk&db=nlabk&AN=156012>
- Kahane, G. (2013). Must Metaethical Realism Make a Semantic Claim? *Journal of Moral Philosophy*, *10*(2), 148–178.
- Klenk, M. (2018). Evolution and Moral Disagreement. *Journal of Ethics and Social Philosophy*, *14*(2), 112–142. <https://doi.org/10.26556/jesp.v14i2.476>
- Mackie, J. L. (1977). *Ethics: Inventing right and wrong*. London: Penguin.
- McNaughton, D. (1988). *Moral vision: An introduction to ethics*. Oxford: Basil Blackwell.
- Nisbett, R. E., & Cohen, D. (1996). *Culture of honor: The psychology of violence in the South*. Boulder, CO: Westview Press.
- Prinz, J. J. (2007). *The emotional construction of morals*. New York, NY: Oxford University Press.

Roeser, S. (2011). *Moral emotions and intuitions*. New York, NY: Palgrave Macmillan.