Dominik Weinhold MA, "Minding the Gap: Moral Disagreement and the Limits of Moral Reasoning"

Opponent's Report

This diploma thesis explores the implications of moral disagreement for the prospects of objectivist accounts of moral judgement, or 'moral realism'. Drawing attention to recent findings in cognitive science, the author argues that moral reasoning is not fully conscious, rational, or formally valid, and therefore he reaches a sceptical conclusion about the possibility of convergence in ethical judgement, even in ideal conditions. This then leads to a wider sceptical perspective on the question of an objective moral order, or of a moral reality existing independently of our diverse attitudes and stances.

The author begins by reflecting on the nature of moral disagreement and makes a range of germane observations and useful distinctions, drawing on the work of numerous philosophers, particularly in the recent analytical tradition. Of cardinal importance here is the distinction between 'normative' and 'meta-ethical' disagreement developed on pages 20-24. Normative disagreements concern judgements about concrete people, acts, and situations, while meta-ethical disagreements concern 'second-order'—peculiarly philosophical—questions about how we should interpret moral value, whether realism or anti-realism obtains, and whether moral judgements have truth-conditions (the question of cognitivism or non-cognitivism).

The distinction between these two kinds of disagreement is of great importance and I would suggest that it could be further clarified in the defence. Above all, it is not clear to me where disagreement between philosophical explanatory theories would fit in. Do competing theories, such as utilitarianism, Kantianism, divine law, and Aristotelian virtue ethics, belong to the normative or meta-ethical side of the distinction? In addition, it was not clear to me how systematically this distinction is observed in later parts of the thesis. For example, on page 38 it is argued that empirical surveys show that moral disagreement is prevalent, but it is not clear whether this research, and the conclusion reached on its basis, is sensitive to the normative-meta-ethical distinction. Again, four different objectivist explanations of moral disagreement are enumerated (pp. 45-6), but these explanations seem to break down into those that deal with normative theory (1-3) and an explanation that deals with meta-ethical theory (4). This fact is not remarked on by the author and, generally, the earlier distinction (pp. 20-24) seems almost to have been forgotten in later parts of the thesis.

Having characterized the different forms of moral disagreement and their relation to philosophical questions such as cultural relativism, moral scepticism and realism, the author conducts a fascinating discussion, in chapter IV, of the relevance to moral philosophy of empirical research in cognitive science and psychology. This is perhaps the most original section of the thesis. Led by the work of different empirical investigators, the author argues that our internal reasoning is much less accessible to introspection than we might ordinarily assume (pp. 64-66), and further argues that reasoning can only be properly surveyed when it is externalised in linguistic inter-personal communication (p. 66). The picture that arises is in line with the 'argumentative theory', put forward by Mercier and Sperber (whose article is cited on page 67, although the author might have used their later and more comprehensive 2019 book *The Enigma of Reason*). Reasoning, on their view, is primarily designed not for individual problem-solving but for persuading and justifying others. This means that our moral reasoning faculties have no special tendency to pursue self-critical objective judgement, but rather they

seek to back up intuitive feelings and hunches: 'people dig in their heels', the author writes, 'and their effort to rationalize and justify their basic moral intuitions turns competitive, leaving purported truth-tracking aims of moral inquiry aside' (pp. 67-68). The implication being that the ethical convergence sought by moral realists is a will-o-the-wisp, and the picture of our rational natures that the realist assumes is 'inconsistent with the empirical knowledge of human psychology' (p. 68).

Throughout this discussion there seems to be a guiding assumption that if the formation of moral judgements takes place unconsciously, then the processes involved must be 'grounded in the affective or conative states of mind' (p. 67). The assumption that unconscious processes must be non-rational, adopted perhaps from a loosely Freudian background (Freud is invoked on p. 65), is not shared by all contributors to cognitive science. For example, Noam Chomsky's widely influential work on language suggests that the processes involved in our parsing and forming syntactical structures are unconscious and not open to introspection. It would be quite wrong, however, to treat these workings of the linguistic faculty as 'grounded in the affective or conative states of mind'. In general, the possibility of unconscious but intelligent and rule-directed mentation is perfectly respectable, and the possibility that there is a moral faculty parallel to the language-faculty is one that has been entertained.

Overall, this is a very good systematic and focused study of the relation between moral disagreement and the question of moral realism or objectivism. It is balanced in its approach, shows wide reading in the specialised literature, and it reaches a reasonable and nuanced conclusion. I would add that the academic English in which it is written is of a very high level.

I unreservedly recommend the thesis for defence, and I suggest the top grade (*výborně*).

Prof. James Hill, PhD