

This is the unproofed author MS. The final and official version can be found at:

<https://www.tandfonline.com/doi/full/10.1080/09515089.2024.2309270>

(Almost) everything you've always wanted to know about moral reasoning and decision making -

The Oxford Handbook of Moral Psychology. Edited by Manuel Vargas and John Doris. Oxford, Oxford University Press 2022, 1120pp., \$190 (hardback), ISBN: 9780198871712

Moral Psychology spans both philosophy and psychology: it addresses questions concerning the role of emotions in moral judgment, the nature of moral motivation, whether human beings are ultimately selfish, why we blame people, and many others. Some of these questions require a descriptive answer that specifies how our moral reasoning and decision-making works. Others are explicitly normative, for example when we ask how people should live (what constitutes a good life), what virtues are, or whether addicts are blameworthy for bad decisions motivated by their desire for drugs. But even more straightforwardly philosophical work in moral psychology that addresses clearly normative questions like 'do mental disorders excuse from responsibility', or 'should negligence be legally culpable', needs to take into account empirical findings relevant to the question at hand. We should therefore expect good work in moral psychology to be interdisciplinary to some extent. The new Oxford Handbook of Philosophy embraces this empirically informed approach to moral psychology wholeheartedly.

The volume does clearly fall on the 'Philosophy' side of the moral psychology literature. It is edited by two philosophers, Manuel Vargas and John M. Doris. The majority of contributors are also philosophers, though there are some psychologists too, and many of the contributing philosophers work across disciplinary boundaries. All the historical figures discussed are philosophers, though this may be because psychology is a comparatively new discipline. It will thus probably also be of greater interest to philosophers than to psychologists. Even so, the editors explicitly commit themselves to a cross-disciplinary approach, pointing out that many philosophers now collaborate on interdisciplinary research. This focus on interdisciplinary research grounds the collection in interesting empirical data that have bearing on moral and metaethical questions. This makes it both more interesting and more relevant, given the fact that moral theorizing needs to be sensitive to the actual psychology of moral agents.

Handbooks are a strange genre, as it's never quite clear whether the target audience are people new to the subject matter or those who are already experts aiming to learn more about different subareas of their field. The Oxford Handbook of Moral Psychology has plenty of material for either demographic, with some contributions more obviously introductory, such as the section on Confucius; while others are more suitable to a specialised audience, for example the discussion of the role of sentiments in David Hume and Adam Smith's theories of moral judgment by Rachel Cohon. However, Cohon does draw out implications from her analysis that apply more broadly to philosophers interested in giving a sentimental account of moral judgement today. I found much to fascinate and educate me in the volume, and the same will be true of other readers with an interest in moral psychology. This handbook provides a valuable resource that the reader can dip into time and again.

Another, related, dimension on which contributions vary is the extent to which the work aims to make a new contribution to the field or merely summarises existing research. Importantly for experts in the field, many articles put forward new arguments that someone working in the area needs to

engage with. For example, Brink and Nelkin's paper on the nature and significance of blame puts forward a new *core and syndrome* account of blame. The paper clearly positions itself in an ongoing lively debate on how best to characterize the nature of blame. According to Brink and Nelkin's proposal, blame essentially involves 'an aversive attitude that is predicated on the belief that the target is blameworthy' (p.178). Different manifestations of blame may involve different further features, such as calling someone out, but according to Brink and Nelkin, all forms of blame have this core feature. The proposal they put forward will be a reference point people working in that area need to take into account, whether or not they agree with it.

The book is organized into three parts, the first one is historical, the second is entitled 'Foundations', the third part looks at applications. The historical section is the shortest one, with seven chapters. It includes a chapter on Confucianism, which I found informative as a complete novice. Other philosophers covered are Plato, Aristotle, Aquinas, Hume and Adam Smith, Kant and Nietzsche. While I hesitate to suggest a further chapter for a book that already has 50 chapters, it would have been nice to include a chapter on a psychologist, or the history of research on moral reasoning and decision making in psychology.

Part 2, *Foundations* deals with perennial and general problems of moral psychology such as moral judgement internalism, the nature of moral reasons, virtue, moral expertise, weakness of the will, self-deception, various topics related to moral responsibility, moral learning, moral nativism, the role of neuroscience in ethics, the evolution of moral psychology but also the moral psychology of animals, the moral psychology of humour, and personal identity. While this last topic might be considered a more fitting topic for a volume on metaphysics, it has clear implication for normative issues such as ongoing obligations that hold between people.

The third section, *Applications*, turns the spotlight onto more specific contexts and problems, such as moral agency in mental disorder, implicit bias, loss of control in addiction, or the relationship between *mens rea* in the criminal law and in moral judgment. The spread of topics is very broad, but one can make out thematic clusters, including one on issues concerning sex, love and marriage, and another one on oppression, victimization, as well as a chapter on how poverty can shape and influence decision making. It is good to see this involvement with the way real-world conditions and the position individuals find themselves in shape reasoning and decision-making processes, especially in a discipline that has historically tended to abstract away from the way human actions are shaped by circumstance.

While there are numerous connections between the different contributions, they are all stand-alone articles that can be read without having read any other part of the book. It is impossible to discuss all fifty contributions to the volume. Instead, I will aim to provide an impression of how the interaction between empirical research and philosophical analysis plays out in different chapters on very different topics.

In their chapter 'Prudential Psychology: Theory, Method and Measurement' Valerie Tiberius and Daniel Haybron give a programmatic overview of different ways philosophical theory and empirical research could interact. The tone of the paper is somewhat didactic, they stress the importance of avoiding cherry-picking of empirical evidence. One might think that in a handbook of this kind, they are preaching to the choir, though the temptation to cherry-pick one's evidence is of course ever-present, both for philosophers and for psychologists. The chapter provides a framework for doing empirically informed philosophy and interdisciplinary research, going through a number of possible ways in which empirical research can inform philosophical theories of what well-being is (high level theories), or how existing measures in different areas of psychology can be recruited for well-being

research. One example the authors cite is using mental health measures from depression, anxiety and stress in looking at emotional well-being. This chapter is very much a 'how to' chapter in style, not one that pushes a certain philosophical and psychological theory.

A chapter on the relevance of neuroscience for ethics by Adina Roskies is similarly broad in scope, outlining work done by her and others on the question whether findings from neuroscience are well suited to adjudicate between different theories of normative ethics – for example deontology and utilitarianism - or metaethical positions such as motivational internalism and externalism. She argues convincingly that data from neuroscience cannot replace substantial ethical assumptions, though they should inform and constrain them. She also shows that finding out more about the brain can change our moral judgments by telling us more about the problem under consideration. So, for example, finding out that some people in permanent vegetative state are still able to process verbal instructions and perform mental tasks tells us something about their level of consciousness. On the assumption that consciousness matters to moral status, this has repercussion for the moral status of some people in a persistent vegetative state. This chapter exemplifies the kind of contribution that makes for an excellent teaching resource or overview for the novice, but will not offer much that is new to those that have been following the debate. One helpful feature of this specific chapter is that it cross-references related contributions in the volume. It would have been nice to see this kind of cross-referencing more consistently throughout the volume.

Other chapters give an overview of relevant research but also explicitly use empirical research to argue for a specific philosophical claim. One particularly engaging example of this is Susana Monso and Kristin Andrews' chapter on Animal Moral Psychologies. They present strong evidence that animals have some capacities relevant for moral judgment and action, namely care, autonomy and normativity. They provide an extended and fascinating discussion of the evidence that animals care for other animals, which is strongly reminiscent of the literature on psychological egoism, as there is overlap in the way experimenters try to disentangle altruistic (caring) and egoistic motivation. Thus, there is a discussion of whether rats help other rats which are in distress because they are genuinely concerned for their wellbeing or merely because they find the cries of distress of their fellow rats aversive. Studies found that rats also helped in situations where they could escape the distressing sounds of other rats in trouble, but that it took them longer to do so. I found this contribution particularly fascinating as it contained much that was new to me, but also linked to long-standing debates in moral psychology. Having contributions that are slightly off the beaten path for many people interested in moral psychology is a real strength of this volume.

Another example of work that draws on empirical research are Emily McTernan's chapter Moral Character, Liberal States and Civic Education which considers the threat that situationism poses to character education and the role alternative strategies such as nudging and shaping people's choice environment play. Empirical research on the contribution of individual traits and external factors to people's behaviour is used to inform thinking on how civic education should proceed.

Chandra Sripada's paper on loss of control in addiction identifies gaps in empirical accounts of addiction which are relevant for the normative question whether addicts are partially or wholly excused for their actions because they cannot control their addictive choices. He argues that the two prominent models of addiction, the choice model and the disease model, both fall short in accounting for some of the data regarding how addiction influences free decision making and don't provide a mechanistic account of how exactly top-down control of desires is impacted in addiction. The choice model emphasises the extent to which even addicts are sensitive to incentives other than their addictive desires and able to stop or at least defer using when suitably incentivized, but neglects the fact that this is also the case in conditions where we do think choice is impaired, such as

Obsessive Compulsive Disorder or Tourette's. The disease model recognises that addictive desires are extremely hard to resist but fails to explain why. Sripada concludes that we need to acknowledge our ignorance and strive for a better understanding of the mechanisms of addictive desire. This paper is thus a call for action for empirical scientists as well, pointing out gaps in empirical knowledge.

There are many more chapters that deserve more detailed discussion, but I hope to have provided a flavour of the very broad range of topics covered and different approaches to moral psychology pursued by different contributors. I look forward to dipping into the Oxford Handbook of Moral Psychology in the coming months and years, and would recommend it as a valuable resource for experts in the field and those interested in finding out more.