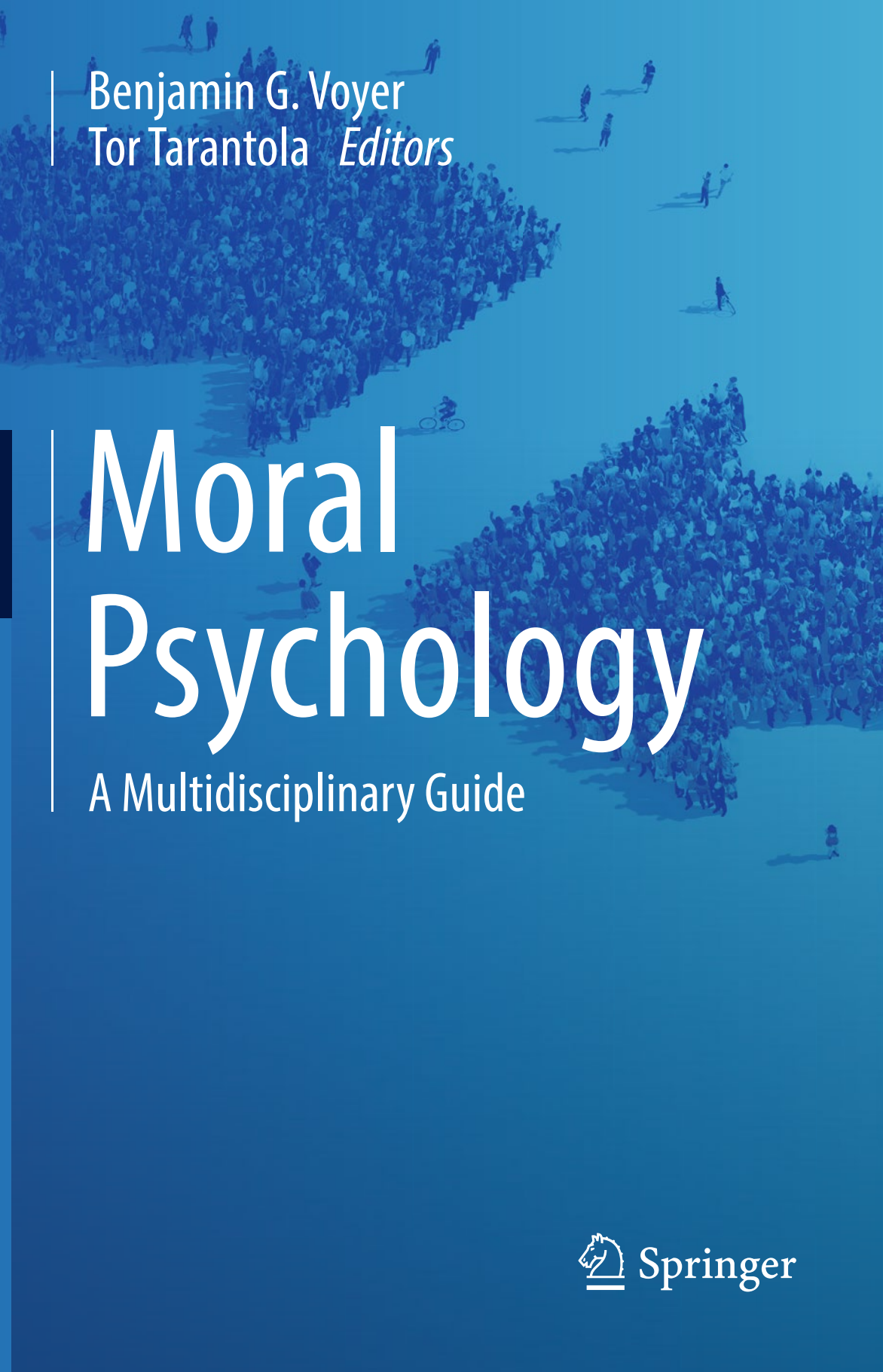
Benjamin G. Voyer
Tor Tarantola Editors
Moral Psychology
A Multidisciplinary Guide
Springer

Moral Psychology

Benjamin G. Voyer • Tor Tarantola
Editors

Moral Psychology

A Multidisciplinary Guide

Springer

Editors
Benjamin G. Voyer
Department of Marketing
ESCP Europe
London, UK

Tor Tarantola
Department of Psychology
University of Cambridge
Cambridge, UK

ISBN 978-3-319-61847-0 ISBN 978-3-319-61849-4 (eBook)
DOI 10.1007/978-3-319-61849-4

Library of Congress Control Number: 2017953195

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer International Publishing AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Acknowledgements

The editors would like to thank Bradley Franks, Jennifer Sheehy-Skeffington, and members of the Department of Psychological and Behavioural Science at the London School of Economics and Political Science. Thank you to our parents.

Contents

Toward a Multidisciplinary Moral Psychology 1
Benjamin G. Voyer and Tor Tarantola

Between Facts and Norms: Ethics and Empirical Moral Psychology 5
Hanno Sauer

An Evolutionarily Informed Study of Moral Psychology 29
Max M. Krasnow

Moral Psychology: An Anthropological Perspective 43
Paolo Heywood

Cognitive and Neural Sciences: Investigating the Moral System 59
Tor Tarantola

(Im)Morality in Political Discourse?: The Effects of Moral Psychology in Politics 81
Nicholas P. Nicoletti and William K. Delehanty

An Open Letter to Our Students: Doing Interdisciplinary Moral Psychology 119
Edouard Machery and John M. Doris

Current Perspectives in Moral Psychology 143
Frans de Waal, Hanno Sauer, Paolo Heywood, Verena Wieser, Edouard Machery, and John M. Doris

Index 163

Contributors

William K. Delehanty Department of Social Sciences, Missouri Southern State University (MSSU), Joplin, MO, USA

John M. Doris Philosophy-Neuroscience-Psychology Program & Philosophy Department, Washington University in St. Louis, St. Louis, MO, USA

Paolo Heywood Division of Social Anthropology, University of Cambridge, Cambridge, UK

Max M. Krasnow Evolutionary Psychology Laboratory, Department of Psychology, Harvard University, Cambridge, MA, USA

Edouard Machery Center for Philosophy of Science, University of Pittsburgh, Pittsburgh, PA, USA

Nicholas P. Nicoletti Department of Social Sciences, Missouri Southern State University (MSSU), Joplin, MO, USA

Hanno Sauer Department of Philosophy and Religious Studies, Utrecht University, Utrecht, The Netherlands

Tor Tarantola Department of Psychology, University of Cambridge, Cambridge, UK

Benjamin G. Voyer Department of Marketing, ESCP Europe, London, UK

Frans de Waal Living Links Center, Emory University, Atlanta, GA, USA

Verena Wieser Department of Strategic Management, Marketing and Tourism, University of Innsbruck School of Management, Innsbruck, Austria

About the Authors

Benjamin G. Voyer is a behavioural scientist and interdisciplinary researcher, working with both quantitative and qualitative research methods to investigate how self-perception and interpersonal relations affect cognition and behaviour in various contexts (consumption, organisational, cross-cultural, etc.). He has authored or co-authored more than 150 scientific contributions to the field of applied psychology (journal articles, conference presentations, case studies, etc.). He is currently L'Oréal Professor of Creativity & Marketing at ESCP Europe Business School and Visiting Fellow at the London School of Economics and Political Science (LSE) in the UK.

Tor Tarantola completed his Ph.D. in psychology at the University of Cambridge and is currently a J.D. candidate at Yale Law School. His research focuses on reinforcement learning, social cognition, and their implications for legal theory and practice. He was formerly a fiscal and policy analyst at the non-partisan Legislative Analyst's Office in California, where he helped advise state lawmakers on criminal justice policy.

Toward a Multidisciplinary Moral Psychology

Benjamin G. Voyer and Tor Tarantola

Humans idolize, condemn, suffer, rejoice, kill, and die in the name of morality. We are a hypersocial species—nearly everything we do involves other people. So it's not surprising that the rules governing our interactions—what we owe to one another and how we ought to treat transgressors—occupy a prime spot in the human psyche. Moral psychology tries to understand how these rules come about, how we perceive and act on them, and how we respond when people violate them.

Despite its name, moral psychology spans every discipline concerned with human interaction—from philosophy and sociology to evolutionary biology and neuroscience. The applications of this research have been significant, helping us better understand the biases of juries, the culpability of minors and the mentally ill, and the public's views on the purposes of criminal and civil sanctions. It also has the potential to guide how governments implement justice policies and to allow us to better negotiate competing norms across cultures.

As the varied disciplines comprising moral psychology have developed, their methods have become increasingly specialized and complex. For example, cognitive scientists have used computational modeling to break new ground in understanding human interaction, but these methods often elude researchers in less mathematical fields. Similarly, academic philosophy often uses frameworks and vocabularies in which scientists are not conversant. As a result, the academy has become increasingly siloed, meaning that significant advances in one discipline can go largely unnoticed in others.

B.G. Voyer (✉)
Department of Marketing, ESCP Europe, London, UK
e-mail: bvoyer@escpeurope.eu

T. Tarantola
Department of Psychology, University of Cambridge, Cambridge, UK
e-mail: tor.tarantola@gmail.com

B.G. Voyer, T. Tarantola (eds.), *Moral Psychology*,
DOI 10.1007/978-3-319-61849-4_1

Our aim in this volume is to help break through the siloes by offering a panoramic view of the trends, methods, and questions that dominate moral psychology research in different disciplines. The breadth and complexity of these questions—within each discipline, let alone across several—preclude us from being comprehensive. Rather, we hope to offer researchers of different backgrounds a taste of the diverse perspectives from which these questions are approached and to show how greater collaboration across disciplines can help advance the field. We also hope this volume will be useful to practitioners—lawyers, clinicians, policymakers, and others—whose work centers on the intersection of moral codes and human behavior.

Philosopher Hanno Sauer begins the volume by tracing the history of moral psychology from David Hume to modern empirical philosophy. For Sauer, the progression of philosophical thinking about morality and the tensions between empirical and normative questions—asking "what is" versus asking "what ought to be"—provide important context for current research. He focuses on what he calls the "gap" between facts and norms and outlines how key debates in philosophy—about moral relativism, free will, leading a virtuous life, and others—can be informed by empirical work.

In the second chapter, evolutionary psychologist Max Krasnow outlines a theory for how the fundamentals of moral behavior—altruism and punishment—evolved as biologically beneficial predispositions. He draws a useful distinction between mechanisms directed at regulating one's own behavior (what he calls "inward-facing mechanisms") and those directed at changing others' behavior ("outward-facing mechanisms"). Consideration of each of these types of mechanisms is necessary, he argues, in order to fully appreciate the evolutionary dynamics that shaped the core of human morality. His theory shows why the empirical and theoretical study of morality can be so difficult—because the ultimate reasons behind a moral decision might often be unknown even to the person who makes it.

In the third chapter, social anthropologist Paolo Heywood offers a critical perspective on the traditional approaches to studying ethics and morality in his discipline. He surveys recent theoretical and ethnographic work, presents the framework for a modern "anthropology of ethics," and tackles a central question facing anthropologists who study moral psychology: how can moral behavior be both universal and culturally specific?

In the fourth chapter, cognitive scientist Tor Tarantola surveys the varied research that has contributed to an emerging cognitive science of morality. He outlines the major advances at three levels of what he calls the "moral system"—the group level, which studies how norms emerge and evolve; the interactive level, which studies the dynamics of interpersonal interactions; and the individual level, which studies the individual cognitive and neural processes that underlie perceptions and behavior. This chapter draws on work from varied fields—from sociology, which has generated important theories about the emergence of norms; to experimental psychology and behavioral economics, which have begun to quantify the factors important to moral judgments; to cognitive neuroscience, which investigates how the brain implements the perceptual and cognitive tasks that underlie moral behaviors and experiences.

Tarantola argues for a more unified cognitive science of morality that uses a mathematical systems approach to integrate insights from each of these three levels.

In the fifth chapter, political scientists Nicholas Nicoletti and William Delehanty connect moral psychology to contemporary politics. In discussing the role of moral judgment in public opinion and political attitude formation, they advance the claim that strong moral commitments can lead to polarization and ultimately hurt political discourse.

In the sixth chapter, philosophers Edouard Machery and John Doris offer some recommendations and guiding principles for conducting interdisciplinary moral psychology research, which will be useful to students and experienced scholars alike. This chapter will be especially helpful to scholars in the humanities looking to apply scientific methods to their work. Drawing from their experience at the intersection of philosophy and the cognitive and neural sciences, Machery and Doris offer valuable advice for how to critically evaluate scientific literature and avoid common pitfalls.

In the final chapter, a number of our contributors and other scholars offer their thoughts on the future of moral psychology. We ask them to comment on the role of empiricism in philosophy, whether moral psychology can help answer moral questions, and the importance and difficulties of being interdisciplinary when doing moral psychology research. They also discuss the implications of this research for public policy and the law.

Our goal in preparing this volume is to begin to synthesize a more multidisciplinary moral psychology. The ideas presented here, while not comprehensive, begin to weave together threads of research that too often progress in isolation. We hope this collection will help catalyze new and creative ways of approaching these fascinating, important, and ancient questions.

Between Facts and Norms: Ethics and Empirical Moral Psychology

Hanno Sauer

A Cold, Hard Look

For most of its history, philosophical moral psychology has been in bad shape. People were asking the right questions, but their methods were questionable: rampant speculation was revised in light of pure guesswork; guesswork had to be amended on the account of arbitrary superstition; superstition was corrected by flimsy moralizing, and the whole thing was rounded off by a healthy dose of wishful thinking. Philosophical theories of human nature had to state how human beings ought to be, rather than how they actually are.

It is not a good idea, generally speaking, to speculate about the nature of the moral mind without systematically investigating how the mind works. Why philosophers failed to appreciate this rather obvious truth is something I can only speculate about myself. The—arguably false—idea that the mind is transparent to itself, and can thus be studied without external aid, may have played a role. We now know that this type of self-transparency is an illusion and that expecting the mind to give honest answers when examined by introspection alone is hopelessly naive.

Perhaps I exaggerate, and it wasn't quite as bad. To find out how moral agents think and act, some philosophers like Aristotle, Hume, or Kant did consult the best science of their time. Then again, this did not necessarily amount to much. Others—Nietzsche comes to mind (Knobe and Leiter 2007)—were in fact pioneers and gave the field of empirically informed moral psychology, most of which was yet to emerge at the time, new directions to pursue, and new questions to address. Yet all too often, philosophers "have been content to invent their psychology […] from scratch" (Darwall et al. 1992, 189). A "cold, hard look at what is known about human nature" (Flanagan 1991, 15) seems to me to be the best cure for this affliction.

H. Sauer (✉)
Department of Philosophy and Religious Studies, Utrecht University, Utrecht, The Netherlands
e-mail: h.c.sauer@uu.nl

B.G. Voyer, T. Tarantola (eds.), *Moral Psychology*,
DOI 10.1007/978-3-319-61849-4_2

The main tension between philosophical and empirical accounts of human moral judgment and agency comes down to the fact that, at the end of the day, philosophers are interested in moral psychology for one thing, and one thing only (I exaggerate again). They want to know what facts about the *psychological* foundations of morality can teach us about the foundations of morality, *period*: how facts about human nature bear on right and wrong, good and bad, and just and unjust. This tension is further aggravated by the fact that many philosophers deem this to be a hopeless endeavor that is doomed to fail from the outset. The problem, these philosophers argue, is that there is no way (no legitimate and informative one, at any rate) to get from an *is* to an *ought*. Rumor has it that facts are different from values. Descriptive statements, it is said, do not entail prescriptive propositions. Empirical information, the story goes, has no normative significance. Nature allegedly has no moral import.

In what follows, I will refer to this problem as *the gap*. In the first section of this chapter, I will briefly explain what the gap is, why it is said to exist, and to what extent it is supposed to pose an obstacle to empirically informed theorizing about ethics.

In the second section, I will take a look at some of the most interesting recent developments in empirical moral psychology and explain what their normative implications are supposed to be. My selection of topics will be somewhat arbitrary, and the discussion I provide by no means is comprehensive. I am not attempting to give an overview of the whole field of contemporary moral psychology. This has already been done elsewhere, by people more qualified to do this than myself (see Doris and Stich 2005; Rini 2015; Kumar forthcoming; Alfano and Loeb 2014; Alfano 2016; Appiah 2008; Tiberius 2014, and the remainder of this book). Instead, I choose a more focused approach and look at the whole field from the perspective of what I take to be the main issue of philosophical interest: my aim is to illustrate how empirical moral psychology might be brought to bear on issues of normative significance—what the virtues are, what makes for a good life, whether free will exists, what role luck plays in morality, what constitutes an action, what it means to be a person, how people arrive at moral judgments, whether these judgments are relative, and whether we are at all competent to make them. My discussion will be arranged around four clusters: normative theory, moral agency, moral and nonmoral judgment, and moral intuition.

In the final section, I will extract some lessons from this discussion. Are the skeptics right? When it comes to figuring out what demands morality makes on us, does empirical information remain thoroughly irrelevant? Or are there grounds for optimism? Do empirically informed ethics have a future after all? I will argue that the normative significance of empirical studies of human moral cognition and behavior, though always indirect, comes in essentially three forms: by debunking the processes on the basis of which we make moral judgments and develop moral concepts; by undermining the empirical presuppositions of some normative theories, vindicating those of others; and by providing tools for the reflective improvement of moral judgment and agency by bringing to light the sometimes egregious mistakes that escape our powers of introspection and the empirically unaided mind.

The Gap

In philosophy, skepticism about the relevance of empirical facts for so-called normative questions—questions about right and wrong, permissible and forbidden, and virtue and vice—can draw on two *loci classici*. One can be found in the third part of David Hume's *Treatise of Human Nature*, where he complains that

> "[i]n every system of morality, which I have hitherto met with, I have always remarked, that the author proceeds for some time in the ordinary way of reasoning, and establishes the being of a God, or makes observations concerning human affairs; when of a sudden I am surpriz'd to find, that instead of the usual copulations of propositions, is, and is not, I meet with no proposition that is not connected with an ought, or an ought not" (Hume 1739/2000, III.I.I)

Hume argued that this transition was as widespread as it was illegitimate; for in his view, and the view of many others, there is no logically valid way to derive a proposition with normative content (it is not ok to lie; drone surveillance is reprehensible; chastity is a virtue; we have a duty to help others, and when doing so, it involves little cost to ourselves) from a set of premises with purely descriptive, factual content (people lie all the time; drones are really useful; your father wants you to be chaste; helping others will make people like you). An inference is logically valid just in case the truth of its premises guarantees the truth of its conclusion. No such inference, Hume thought, could ever take you from an *is* to an *ought*.

The second go-to place for friends and foes of *the gap* is G. E. Moore's (1903) *Principia Ethica*. Here, Moore coined the term "naturalistic fallacy" (Moore 1903) to refer to attempts to identify the property of being *good* with any natural property, such as being *useful*, or *maximizing pleasure*, or being *economically efficient*, or being *sanctioned by the state*. Moore's point was that *good* and *bad* cannot be defined in natural terms, because if they could, then whenever we had found some action or event instantiating the natural property picked out by our definition (given that said definition is correct), the question whether the action or event is also good would necessarily be *closed* to anyone but the conceptually confused. Centaurs, and only centaurs, are creatures with an anthropic upper and hippic lower half; if I manage to show you such a thing, the question whether it is also a centaur is *closed*. Now Moore argued that for every proposed natural definition of the good—say "the good which maximizes pleasure"—it always remains possible to ask whether something instantiating the natural property specified in the definiendum is also good. "It maximizes pleasure, but is it also good," or "it is loved by the Gods, but is it also good," or "it is useful for society, but is it also good," and so on. These questions all make sense, and the property of being good cannot be conceptually reduced to other natural properties. This is Moore's famous "open question argument."

The naturalistic fallacy is not strictly speaking a fallacy, and as we have seen, the term was originally supposed to refer not to *the gap*, but an entirely different, semantic point. Then again, people love to accuse one another of fallacious reasoning, and the term is catchy, so "naturalistic fallacy" stuck around and is now widely

used for illicit attempts to bridge *the gap*. Examples for naturalistic fallacies are ridiculously easy to find and are especially common in debates on evolutionary psychology, sexual morality, and most other topics in applied ethics. I will not cite any sources here, as the research would have been too depressing. But I *can* give a few examples of the kind of reasoning I have in mind and which we are all too well acquainted with: evolution favors the selfish and competitive, so that is how we, too, ought to act; homosexuality is unnatural and should thus be banned; humans are the only animals with the power to reason, so the rational life is best for humans; people have always killed animals for food, and women were always discriminated against, so clearly there is nothing wrong with those things. Never mind whether these inferences get the facts right or not—because even if they did, they would fail to establish their conclusion on account of *the gap*.

On the other hand, it seems hard to see how empirical facts could *always* remain *thoroughly* irrelevant to normative inquiry. Whether or not abortion is permissible, and under what conditions, it will surely depend on what kind of being a fetus is and whether it can feel pain or has interests and conscious experiences. Likewise, my indignation toward the man I believe my wife cheated on me with, and which I am about to punch in the face, will readily switch its target once I have found out that *this* man isn't the culprit, but the pathetic scoundrel standing next to him. What should be done about climate change, or whether anything should be done at all, cannot be assessed without factual knowledge. And whether or not you should perform that tracheotomy to save your suffocating friend will depend on how likely it is that you will succeed. In all these cases, empirical facts have bearing on issues of normative significance, if only via the nonmoral facts upon which moral facts are grounded.

Moreover, many normative moral theories seem to make rather straightforward assumptions about what kinds of agents we are, assumptions which are far from empirically innocent. For instance, some Kantians argue that moral norms are prescriptive rules whose authority does not depend on whether or not one is already motivated to conform to them: these rules are supposed to be motivating *independently* of an agent's desires and goals, simply by virtue of the fact that they specify what it means to be an agent (Korsgaard 1996; Velleman 2011). But what if this paints an unrealistic picture of how motivation works and of what constitutes an agent? Virtue ethicists often claim that a good person is a person with a coherent set of laudable character traits (Hursthouse 1999; Foot 2001). Does this account rely on an erroneous idea of how people function and how well their personalities are integrated? Some consequentialists hold that the right action—the one we ought to choose—is the unique action that has the best consequences. But what if figuring out which action is beyond human deliberative powers (Mason 2013)? In all these cases, normative theories make empirical presuppositions.

The question, then, is this: despite the fact that no ought ever follows from an is and despite the fact that the concept of the good cannot be identified with any empirical property, how should we understand the *normative relevance of empirical facts* in light of the *empirical presuppositions of normative theories*?

Normative Theory

(i) *Consequentialism and Deontology*. Contemporary normative ethics is organized around a distinction that manages at the same time to be one of the least well liked and yet one of the most popular in all of philosophy: the distinction between *consequentialism* and *deontology*. Consequentialist moral theories hold that the rightness or wrongness of an action is determined *only* by its (actual or expected) consequences. Deontological moral theories deny this. Some deontologists hold that intentions matter for the moral evaluation of an action as well, others argue that there are certain side constraints (such as individual rights) on the maximization of the good, and that it can make a moral difference whether one actively does something or merely allows it to happen or whether someone uses someone else as a mere means to an end rather than an end in his/herself. There is plenty of evidence that on an intuitive level, people take deontological considerations to be morally relevant (Young et al. 2007). Often, their judgments conform to deontological rules such as the doctrine of double effect (according to which harming someone can be permissible when it is an unintended but foreseen side effect, rather than when the harm is directly intended, Mikhail 2007, Kamm 2007), even though such slightly more sophisticated principles may remain ineffable.

What about *the gap*? Can empirical data shed light on which theory is correct? One way to model the difference between consequentialism and deontology is to look at sacrificial dilemmas involving urgent trade-offs between harming an individual person and promoting the greater good and to see which conflicting actions consequentialism and deontology classify as right and wrong, respectively, when doing what's best overall clashes with certain intuitively plausible moral rules. Moral emergencies (Appiah 2008, 96ff.) of this sort form the basis of what is perhaps the single most thriving and controversial research program in normatively oriented empirical moral psychology: Joshua Greene's *dual process* model of moral cognition (Greene 2014). According to this model, cognitive science can show that one of the two normative theories is superior to the other. Consequentialism, the evidence is purported to show, engages more rational parts of the brain and more sophisticated types of processing than deontology, which is associated with more emotional parts of the brain and more crude forms of cognition (Greene 2001, 2004). When people judge it impermissible, for instance, to kill one person to save five others (thereby endorsing the deontological option), they arrive at this judgment via a more emotional and less calculating route. Deontological moral theory, then, amounts to little more than post hoc rationalizations of those brute, alarm-like responses (Greene 2008; see chapter "Cognitive and Neural Sciences: Investigating the Moral System" for a more thorough discussion of the neuroscience of moral judgment).

The dual process model's main normative upshot is supposed to be a vindication of consequentialist and a debunking of deontological intuitions on the basis of empirical evidence regarding the cognitive processes that produce these two types of moral intuitions. But it remains unclear whether the way people arrive at their

consequentialist responses deserves to be described as consequentialist reasoning at all, rather than an ordinary weighing of competing considerations for and against a proposed action (Kahane 2012). Even worse, the consequentialist judgments that some people end up endorsing do not seem to be based on an impartial concern for the greater good, but on much more sinister dispositions (Kahane et al. 2015). Perhaps most importantly, the connection between consequentialist judgments and controlled, System II processing on the one hand, and deontological judgments and automatic, System I processing on the other hand (Evans 2008; Stanovich 2011; Kahneman 2011), seems to be due to the fact that in Greene's original studies, the consequentialist option always *happened to be* the counterintuitive one. When this confound is removed and counterintuitive deontological options are included, the pattern is reversed (Kahane et al. 2012; cf. Greene et al. 2014).

Dual process theory continues to be haunted by *the gap*. Empirical data on which type of process, or which brain region, is involved in the production of a moral judgment tells us very little about whether or not this judgment is justified or not—unless we *already* know which processes are unreliable and which aren't, which we arguably do not. Now the dual process model's two best shots are an *argument from morally irrelevant factors* and an *argument from obsoleteness*. Firstly, it could be shown that regardless of whether people arrive at them through emotion or reasoning, deontological intuitions pick up on *morally irrelevant factors*, such as whether an act of harming someone has been brought about in a distal or proximal way. Such sensitivity to morally extraneous features is often sufficient to indict a particular type of judgment as unreliable. Secondly, one could argue that some moral intuitions are generated on the basis of processes which are unlikely to deliver correct results under conditions they have neither evolved nor have been culturally shaped in (Singer 2005). For instance, moral cognition may be good at dealing with how to secure cooperation in stable small-scale communities. Dynamic, large-scale societies comprised of strangers and organized on the basis of complex economic and political institutions may constitute a hostile environment for the cognitive processes our ancestors bequeathed to us. Since a similar story may be true of the processes responsible for deontological intuitions and the conditions we currently inhabit, this, too, could help undermine the credibility of those intuitions via those processes (Nichols 2014).

The problem with these arguments, however, is that it is far from clear which role empirical evidence has to play in them at all, and whether most or all of the normative heavy lifting isn't done by armchair theorizing about what does and what doesn't count as morally relevant—which is to say, by moral philosophy (Berker 2009; Sauer 2012b). As for the second point, it has to be emphasized that the primitive cognitive processes of modern social conditions with their dynamic, large, anonymous societies and complex technological challenges do not exclusively deliver deontological intuitions. Conversely, the cognitive processes that are required to successfully navigate such conditions are not exclusively consequentialist in nature. As far as the consequentialism/deontology distinction is concerned, dual process theory is thus neither here nor there. What remains of its steep ambitions may simply be that some moral judgments are produced by

automatic and some by controlled cognitive processes, together with the claim that under certain conditions, the former are less likely to produce correct responses than the latter.

(ii) *Moral Relativism*. But why speculate about the correct normative theory, when it is far from clear whether moral problems have correct solutions at all? Isn't it clear that people have widely diverging and irreconcilable views about what morality requires? One does not need empirical research to confirm that people disagree about morality. The so-called argument from disagreement (Brink 1984; Mackie 1977) is supposed to use this fact of life to make the case for moral relativism, the view that there is no single true morality and that moral norms and values are only ever valid relative to some individual, social, or cultural context.

The problem with this argument is that there is a rather obvious objection to it. Disagreement does not entail relativity (Enoch 2009): people disagree about all kinds of things, but this doesn't mean that there are no facts of the matter about which one side is right and the other wrong. Non-relativists like to point out that what is needed for the argument from disagreement to get off the ground is a case of intractable *fundamental* moral disagreement—disagreement that would persist even under ideal conditions of full information and flawless reasoning on the part of those disagreeing. Non-fundamental disagreement, the kind that is purportedly not damaging to moral universalists, is disagreement for which a so-called defusing explanation can be given. Such disagreement can be due, among other things, to disagreement about the underlying facts, special pleading, or irrationality. Special pleading occurs when people refuse to apply a value consistently, trying to make an exception for themselves (e.g., endorsing the death penalty except when oneself is to be executed); irrationality can occur when people fail to appreciate what their values entail (e.g., wanting to reduce the suffering of sentient beings, but not adjusting one's diet in light of this goal).

What can empirical data contribute to this debate? Recently, Doris and Plakias (2008) have tried to revive the argument from disagreement by bringing evidence from cultural psychology to bear on the issue of whether it is possible to identify a case of fundamental moral disagreement for which no defusing explanation seems to be available. For instance, Doris and Plakias draw heavily on Nisbett and Cohen's (1996) "culture of honor" explanation for differences in attitudes toward violence between people from the American North and South. Evidence from criminal statistics, legal decisions, lab experiments, and field studies all point in the direction that Southerners are both more prone to violence and more tolerant of it. Nisbett and Cohen attribute this tendency, which is restricted to violence in response to threats, insults, and other violations of *honor*, to the reputational demands of herding economies. In contrast to economies based on farming or trade, a herding economy is a high-stakes environment in which a person's entire assets could be stolen, which made it necessary for individuals to convey that they would be willing to respond violently to threats. Others (Fraser and Hauser 2010) have argued that some cultures (e.g., rural Mayans) do not see a morally relevant difference between acts and omissions, which is another promising candidate for a fundamental moral disagreement.

Does this type of argument succeed in bridging *the gap*? Doris and Plakias argue that none of the aforementioned defusing explanations plausibly account for differences in Southern and Northern attitudes toward violence. If true, this would support their case for moral relativism. However, there are reasons for doubt. To a large extent, cross-cultural agreement about certain general prima facie duties is compatible with seemingly dramatic disagreement about all-things-considered obligations (Meyers 2013). Many disagreements concern *how* wrong or right something is and do not involve one party thinking that something is completely wrong which the other thinks is completely innocuous. That Southerners behave more violently and are more likely to condone violence does not mean that they take it to be more permissible (Leiter 2007). Moreover, most disagreements vanish under close scrutiny: when they are subjected to the sort of inquiry moral universalists favor, moral disputes tend to disappear (hint: less rural/more formally educated Mayans *do* see a difference between doing and allowing). The disagreements Doris and Plakias base their argument on can be located at the level of unreflective System I responses, where they inflict hardly any damage on non-relativists (Fitzpatrick 2014). If Southerners were informed about Nisbett and Wilsons's "culture of honor" explanation *itself*, and thus about the fact that the original economic rationale for their attitudes no longer obtains, they may well be inclined to change those attitudes (Sneddon 2009). This sort of genealogical defeater is demonstrably effective (Paxton et al. 2012). The issue of moral relativism thus can be addressed empirically, at least as long as its defenders and opponents are willing to make clear predictions on how much convergence or divergence in people's moral views, and of what sort, to expect if their respective positions are true.

Moral Agency

(iii) *Character, Situation, and Virtue*. The fourth main player in normative ethics besides consequentialism, deontology, and moral relativism—*virtue ethics*—does not merely incur, as it were by accident, empirical presuppositions regarding what kinds of agents we are. Rather, its normative criteria are straightforwardly built upon an account of agency, thereby rendering it particularly hostage to empirical fortune. The rightness of an action does not, on this account, lie in the extent to which it satisfies some principled criterion of rightness. The right action, virtue ethicists argue, is the one the virtuous person would perform under the circumstances. The virtuous person is a person of good character, that is, an agent who possesses an assortment of praiseworthy traits such as honesty, courage, persistence, tranquility, magnanimity, and other quaint things.

It has long seemed fair to empirically minded philosophers (Harman 1999; Doris 2002) to ask whether this account of human agency is at all realistic. Perhaps unsurprisingly, they have been keen to show that it is not (Ross and Nisbett 1991; Doris 2009). The evidence—ranging from landmark experiments, such as Milgram's obedience studies, Zimbardo's prison experiment, and various studies on helping

behavior (Isen and Levin 1972; Darley and Batson 1973), to real-life atrocities such as the massacre of My Lai, the Rwandan genocide, or the violent and humiliating abuse of prisoners at Abu Ghraib (Doris and Murphy 2007)—consistently suggests that cross-situationally stable character traits of the kind postulated by virtue ethicists are nowhere to be found. The influence of frequently subtle and seemingly insubstantial situational features towers over that of internal dispositions.

However, even in this seemingly open and shut case in favor of situationism, the gap is not bridged without resistance. Some virtue ethicists have argued that character traits need to be construed differently (Kristjánsson 2012; Webber 2013), sought elsewhere (Merritt 2009), or that there is contrary evidence pointing toward the existence of virtues (Vranas 2005). Others chose to insist on the fact that the acquisition of virtues was always supposed to be a rare ideal, so that evidence for the rarity of virtuous agency cuts no ice (Miller 2003). Then again, few are comfortable defending unattainable ideals, and rightly so.

Among the more radical friends of situationism, some have suggested that we should abandon futile character education in favor of effective situation management (Harman 2009). Others have advocated a different form of moral technology that relies on the factitiousness of virtue: the nonexistence of global traits gives us no reason to abandon trait talk, which can function as a self-fulfilling prophecy. This suggests that we stop attributing only undesirable traits (Alfano 2013). Finally, some have argued that virtue ethics fails even if traits are real (Prinz 2009), because its normative authority rests upon an account of universal human nature that is debunked by cultural psychology.

(iv) *Freedom of the Will*. Virtue ethics is perhaps the clearest example of a normative theory that can be assessed in light of empirical facts. Other aspects of moral agency, such as freedom of the will, are harder to pin down; after all, many philosophers believe that free will just isn't the kind of thing that can be studied empirically.

The contemporary debate on the nature and existence of freedom of the will, perhaps one of the most mature in all of philosophy, cannot be adequately summarized here. Instead, I wish to mention two types of empirically supported challenges to free will and moral responsibility and to see what may follow from them normatively. One has to do with the *timing* of choice, the other with whether we have reason to believe conscious intentions ever really cause actions at all.

The first challenge, and arguably the more famous one, aims to show that people's conscious intentions do not initiate their actions (Libet 1985). In a series of experiments, Benjamin Libet could show that people's decision to execute a simple motor action is preceded, in the range of an average 350 ms, by a readiness potential (measured via EEG) initiating the action before people become aware of it. Other studies (Soon et al. 2008) report that it is possible to predict, with above chance accuracy, which of two simple actions an individual will perform up to 10 s before a subject's conscious decision. This makes it hard to see how conscious intentions could be responsible for action initiation.

According to the second challenge, a range of phenomena such as illusions of control, where people have the feeling of agency without any actual causal impact; episodes of confabulation, where people make up reasons for their actions that

couldn't possibly have played a motivating role; or certain pathological conditions such as utilization behavior or alien hand syndrome and, in general, the pervasive automaticity of human behavior (Bargh and Chartrand 1999) support the view that mental causation and the experience of conscious willing are illusory (Wegner 2002). In particular, people can have a sense of agency when their agency couldn't possibly have made a difference and are more than happy to come up with reasons for their actions that couldn't possibly have played a role in why they did what they did.

Both challenges are taken to suggest that our actions are determined by unconscious processes beyond our conscious awareness and control. I wish to remain agnostic about whether or not these challenges to free will are ultimately successful. But let me emphasize that the evidence also suggests that, at the very least, people retain a form of veto control over their actions (Schultze-Kraft et al. 2016). An unfree will may not be so hard to swallow if we at least have a free unwill.

Moreover, the Libet experiment (a) only concerns intentions *when* to perform a certain preselected action, and says nothing about decisions regarding *what* to do (however, see Haggard and Eimer 1999); (b) only investigates *proximal*, but crucially depends on the causal efficacy of *distal* intentions to follow the instructions of the experiment (Schlosser 2012a); and (c) presents only insignificant options which subjects have no good reasons to choose either way (Schlosser 2012b, 2014).

The normative problem of free will has two main aspects. One has to do with the consequences of people believing or disbelieving in free will. The other is about how we, individually and socially, should respond if free will turned out to be an illusion or to be much less free than we intuitively suppose. Firstly, people who have been primed to believe in determinism (which many, though importantly not all, hold to be incompatible with free will) are more likely to cheat on a subsequent task. Other studies suggest that disbelief in free will increases aggressiveness and reduces helping (Baumeister et al. 2009). On the other hand, a belief in free will need not have only desirable consequences, as it can make people more punitive and judgmental (Clark et al. 2014).

Secondly, and in line with the last point, the close tie between free will and moral responsibility entails that the nonexistence of free will has important ramifications for our social practice of punishment. To be sure, free will skepticism would leave three of the four functions of punishment—deterrence, protection, and rehabilitation—untouched, at least in principle. If free will does not exist, however, it may well turn out that all forms of *retributive* punishment are severely wrong (Zimmerman 2011). At the very least, it would open our punitive practices up for a sober empirical assessment in light of their consequences; drastically less harsh punishments, and perhaps even positive incentives to refrain from crime, are likely to be the upshot (Levy 2015). Retributive punishment has many undesirable consequences both for the punished and for society, which has to pay for expensive incarceration and deal with people who leave prison traumatized, stigmatized, and unemployable. When practices of punishment are assessed in light of their consequences rather than what wrongdoers allegedly deserve, these costs could be avoided.

Moral and Nonmoral Judgment

(v) *Personal Identity*. If situationists and free will skeptics are right, we are patchy puppets. Now what? Entities who are candidates for possessing free will or character traits are called *persons*. Persons, in turn, are the primary bearers of moral status: the coveted privilege of belonging to the circle of beings who enjoy special moral consideration in the form of rights and the dreaded burden of being the addressee of corresponding duties.

What does it take to be a person with an identity that remains stable over time? Either physical (the "stuff" people are made of) or psychological (people's "soul") continuity has been emphasized as the feature that decides what makes a person persist as one and the same (Martin and Barresi 2003). However, there is now a wealth of evidence suggesting that this is not how people think about personal identity.

Many concepts previously thought to be nonevaluative in character are actually downstream from people's moral assessments (the most famous perhaps being the concept of intentionality; more on this below). Personal identity is one such concept. For instance, people think that changes to a person's moral traits matter the most for whether a person stays the same or not (Strohminger and Nichols 2014). Moral judgments also influence how people think about what constitutes a person's true self, rather than more superficial aspects of their personality. First of all, people think that a person's core self is *fundamentally good* (Newman et al. 2015). This means that whether they take, say, an individual's inner dispositions or her explicit beliefs to constitute this core will depend on *their own* moral judgments: conservatives are more likely to think that a person's explicit beliefs form her true self when these beliefs display an aversion to homosexuality, but less likely to think so when those beliefs are pro-gay, and the other way around for a person's feelings of attraction. This leads to what is now sometimes referred to as the *Phineas Gage effect* (named after Phineas Gage, a nineteenth century railroad worker who allegedly underwent a drastic change of character after sustaining brain injury, Tobia 2015): changes for the better are seen as moves *toward* and changes for the worse as moves *away from* a person's true identity.

What is the normative relevance of this type of evidence? Of the many pressing moral issues for which personal identity is very important—how should we treat people's past wishes? what is the moral relevance of people who do not yet exist?—let me mention only one. A standard objection to utilitarianism has it that it licenses illicit trade-offs between people when aggregate welfare is maximized. As long as many can enjoy a life of leisure, it is palatable for a few to toil and drudge. But this, many think, ignores the essential separateness of persons: *inter*personal trade-offs, where a cost to one person is supposedly compensated by a larger benefit to another, should not be assimilated to *intra*personal trade-offs, where a cost incurred *now* can be outweighed by a later benefit to the same person. But if our intuitions about personal identity—the basic moral unit, as it were—are themselves shaped by moral intuitions, then our judgments about whom we are inclined to treat as a person at all, how to draw the lines between persons, and about the extent to which such lines carry moral weight may be deeply called into question.

(vi) *Intentionality*. Personal identity is only one of the domains where our thinking is influenced by moral considerations. In fact, some have suggested that the influence of moral judgments on the application of seemingly nonmoral concepts is pervasive: we are moralizers through and through (Pettit and Knobe 2009).

The most famous example is perhaps the concept of intentionality. Numerous studies confirm the basic asymmetric pattern: people are more likely to attribute intentionality for bad side effects than for good ones (Knobe 2003). When asked about whether the chairman of a company intentionally brought about a side effect to the environment, people are more likely to answer affirmatively when said side effect is bad rather than good. But why is this, when we tend to think that we need to establish intentionality first, to judge the morality of those intentional actions later?

And intentionality isn't the only concept people attribute asymmetrically when something of normative significance is at stake. Far from it, plenty of studies—on the doing/allowing distinction, the means/end distinction, knowledge, causality, free will, happiness, and many more (Cushman et al. 2008; Cova and Naar 2012; Beebe and Buckwalter 2010; Nichols and Knobe 2007; Phillips et al. 2011; Pettit and Knobe 2009; Knobe and Fraser 2008)—show that a host of other cognitive domains are susceptible to the same striking effect.

Knobe's surprising claim has long been that this influence of moral considerations on seemingly nonmoral issues is not a contaminating one where an otherwise value-neutral process is derailed, distorted, and illegitimately biased by people's moral beliefs (Knobe 2010; Sauer and Bates 2013). Rather, he has argued that moral judgments kick in at a deeper level, for instance, when setting the defaults against which intentionality and other psychological categories are assessed. In the case of the environment, the default is to be somewhat in favor of helping it; not caring about helping it at all, as the chairman is described in the original vignette, thus falls under this threshold. With respect to harming the environment, the default is to be against it; so in this case, not caring about harming it at all surpasses this threshold—hence the attribution of intentionality. Others have proposed that the aforementioned asymmetries are driven by judgments about norms more generally (Robinson et al. 2015; Holton 2010) or about people's so-called deep selves (Sripada and Konrath 2011).

Whatever the scope and substance of the correct explanation (Sauer 2014), the normative implications of the effect are potentially far-reaching and deeply revisionary. Outcomes which were brought about intentionally may not be worse than merely foreseen ones—worse outcomes would simply count as more intentional. Virtually all cases where intentionality is supposed to make a moral difference are affected by the asymmetry. Finally, the asymmetry may make it exquisitely difficult for jury members to accurately establish intentionality when immoral acts such as murder or rape are at issue (Nadelhoffer 2006). The very concepts we base our moral judgments upon may be suffused with morality from the outset. This would require us to reshape not just the way we think about a good deal of our practices, but those practices themselves.

(vii) *Moral Luck*. Other asymmetries are just as puzzling. A father whose children drown in the tub seems dramatically more blameworthy than one whose kids do not, even when both have exerted the same amount of care (or negligence) and one merely had good, the other bad luck. A drunk driver who happens to hit and injure someone is seen as a bad person, but millions of drunk drivers who simply had more luck are cut quite a bit of slack.

Moral luck is the degree to which luck affects the moral status of an action or person. The *problem* of moral luck, then, is how to reconcile the intuitive difference between lucky and unlucky fathers and drivers with the idea that people cannot be blame or praiseworthy for things beyond their control. Brute outcomes should make no moral difference.

Normatively speaking, the issue comes down to whether we should think moral luck is real, or whether it is a mistake to let luck play any role in our moral assessment of people and their actions. Some have argued that moral luck is the result of hindsight bias: after the fact, people think that an outcome was more likely to happen simply because it did happen, which biases their moral verdict. Others have favored various forms of epistemic reductionism (Schinkel 2009); moral luck intuitions could be explained by the fact that what we are after when we make moral judgment is people's intentions, but that we use outcomes as *evidence* for people's intentions. Alternatively, these intuitions may be based on knowledge attributions; unlucky drivers and fathers hold false beliefs about the future outcomes of their actions, which may make us view them as more morally blameworthy (Young et al. 2010).

How do these explanations bear on *the gap*? Recently, people have turned to an evolutionary perspective for answers. Here, the idea is that blame and punishment serve an adaptive function: they are supposed to provide a learning environment that favors cooperation and pro-social dispositions at the expense of free-riding and antisocial tendencies. Now, the empirical evidence suggests that only rigid punishment based on outcomes rather than intentions or the goal of deterrence can do this (Cushman 2008, 2013, 2015). Perpetrators can deceive others about their intentions, which always remain somewhat opaque; moreover, they can strategically disincentivize punishment by indicating that they are unwilling to learn, thereby ruling out deterrence as a possible rationale for punishing. Only outcome-based punishment escapes these two problems. Sensitivity to resultant luck thus makes evolutionary sense.

This suggests that moral luck is justified for consequentialist reasons which used to obtain in our environment of evolutionary adaptedness (Kumar 2017). Interestingly, some people have used similar evolutionary arguments to make the opposite point: in assigning blame, it used to make sense to rely on proxies for potential wrongdoers' mental states which are hard, if not impossible, to access directly (Levy 2016). However, this also means that whenever we have more direct and reliable evidence regarding people's mental states, these more informed judgments should trump those which are based on less trustworthy proxies.

Moral Intuition

(viii) *Rationalism and Sentimentalism*. Should we think of the influence of moral judgments on seemingly nonmoral concepts as a pernicious one? Obviously, this does not merely depend on the relevance of moral judgments for those other cognitive domains, but also on whether moral judgments themselves have a sound basis.

For an astonishingly long time, philosophers have thought that the question whether moral judgments can be trusted or not could be substituted for the question whether these judgments were based on emotion or reason. Some sentimentalists, such as Hume, thought moral judgments had to be grounded in the former. Reason, his argument went, was in the business of determining facts; moral judgments, on the other hand, were capable of motivating people to act. But, Hume also argued, only feelings and desires have such motivational force; and since feelings and desires do not have the right "direction of fit" (Smith 1987), they are not in the business of determining facts. Hence, moral judgments could not be based on reason. Others, such as Kant, argued that this could not be true, since moral judgments were supposed to have unconditional authority, which emotion could not deliver. They thus went looking for a purely rational justification of moral requirements that was cleansed of all emotional impurity.

I say "astonishingly long time" because on closer inspection, the idea that reason and emotion are somehow opposed forces has little to commend it and tends to evaporate rather quickly. And yet for the most part, empirically informed philosophers have not just sided with the sentimentalist tradition (Nichols 2004; Prinz 2006, 2007), but continued to dress up their sentimentalism—the claim that moral judgments are based on emotion—as an alternative to rationalism.

As far as the empirical evidence is concerned, this meant showing that emotions do not merely accompany moral judgments, but properly constitute them. One way to do this is to show that reasoning doesn't produce moral judgments. Emotionally charged intuitions take primacy, which reason merely rationalizes after the fact. When people's reasoning is debunked, they tend not to give up their moral intuitions, but enter a state of "moral dumbfounding" (Haidt 2001). It is true in general that people only have poor introspective access into what drives their moral judgments (Uhlmann et al. 2009; Hall et al. 2012). Moreover, emotions seem to be both necessary and sufficient for moral judgment (Prinz 2006). Evidence from psychopathic individuals suggests that impaired emotion leads to impaired moral judgment (Blair 1995). Emotion manipulation studies seem to demonstrate that changing people's emotions changes their moral beliefs as well (Schnall et al. 2008; Wheatley and Haidt 2005; Valdesolo and DeSteno 2006). Then again, more recent studies suggest that psychopaths, though suffering from diminished empathy, guilt, and remorse, are indeed able to draw the distinction between moral and conventional norms (Aharoni et al. 2012). The aforementioned emotion manipulation studies, in turn, are problematic in that they focus on very specific subgroups of the population (e.g., highly hypnotizable subjects), find statistically significant effects

only for some vignettes, and, perhaps most importantly, fail to alter the polarity of people's moral beliefs (e.g., from "X is right" to "X is wrong"; May 2014). But even if it had been shown that moral judgments are thoroughly saturated with emotion, it remains unclear why this would have any implications for how trustworthy they are (Sauer 2012a).

(ix) *Evolutionary Debunking*. What other grounds, besides an obsolete commitment to the incompatibility of emotion and reason and the shaky evidence adduced to support it, are there for believing that moral intuition may be a poor guide to the moral truth?

Evolution—of course. I have already mentioned one example for how evolutionary considerations can be used to undermine a subset of moral intuitions: the Greene/Singer strategy of debunking deontological intuitions as alarm-like responses to morally irrelevant factors such as up-close-and-personal harm that were selected for in an environment we no longer inhabit (see section (i) above).

But so-called evolutionary debunking arguments (Kahane 2011) can be generalized to cover all moral judgments. The basic strategy is this: many, if not all, of our moral judgments can in some way be traced back to a few basic evaluative dispositions. We want to avoid pain, punish evildoers, sympathize with vivid suffering, care about our kin, like to reciprocate favors, and dislike cheaters. It is overwhelmingly plausible that evolution has something to do with why we hold these values and not their opposites, or something else entirely (such as "the fact that something is purple is a reason to scream at it," Street 2006, 133). Now, suppose there are certain objective moral facts: facts about right and wrong, or about what we have most moral reason to do. How likely is it that we are in a position to know these facts when relying on our basic evaluative dispositions?

Spectacularly unlikely, some have argued (Joyce 2006). In fact, it would be pure serendipity for our moral beliefs to hit upon the moral truth by accident, given that the mechanism that shaped the dispositions we rely upon in making those judgments bore no connection whatsoever to their truth. Evolutionary pressures select for traits which are adaptive; but unlike in the nonmoral case, where false beliefs can get you killed, moral beliefs don't have to be true to allow you (and a fortiori your genes) to survive. Unless we have something else to go on—which we do not—this insight thoroughly undermines our moral intuitions.

I cannot summarize the rich literature on this topic here, so let me just hint at some possible responses, always keeping an eye on *the gap*. Evolutionary debunking arguments pose a reliability challenge—the processes that produce our moral judgments do not aim at truth, but at increasing the frequency of our genes in a given population. Now, some have argued that the evolutionary challenge can be met (Huemer 2005; Fitzpatrick 2014): our capacity to make moral judgments may be the upshot of a more general capacity, such as reason or intelligence, for which there *is* an evolutionary rationale. Some have tried to show that the challenge overgeneralizes in various unwelcome ways. After all, what reason is there to believe that evolution has given us the capacity to recognize mind-independent mathematical truths (Clarke-Doane 2012)? Some have

suggested that the challenge can be redirected. According to evolutionary debunkers, moral judgments are produced by off-track processes. But what if there is no track to be *on* at all? If moral judgments do not aim at discerning any mind-independent moral truths to begin with, then the threat of moral skepticism is disarmed (Street 2006). Finally, some have argued that there is a class of moral beliefs that remains immune to debunking, because it cannot be explained on evolutionary grounds (de Lazari-Radek and Singer 2012). An attitude of universal and impartial benevolence, for instance, seems to confer no fitness benefits. The debate on evolutionary debunking shows, at any rate, how tightly connected normative and so-called *metaethical* questions regarding the nature of moral values and value judgments are.

(x) *The Reliability of Intuition*. Distal causes such as evolution are not the only ones to cast doubt on the trustworthiness of our moral intuitions. Proximal ones, such as the susceptibility of those intuitions to irrelevant features of the situation, seem to provide more direct and less speculative grounds for skepticism toward the reliability of moral cognition.

For instance, people's moral beliefs appear to be subject to order effects (Liao et al 2012; Schwitzgebel and Cushman 2012). For instance, subjects are more likely to judge it permissible to push a person to her death to save five others when the respective scenario was presented before a similar one in which a runaway trolley had to be redirected using a switch to achieve the same result. This effect holds even for professional philosophers among which some familiarity with the scenarios given can be presumed. Framing effects, in which people's moral judgments are affected by how and in what context an option is presented, are also frequently cited as an unwelcome influence on our moral thinking (Sinnott-Armstrong 2008).

These findings lead us to a possible skeptical argument. In making moral judgments, we rely on moral intuitions. But if, as the evidence suggests, these intuitions are sensitive to morally extraneous factors the presence of which we are frequently unaware of and sometimes cannot rule out, then our intuitions require confirmation. But the only thing we have to confirm our moral intuitions are *more moral intuitions*. The justification of our moral beliefs seems to have no hinges to turn on.

How unreliable do framing effects make moral judgments? According to one very reasonable measure of reliability, the mean probability that a subject will *not* change her moral judgment depending on framing or order is 80%—not so bad (Demaree-Cotton 2016; cf. Andow 2016). Moreover, as in the case of emotion manipulation studies more generally, effect sizes tend to be small, and framing effects rarely alter the polarity of people's judgments. That is to say, subjects' judgments are somewhat affected by being in one frame or another, but people do not, strictly speaking, change their minds.

Moreover, debunking arguments aiming to show that moral intuitions are unreliable face one crucial limitation: they rely on moral intuitions themselves, in particular regarding which factors count as morally irrelevant and which do not (Rini 2015). In order for such arguments to get off the ground, then, at least some moral judgments *must* be considered reliable.

Bridging the Gap

The guiding question of this chapter was: given the empirical presuppositions of normative theories of moral judgment and agency, what is the normative significance of empirical facts about our moral psychology? Most importantly, how should we think about the relationship between the two in light of *the gap*? Let's provide at least a tentative answer to this question.

I have surveyed a variety of topics that moral psychologists and empirically informed philosophers are currently working on, ranging from more specific issues such as which normative ethical theory fares best in light of empirical scrutiny, to whether human beings tend to have what it takes to satisfy the requirements of moral agency, to the influence of moral judgment on nonmoral thinking and, finally, the reliability of moral cognition in general.

It is rather clear that, though empirical research has no *direct* normative implications, there are ways to make empirical research normatively *relevant*. Empirical information always needs to be coupled with normative bridging principles to develop genuine moral impact. Note, however, that this is not an indictment of empirically informed moral philosophy, as the situation is exactly symmetrical with respect to purportedly "pure" normative inquiry, which equally fails to have any genuine normative implications unless coupled with empirical bridging principles that connect it to the real world.

In addition to the three positive ones mentioned below, I have one negative lesson to offer about trying to make empirical data normatively significant. It may seem trivial, but is easily—and frequently—ignored: avoid hasty, sweeping generalizations. Claims such as "moral intuitions are unreliable/reliable," "people are free/there is no such thing as free will," or "people are essentially good/bad" are unlikely to be true unless appropriately qualified to add nuance, in which case the bolder version of the claim turns out to be not just untrue and imprecise, but also unhelpful. Rather, empirically informed normative inquiry should be conducted in a piecemeal fashion. Exactly how, and to what extent, are intentionality attributions driven by normative judgment? How strong is the influence of framing effects on moral beliefs? In what sense may people have stable or fragmented personality traits, and how do they manifest? How does human decision-making work, when does it break down, and what causes it to do so? These complex questions cannot be answered with bold, attention-grabbing slogans—not correctly, at any rate.

Here is the first positive lesson I believe can be drawn: empirical data can develop normative relevance by *undermining the empirical presuppositions of various normative ethical theories regarding what kind of creature we are*. This means that when it comes to *the gap*, the *ought implies can* principle is at least as important as the *no ought from an is* principle. If we literally cannot act in the way postulated by a moral theory, then it cannot be the case that we ought to act in that way. To be sure, it is true that moral theories are not in the business of merely describing the world. Ultimately, normative inquiry is about what is good or right, and the normative power of the factual only goes so far. But it makes little sense to come up with fancy

ideals no one can bring herself to care about, while ignoring the things we do care about because they do not comport with the clever principles we came up with in our study. This point has been very clearly articulated by Owen Flanagan (1991), who calls it the "principle of minimal psychological realism" (32ff.). We see it at work, for instance, in sections (iii) and (iv) above.

My second lesson has it that empirical moral psychology can uncover that *the etiology of our moral intuitions sometimes undermines their justification*. Psychological debunking arguments of this sort all share the same basic structure: (1) There is a class C of moral judgments that is generated by cognitive process P. (2) P is unreliable with respect to C. (3) C is unjustified. (Or, alternatively, a subject S would be unjustified in holding a belief out of C if S arrived at that belief on the basis of P.)

Actually, debunking arguments are a motley bunch rather than a monolithic strategy. All debunking arguments try to show that a given belief has been generated by dubious processes. But there are various ways of spelling out this dubiousness. It is useful to distinguish six different types of debunking: (a) off-track debunking: a moral belief is based on a cognitive process that does not track the (moral) truth, e.g., evaluative tendencies that are evolutionarily adaptive, but not morally trustworthy (see section (ix) above). (b) Hypersensitivity debunking: many moral judgments are driven by feelings of disgust. But disgust is a hypersensitive "better safe than sorry" mechanism that generates an unhealthy amount of false positives and should thus be viewed with skepticism (Kelly 2011). (c) Hyposensitivity debunking: empathy is the (potential) source of at least as many moral judgments as disgust. But empathy is a hyposensitive mechanism that generates many false negatives due to its inherent partiality toward the near and dear (Prinz 2011). (d) Obsoleteness debunking: some judgmental processes used to be epistemically viable, but no longer are because the natural and social scaffolding they used to fit has disappeared. Our intuitive morality has been shaped to deal with the demands of stable, intimate, small-scale tribal groups in the Pleistocene. We are ill-equipped to deal with environments very unlike this one—namely, the one we currently happen to live in (Greene 2013). (e) Inconsistency debunking: in some cases, we can build inconsistent pairs of moral judgments, one or both of which we thereby know has to be given up because the difference between the two moral judgments may be based on nothing but a morally irrelevant factor (Campbell and Kumar 2012). (f) Ignoble origins debunking: this is the "original" type of debunking made famous by nineteenth (and early twentieth) century renegades such as Marx, Nietzsche, and Freud. It aims to uncover the ugly distal history of certain moral views by showing that they originated in processes, events, or dispositions that are either inherently undesirable or at least inconsistent with the targeted moral outlook. Christianity preaches love and compassion, but is founded on resentment and envy; capitalism is founded on the ideal of equal rights and fairness, but these ideals actually just serve the interests of the ruling class; and so on (Prinz 2007, 215ff.). The power of debunking arguments is discussed in sections (i), (ii), and (viii)–(x).

A final lesson is this: often, but certainly not often enough, empirical information can develop normative significance by enabling us to use this information for the *reflexive improvement of moral judgment and agency* (Rini 2013). We cannot discount

implicit biases unless we know how, why, when, and under what conditions they operate. Empirical research can tell us when and how the tools we wish to deploy in moral cognition and action are unsuitable or likely to be broken. Sections (i), (ii), and (v)–(x) nicely illustrate the usefulness of this lesson.

The problem is that we have no way of knowing introspectively when this is the case. In fact, we have no way of knowing, in general, what causes our thoughts and desires, and our folk theories of how our thinking works are often hopelessly inadequate. Empirical research is essential for this reflexive purpose, and ignoring or dismissing it reckless and foolish.

References

Aharoni, E., Sinnott-Armstrong, W., & Kiehl, K. A. (2012). Can psychopathic offenders discern moral wrongs? A new look at the moral/conventional distinction. *Journal of Abnormal Psychology, 121*(2), 484.

Alfano, M. (2013). *Character as moral fiction*. Cambridge University Press.

Alfano, M. (2016). *Moral psychology: An introduction*. Cambridge: Polity.

Alfano, M., & Loeb, D. (2014). Experimental moral philosophy. *Stanford Encyclopedia of Philosophy*, 1–32.

Andow, J. (2016). Reliable but not home free? What framing effects mean for moral intuitions. *Philosophical Psychology, 29*(6), 904–911.

Appiah, A. (2008). *Experiments in ethics*. Cambridge, MA: Harvard University Press.

Bargh, J. A., & Chartrand, T. L. (1999). The unbearable automaticity of being. *American Psychologist, 54*, 462–479.

Baumeister, R. F., Masicampo, E. J., & Nathan DeWall, C. (2009). Prosocial benefits of feeling free: Disbelief in free will increases aggression and reduces helpfulness. *Personality and Social Psychology Bulletin, 35*(2), 260–268.

Beebe, J. R., & Buckwalter, W. (2010). The epistemic side-effect effect. *Mind & Language, 25*(4), 474–498.

Berker, S. (2009). The normative insignificance of neuroscience. *Philosophy and Public Affairs, 37*(4), 293–329.

Blair, R. J. R. (1995). A cognitive developmental approach to morality: Investigating the psychopath. *Cognition, 57*(1), 1–29.

Brink, D. O. (1984). Moral realism and the sceptical arguments from disagreement and queerness. *Australasian Journal of Philosophy, 62*(2), 111–125.

Campbell, R., & Kumar, V. (2012). Moral reasoning on the ground. *Ethics, 122*(2), 273–312.

Clark, C. J., et al. (2014). Free to punish: A motivated account of free will belief. *Journal of Personality and Social Psychology, 106*(4), 501–513.

Clarke-Doane, J. (2012). Morality and mathematics: The evolutionary challenge. *Ethics, 122*(2), 313–340.

Cova, F., & Naar, H. (2012). Side-effect effect without side effects: The pervasive impact of moral considerations on judgments of intentionality. *Philosophical Psychology, 25*(6), 837–854.

Cushman, F. (2008). Crime and punishment: Distinguishing the roles of causal and intentional analyses in moral judgment. *Cognition, 108*(2), 353–380.

Cushman, F. (2013). The role of learning in punishment, prosociality, and human uniqueness. In K. Sterelny, R. Joyce, B. Calcott, & B. Fraser (Eds.), *Cooperation and its evolution*. Cambridge, MA: MIT Press.

Cushman, F. (2015). Punishment in humans: From intuitions to institutions. *Philosophy Compass, 10*(2), 117–133.

Cushman, F., Knobe, J., & Sinnott-Armstrong, W. (2008). Moral appraisals affect doing/allowing judgments. *Cognition, 108*(2), 353–380.

Darley, J. M., & Batson, C. D. (1973). From Jerusalem to Jericho: A study of situational and dispositional variables in helping behavior. *Journal of Personality and Social Psychology, 27*(1), 100–108.

Darwall, S., Gibbard, A., & Railton, P. (1992). Toward Fin de Siècle Ethics: Some trends. *Philosophical Review, 101*(1), 115–189.

Demaree-Cotton, J. (2016). Do framing effects make moral intuitions unreliable? *Philosophical Psychology, 29*(1), 1–22.

Doris, J. M. (2002). *Lack of character: Personality and moral behavior*. Cambridge: Cambridge University Press.

Doris, J. M. (2009). Skepticism about persons. *Philosophical Issues, 19*(1), 57–91.

Doris, J. M., & Murphy, D. (2007). From My Lai to Abu Ghraib: The moral psychology of atrocity. *Midwest Studies in Philosophy, 31*(1), 25–55.

Doris, J., & Plakias, A. (2008). How to argue about disagreement: Evaluative diversity and moral realism. In W. Sinnott-Armstrong (Ed.), *Moral psychology*, *The cognitive science of morality: Intuition and diversity* (Vol. 2, pp. 303–331). Cambridge, MA: MIT Press.

Doris, J. M., & Stich, S. P. (2005). As a matter of fact: Empirical perspectives on ethics. In F. Jackson & M. Smith (Eds.), *The Oxford handbook of contemporary philosophy*. Oxford: Oxford University Press.

Enoch, D. (2009). How is moral disagreement a problem for realism? *The Journal of Ethics, 13*(1), 15–50.

Evans, J. (2008). Dual-processing accounts of reasoning, judgment, and social cognition. *Annual Review of Psychology, 59*, 255–278.

Fitzpatrick, S. (2014). Moral realism, moral disagreement, and moral psychology. *Philosophical Papers, 43*(2), 161–190.

Flanagan, O. J. (1991). *Varieties of moral personality: Ethics and psychological realism*. Cambridge, MA: Harvard University Press.

Foot, P. (2001). *Natural goodness*. Oxford: Oxford University Press.

Fraser, B., & Hauser, M. (2010). The argument from disagreement and the role of cross-cultural empirical data. *Mind & Language, 25*(5), 541–560.

Greene, J. D. (2008). The secret joke of Kant's soul. In W. Sinnott-Armstrong (Ed.), *Moral psychology: Vol. 3. The neuroscience of morality: Emotion, brain disorders, and development*. Cambridge, MA: MIT Press.

Greene, J. D. (2013). *Moral tribes: Emotion, reason, and the gap between us and them*. New York: Penguin Press.

Greene, J. D. (2014). Beyond point-and-shoot morality: Why cognitive (Neuro)science matters for ethics. *Ethics, 124*(4), 695–726.

Greene, J. D., et al. (2014). Are "counter-intuitive" deontological judgments really counter-intuitive? An empirical reply to Kahane et al. (2012). *Social Cognitive and Affective Neuroscience, 9*(9), 1368–1371.

Greene, J. D., Nystrom, L. E., et al. (2004). The neural bases of cognitive conflict and control in moral judgment. *Neuron, 44*, 389–400.

Greene, J. D., Sommerville, B. D., et al. (2001). An fMRI investigation of emotional engagement in moral judgment. *Science, 293*, 2105–2108.

Haggard, P., & Eimer, M. (1999). On the relation between brain potentials and the awareness of voluntary movements. *Experimental Brain Research, 126*(1), 128–133.

Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review, 108*(4), 814.

Hall, L., Johansson, P., & Strandberg, T. (2012). Lifting the veil of morality: Choice blindness and attitude reversals on a self-transforming survey. *PloS One, 7*(9), e45457.

Harman, G. (1999). Moral philosophy meets social psychology: Virtue ethics and the fundamental attribution error. *Proceedings of the Aristotelian Society, 99*(1999), 315–331.

Harman, G. (2009). Skepticism about character traits. *The Journal of Ethics, 13*(2/3), 235–242.

Holton, R. (2010). Norms and the Knobe effect. *Analysis, 70*(3), 1–8.

Huemer, M. (2005). *Ethical intuitionism*. New York: Palgrave Macmillan.

Hume, D. (1739/2000). *A treatise of human nature*. Oxford: Oxford University Press.
Hursthouse, R. (1999). *On virtue ethics*. Oxford: Oxford University Press.
Isen, A. M., & Levin, P. F. (1972). Effect of feeling good on helping: Cookies and kindness. *Journal of Personality and Social Psychology, 21*(3), 384–388.
Joyce, R. (2006). *The evolution of morality*. Cambridge: MIT Press.
Kahane, G. (2011). Evolutionary debunking arguments. *Noûs, 45*(1), 103–125.
Kahane, G. (2012). On the wrong track: Process and content in moral psychology. *Mind & Language, 27*(5), 519–545.
Kahane, G., et al. (2012). The neural basis of intuitive and counterintuitive moral judgement. *Social Cognitive and Affective Neuroscience, 7*(4), 393–402.
Kahane, G., Everett, J. A. C., Earp, B. D., Farias, M., & Savulescu, J. (2015). "Utilitarian" judgments in sacrificial moral dilemmas do not reflect impartial concern for the greater good. *Cognition, 134*, 193–209. doi:10.1016/j.cognition.2014.10.005.
Kahneman, D. (2011). *Thinking, fast and slow*. London: Macmillan.
Kamm, F. M. (2007). *Intricate ethics: Rights, responsibilities, and permissible harm*. New York: Oxford University Press.
Kelly, D. (2011). *Yuck!: The nature and moral significance of disgust. A Bradford book*. Cambridge, MA: MIT Press.
Knobe, J. (2003). Intentional action and side effects in ordinary language. *Analysis, 63*(3), 190–194.
Knobe, J. (2010). Person as scientist, person as moralist. *Behavioral and Brain Sciences, 33*(4), 315–329.
Knobe, J., & Fraser, B. (2008). Causal judgment and moral judgment: Two experiments. In W. Sinnott-Armstrong (Ed.), *Moral psychology* (Vol. 2). Cambridge, MA: MIT Press.
Knobe, J., & Leiter, B. (2007). The case for Nietzschean moral psychology. In B. Leiter & N. Sinhababu (Eds.), *Nietzsche and morality*. Oxford: Oxford University Press.
Korsgaard, C. M. (1996). *The sources of normativity*. Cambridge: Cambridge University Press.
Kristjánsson, K. (2012). Situationism and the concept of a situation. *European Journal of Philosophy, 20*(S1), E52–E72.
Kumar, V. (2017). Moral vindications. *Cognition*. Vol. 167, 124–134.
Kumar, V. (forthcoming). The ethical significance of cognitive science. In S.-J. Leslie & S. Cullen (Eds.), *Current controversies in philosophy of cognitive science*. Routledge.
de Lazari-Radek, K., & Singer, P. (2012). The objectivity of ethics and the unity of practical reason. *Ethics, 123*(1), 9–31.
Leiter, B. (2007). Against convergent moral realism: The respective roles of philosophical argument and empirical evidence. In W. Sinnott-Armstrong (Ed.), *Moral psychology*, *The cognitive science of morality: Intuition and diversity* (Vol. 2, pp. 333–337). Cambridge, MA: MIT Press.
Levy, N. (2015). Less blame, less crime? The practical implications of moral responsibility skepticism. *Journal of Practical Ethics, 3*(2), 1–17.
Levy, N. (2016). Dissolving the puzzle of resultant moral luck. *Review of Philosophy and Psychology, 7*(1), 127–139.
Liao, S. M., Wiegmann, A., Alexander, J., & Vong, G. (2012). Putting the trolley in order: Experimental philosophy and the loop case. *Philosophical Psychology, 25*(5), 661–671.
Libet, B. W. (1985). Unconscious cerebral initiative and the role of conscious will in voluntary action. *Behavioral and Brain Sciences, 8*(4), 529–566.
Mackie, J. L. (1977). *Ethics: Inventing right and wrong*. New York: Penguin.
Martin, R., & Barresi, J. (2003). Personal identity and what mattes in survival: An historical overview. In R. Martin & J. Barresi (Eds.), *Personal identity*. Oxford: Blackwell.
Mason, E. (2013). Objectivism and prospectivism about rightness. *Journal of Ethics and Social Philosophy, 7*(2).
May, J. (2014). Does disgust influence moral judgment? *Australasian Journal of Philosophy, 92*(1), 125–141.
Merritt, M. (2009). Aristotelean virtue and the interpersonal aspect of ethical character. *Journal of Moral Philosophy, 6*(1), 23–49.
Meyers, C. D. (2013). Defending moral realism from empirical evidence of disagreement. *Social Theory and Practice, 39*(3), 373–396.

Mikhail, J. (2007). Universal moral grammar: Theory, evidence, and the future. *Trends in Cognitive Sciences, 11*(4), 143–152.
Miller, C. (2003). Social psychology and virtue ethics. *The Journal of Ethics, 7*(4), 365–392.
Moore, G. E. (1903). *Principia Ethica*. New York: Dover Publications.
Nadelhoffer, T. (2006). Bad acts, blameworthy agents, and intentional actions: Some problems for juror impartiality. *Philosophical Explorations, 9*(2), 203–219.
Newman, G. E., De Freitas, J., & Knobe, J. (2015). Beliefs about the true self explain asymmetries based on moral judgment. *Cognitive Science, 39*(1), 96–125.
Nichols, S. (2004). *Sentimental rules: On the natural foundations of moral judgment*. New York: Oxford University Press.
Nichols, S. (2014). Process debunking and ethics. *Ethics, 124*(4), 727–749.
Nichols, S., & Knobe, J. (2007). Moral responsibility and determinism: The cognitive science of folk intuitions. *Noûs, 41*(4), 663–685.
Nisbett, R. E., & Cohen, D. (1996). *Culture of honor: The psychology of violence in the South*. Boulder, CO: Westview Press.
Paxton, J. M., Ungar, L., et al. (2012). Reflection and reasoning in moral judgment. *Cognitive Science, 36*(1), 163–177.
Pettit, D., & Knobe, J. (2009). The pervasive impact of moral judgment. *Mind & Language, 24*(5), 586–604.
Phillips, J., Misenheimer, L., & Knobe, J. (2011). The ordinary concept of happiness (and others like it). *Emotion Review, 71*(3), 929–937.
Prinz, J. (2006). The emotional basis of moral judgments. *Philosophical Explorations, 9*(1), 29–43.
Prinz, J. (2007). *The emotional construction of morals*. Oxford: Oxford University Press.
Prinz, J. (2009). The normativity challenge: Cultural psychology provides the real threat to virtue ethics. *The Journal of Ethics, 13*(2–3), 117–144.
Prinz, J. (2011). Against empathy. *Southern Journal of Philosophy, 49*(s1), 214–233.
Rini, R. A. (2013). Making psychology normatively significant. *The Journal of Ethics, 17*(3), 257–274.
Rini, R. A. (2015). Morality and cognitive science. *Internet Encyclopedia of Philosophy*.
Robinson, B., Stey, P., & Alfano, M. (2015). Reversing the side-effect effect: The power of salient norms. *Philosophical Studies, 172*(1), 177–206.
Ross, L., & Nisbett, R. E. (1991). *The person and the situation*. Philadelphia: Temple University Press.
Sauer, H. (2012a). Psychopaths and filthy desks: Are emotions necessary and sufficient for moral judgment? *Ethical Theory and Moral Practice, 15*(1), 95–115.
Sauer, H. (2012b). Morally irrelevant factors: What's left of the dual process-model of moral cognition? *Philosophical Psychology, 25*(6), 783–811.
Sauer, H. (2014). It's the Knobe effect, stupid! *Review of Philosophy and Psychology, 5*(4), 485–503.
Sauer, H., & Bates, T. (2013). Chairmen, cocaine, and car crashes: The Knobe effect as an attribution error. *The Journal of Ethics, 17*(4), 305–330.
Schinkel, A. (2009). The problem of moral luck: An argument against its epistemic reduction. *Ethical Theory and Moral Practice, 12*(3), 267–277.
Schlosser, M. E. (2012a). Free will and the unconscious precursors of choice. *Philosophical Psychology, 25*(3), 365–384.
Schlosser, M. E. (2012b). Causally efficacious intentions and the sense of agency: In defense of real mental causation. *Journal of Theoretical and Philosophical Psychology, 32*(3), 135–160.
Schlosser, M. E. (2014). The neuroscientific study of free will: A diagnosis of the controversy. *Synthese, 191*(2), 245–262.
Schnall, S., et al. (2008). Disgust as embodied moral judgment. *Personality and Social Psychology Bulletin, 34*(8), 1096–1109.
Schultze-Kraft, M., et al. (2016). The point of no return in vetoing self-initiated movements. *Proceedings of the National Academy of Sciences, 113*(4), 1080–1085.

Schwitzgebel, E., & Cushman, F. (2012). Expertise in moral reasoning? Order effects on moral judgment in professional philosophers and non-philosophers. *Mind & Language, 27*(2), 135–153.
Singer, P. (2005). Ethics and intuitions. *The Journal of Ethics, 9*(3–4), 331–352.
Sinnott-Armstrong, W. (2008). Framing moral intuitions. In W. Sinnott-Armstrong (Ed.), *Moral psychology*, *The cognitive science of morality* (Vol. 2, pp. 47–76). Cambridge, MA: MIT Press.
Smith, M. (1987). The human theory of motivation. *Mind, 96*(381), 36–61.
Sneddon, A. (2009). Normative ethics and the prospects of an empirical contribution to the assessment of moral disagreement and moral realism. *Journal of Value Inquiry, 43*(4), 447–455.
Soon, C. S., Brass, M., Heinze, H. J., & Haynes, J. D. (2008). Unconscious determinants of free decisions in the human brain. *Nature Neuroscience, 11*(5), 543–545.
Sripada, C., & Konrath, S. (2011). Telling more than we can know about intentional action. *Mind & Language, 26*(3), 353–380.
Stanovich, K. (2011). *Rationality and the reflective mind*. New York: Oxford University Press.
Street, S. (2006). A Darwinian dilemma for realist theories of value. *Philosophical Studies, 127*(1), 109–166.
Strohminger, N., & Nichols, S. (2014). The essential moral self. *Cognition, 113*(2014), 159–171.
Tiberius, V. (2014). *Moral psychology: A contemporary introduction*. Routledge.
Tobia, K. P. (2015). Personal identity and the phineas gage effect. *Analysis, 75*(3), 396–405.
Uhlmann, E. L., Pizarro, D. A., Tannenbaum, D., & Ditto, P. H. (2009). The motivated use of moral principles. *Judgment and Decision making, 4*(6), 479.
Valdesolo, P., & DeSteno, D. (2006). Manipulations of emotional context shape moral judgment. *Psychological Science, 17*(6), 476–477.
Velleman, J. D. (2011). *How we get along*. Cambridge: Cambridge University Press.
Vranas, P. B. (2005). The indeterminacy paradox: Character evaluations and human psychology. *Noûs, 39*(1), 1–42.
Webber, J. (2013). Character, attitude and disposition. *European Journal of Philosophy, 21*(1), 1082–1096.
Wegner, D. M. (2002). *The illusion of conscious will*. Cambridge, MA: MIT Press.
Wheatley, T., & Haidt, J. (2005). Hypnotic disgust makes moral judgments more severe. *Psychological Science, 16*(10), 780–784.
Young, L., Cushman, F., Hauser, M., & Saxe, R. (2007). The neural basis of the interaction between theory of mind and moral judgment. *Proceedings of the National Academy of Sciences, 104*(20), 8235–8240.
Young, L., Nichols, S., & Saxe, R. (2010). Investigating the neural and cognitive basis of moral luck. *Review of Philosophy and Psychology, 1*(3), 333–349.
Zimmerman, M. J. (2011). *The immorality of punishment*. Buffalo, NY: Broadview Press.

An Evolutionarily Informed Study of Moral Psychology

Max M. Krasnow

As this volume attests, the study of morality is a difficult task. Philosophers, research scientists and scholars across the academy have been wrestling with how to define, measure, and think about morality for thousands of years. Yet, despite this effort, answers to fundamental questions have been devilishly elusive, with researchers still debating what morality even is. Why is it that the study of morality is such a difficult task? Perhaps the answer is less a result of the subject material than of the minds of those studying it. While the human mind is not usually considered an impediment to scientific progress, it may present particular barriers to accurate models of the nature of morality and moral psychology.

This is not the first research question that has been hampered by the fact that science is done by humans. Often, the problem is that we have a powerful intuition or perception of how the world seems or ought to be that gets in the way of scientifically understanding how the world really is. For instance, unassisted by technology, our eyes look to the horizon and see the Earth stretching out as if in a plane. And indeed, flat Earth theory was held in many cultures around the world for hundreds of years. Scientists had been studying gravity for centuries before Einstein's formulation of gravity as a distortion of space-time geometry gave us a more accurate, though far less intuitive, theory: general relativity. Here and elsewhere, the fact that human intuition or perception does not well map the real world has made humans worse at science. But the psychological barriers to understanding morality may not merely be a problem of this kind.

Whatever the specific design of our psychology turns out to be, results have been collected that make it very hard to believe that this design is simply a machine for uncovering the objective truth of the world. Whether it is the emotional dog wagging the rational tail (Haidt 2001), our heuristics and biases (Gilovich et al. 2002),

M.M. Krasnow (✉)
Evolutionary Psychology Laboratory, Department of Psychology, Harvard University,
980 William James Hall, Cambridge, MA, USA
e-mail: krasnow@fas.harvard.edu

B.G. Voyer, T. Tarantola (eds.), *Moral Psychology*,
DOI 10.1007/978-3-319-61849-4_3

or our susceptibility to visual illusions (Gregory 1997) and illusory correlations (Whitson and Galinsky 2008), it very often seems to be that the inferences the brain reliably makes are not always targeted at the objective truth. But, should we expect the mind to be designed for discovering objective truths? The human brain is a product of natural selection just like any other organ of the body, and just like every other organ of the body, the design features it has are there because they solved adaptive problems in the past (Tooby and Cosmides 1992). The brain, like any other feature of an organism, is the result of generations of successive filters on potential designs, filters that preferentially maintained those designs that optimized reproduction against the backdrop of environments experienced by the organism's ancestors. We should only expect design for uncovering objective truth to be a reliable feature of the human brain if doing so was a reliable solution to one of these filters by contributing to reproduction or by hitchhiking along as the by-product of another design that did. For many of the designs that make up human psychology, neither of these options is very likely.

In contrast, it is eminently more plausible that in many cases designs that vied away from objective truth seeking in the direction of inferences and behaviors that reliably contributed to reproductive fitness were the ones that better survived the various filters. We should expect this for three distinct but convergent reasons. First, there are likely many inferences for which knowing the true state of the world caries absolutely no fitness gain. For example, for a terrestrial primate, perceiving gravity as a distortion of space-time and not merely a force that pulls objects down toward the Earth cannot plausibly have influenced anyone's fitness over ancestral conditions; this information is irrelevant in the extreme. If there is no selection pressure that would maintain the focus of an inference system on reliably picking out the true state of the world, then randomizing forces like mutation should degrade its precision and allow drift over time. Or, if information necessary to make a certain inference was not reliably available ancestrally, then there would not have been any filter on designs that picked up on this information and used it in the right way. For example, it is inarguable that humans have a richly articulated mating psychology that uses information about prospective mates (cues of age, health, status, trustworthiness, fertility, etc.) to make mating decisions (Buss and Schmitt 1993). For men at least, a lot of this design is aimed at picking fertile targets of mating effort out of the sea of nontargets. But, there was no selective filter on mechanisms that meta-represented that this was their function; designs could increase reproduction by assigning high mate value in decision-making to bearers of particular cues like youthful appearance, low fluctuating asymmetry, etc., but there was no gain from representing that these inferences were about reproduction per se. So, while male mating psychology has design for inferring fertility and using it to inform mating decisions, it does not meta-represent what it is about and so does not accurately assess relevant fertility across all situations. That is why plastic surgery can maintain youthful attractiveness divorced from objective age, why men don't line up at sperm banks competing to fertilize as many eggs as possible, and why men continue to find their female partners attractive even when they are contracepting. Many mechanisms are likely to be under a similar filter, where solving a problem can be

done without the solution necessarily explicitly considering what the problem is and how the solution solves it. For at least these mechanisms, objective truth seeking per se is simply not the problem to be solved.

Second, there are likely many inferences for which the costs of getting the inference wrong are asymmetrical—that is, the false positives are more or less costly than the misses (Delton et al. 2011; Haselton and Buss 2000; Johnson et al. 2013). Taking again the example of a terrestrial primate, mistaking a bit of ground-level motion at your peripheral vision for a snake and deploying an evasive response is minimally costly—regardless of whether you are actually avoiding a snake or a harmless breeze, the energy expended is relatively minimal. Alternately, failing to detect a poisonous snake if it actually is there and thus getting bitten is a very costly error. Many ancestral problems are likely to have this asymmetric costs profile, and so mechanisms designed to infer the state of the world relevant for these problems are likely to incorporate design that makes the expensive error relatively less likely than the cheap error. For a snake-avoidance mechanism, the goal isn't to be as accurate as possible but to be as unbitten as is reasonable.

Third, the social world is not a solitary game: my behavior can influence others' behavior which can then impact my fitness. The beliefs I hold, my motivations for action, the things I value, and how I act can all have consequences, and can be relevant to others and how they treat me. For example, both game theoretic modeling and simple intuition predict that we will like others who have a history of good deeds, preferring them as social partners and treating them better than we would otherwise (Axelrod and Hamilton 1981; Barclay 2013; Trivers 1971). As such, design to be seen to do good deeds (but probably not to be seen being seen) has very likely been under selection. Behaviors, attitudes, ideas, and opinions that signal this kind of disposition (or other dispositions of similar relevance) are all potential targets of selection for expressing them, especially when others can see. To the extent that these targets (particular kinds of beliefs, opinions, etc.) are different than the objective state of the world, we should for this reason not expect the respective psychologies to be designed to be aligned to the objective truth. Who among us hasn't genuinely felt and then told a romantic partner that they are the most beautiful man or woman in the world? We can't all be right, but it sure feels nice to have a partner say so. In the moral domain, the selection pressures responsible for our moral sentiments—our concern for the sick, our outrage at the oppressor, etc.—may be more about what these sentiments signal to others than anything to do with objective truth seeking.

Taking these points together—that the objective truth is often fitness irrelevant, that the right kind of error is often ecologically rational, and that the adaptive problem is at least sometimes about changing someone else's behavior—helps suggest a program for an evolutionarily informed study of human moral psychology. The first task is to identify the major filters—that is, the adaptive problems—that components of moral psychology have been designed to solve. Considering the ecology of ancestral hominins, more than a few adaptive problems stand out as both presenting substantial selection pressure and potentially producing morality-relevant psychological design, including but by no means limited to optimally allocating resources

between the self and others (Delton and Robertson 2016), attracting and keeping cooperative partners (Delton and Robertson 2012; Krasnow et al. 2012), marshalling allies and maintaining one's group membership (Pietraszewski 2016; Tooby et al. 2006), and preventing yourself and those you value from being exploited (Krasnow et al. 2016; Sell et al. 2009). The task is to get specific about the recurring features of the ancestral environment that a design solution could use to solve the problem, including what kinds of information a behaving organism would have access to and how the organism's behavior could affect its outcome. Importantly, the points above suggest that our moral intuitions and reasoning about moral problems are likely the result of mechanisms shaped by one or more of the above selection pressures and therefore that they may be systematically biased in directions away from what might otherwise be considered the objective truth or normative moral correctness and toward what were ancestrally fitness maximizing conclusions.

There is a large and growing literature that can be analyzed (or reanalyzed) using the evolutionary framework suggested here. Rather than attempting an exhaustive review, below I sketch what I see as a major dividing line in the space of adaptive problems involved in morality and discuss research that exemplifies the distinction in ways I hope will be helpful to researchers going forward.

Inward- vs. Outward-Facing Mechanisms

Regardless of the other features of an adaptive problem, the solution is either to regulate my behavior (solved by what I will call here an *inward-facing mechanism*) or to influence the behavior of others (solved by what I will call here an *outward-facing mechanism*) or both. Typically, the former category has been the main focus of work in moral psychology. For example, a lot of work looks at what it means to do good or to be altruistic. Philosophers ask on what theory "good" is measured, be it a utilitarian calculus or some other set of ideals. Economists propose elements of subjective utility (e.g., the "warm glow") that could compensate for otherwise costly behavior (Andreoni 1990; Fehr and Schmidt 1999). Cognitive psychologists and neuroscientists ask what are the proximate mechanisms of this decision and ask whether it is reflexive or deliberative (Rand et al. 2012) and if neural reward centers are involved (Decety et al. 2004). But, misconstruing the target of the mechanism across the distinction of inward-vs.-outward facing can have major consequences for our understanding of the psychology. If doing good has been selected because being seen by others to be doing good resulted in being chosen for more or better cooperative relationships—that is, results from an outward-facing mechanism—then researchers have been looking for the benefits to balance the equation in the wrong place. You would never intuit your way to this answer by introspecting on the experience of doing good; you would just conclude that you do good because it is the right thing to do, because it feels good, because it triggers a dopamine pulse, etc. Problematically, when the target of a moral adaptive problem is to influence

another's behavior, one's own representation of one's motives is likely to be especially suspect. As discussed below, mistaking the target as inward rather than outward facing may be an especially likely mistake for humans to make.

I should note this inward- vs. outward-facing dimension is very similar to the distinction DeScioli and Kurzban (2009) made between what they term "condemnation" (moral adaptations for judging others' bad behavior) and "conscience" (moral adaptations for governing one's own behavior in order to preempt others' condemnation). DeScioli and Kurzban construe the ecology of moral problems as involving perpetrators, victims, and observers, with condemnation resulting from the interest of observers and conscience resulting from the preemptive response of potential perpetrators. Yet, for conscience to preempt condemnation, it can use outward-facing expressions (contrast the mere private experience of guilt with a guilty expression). Condemnation can result from inward-facing design (contrast outwardly concerned rehabilitative punishment with the private orientation of ostracism). Dissecting the problem space as inward vs. outward facing, I believe, more cleanly aligns the adaptive problems with the mechanistic design features that solve them. Below I review a selection of morally relevant psychological mechanisms to hopefully illustrate the utility of this alternative inward- vs. outward-facing distinction.

Moral Sentiments Regarding When to Be Nice

Inward-Facing Mechanisms

Codes of morality around the world are filled with proscriptions concerning when and how to be nice to others, when and with whom to share, and who is entitled to being helped. But how to optimally share resources is not just an abstract moral question; who gets what is inherently fitness relevant. In a world where others in your environment may share genes in common with you by virtue of recent common descent (i.e., kin), traits that allocate resources in ways that maximize the likelihood of your genes reproducing will be favored by natural selection; this is Hamilton's theory of cooperation via kin selection (Hamilton 1964). In a world where there are gains in trade to be had by pooling or exchanging your resources with others, then traits that maximize these gains will be favored by natural selection; this is Trivers' theory of reciprocal altruism (Trivers 1971). In a world where others represent unique value to you—such as unique constellations of mutual interest—then traits that tend to keep them around and in good shape would be favored by natural selection; this is Tooby and Cosmides' theory of deep engagement (Tooby and Cosmides 1996). On these and other theories, the mind should embody design that, at least in some circumstances, favors giving resources away to others and being perfectly happy to do so.

Recent work has asked, "What kind of psychology could embody strategies that produce these other-favoring effects?" In answering this question, researchers have

considered features of the ancestral ecology that simple heuristics could exploit to produce, on average, a good approximation of a solution. While a great deal about the ancestral world cannot be known, certain features can be safely assumed. For example, the ancestral social world was filled with different kinds of relationships; some people were strangers to you, and others were your family, friends, or cooperative partners. While there were doubtlessly many features that discriminated these categories from each other, it can be safely assumed that our ancestors could not predict the future with perfect certainty. At least sometimes, someone who at first blush appeared to be an irrelevant stranger never to be seen again actually became a relevant social partner (Krasnow et al. 2013). Moreover, for a hunting and gathering hominid with a specialized division of labor, long periods of childcare, and the ability to both accumulate and transmit cultural knowledge, there were likely many gains in trade possible between our ancestors where the gains were potentially lucrative. A tendency to trust others on the chance that a mutually beneficial relationship could develop—that is, a psychology of default trust—would be optimal social foraging in such an environment. While default trust is risky, as some investments would not pay off, the long-term rewards should be higher than those of a safer, asocial strategy (Delton et al. 2011; Delton and Robertson 2012; Rand et al. 2014). There are many ways such a design could be implemented. Just as our mechanisms of animacy and agency detection seem to be hypersensitive, attributing these features even when the evidence is scanty or absent, our mechanisms of social foraging could be designed to err on the side of treating even strangers as if they could be long-term cooperative partners. And just as our mechanisms of animacy detection help coordinate our behavior without going through explicit cognition—jumping away from a rustle you thought hid a snake did not require you to explicitly represent the propositions "this is a snake," "snakes are dangerous," and "snake danger can be mitigated by avoiding proximity"—our mechanisms of social foraging could plausibly be designed to effect behavior in the absence of explicit representations like "I might see this person again," "this person may be able to help me out later," and "if I don't see them again, at least I'm not risking much."

Relatedly, it is safe to assume that not all social partners were created equal; some were more trustworthy than others. When presented with attractive outside options (a more profitable partner to trade with, a tempting reason to cheat, etc.), some partners would have been more likely to take the option than others. A partner who simply doesn't consider these outside options should be more trustworthy than one who does, and to the extent that we can perceive cues to this disposition—such as a friend immediately agreeing when asked for help—a psychology that was sensitive to these cues and found them appealing in others would be favored by selection (Hoffman et al. 2015). Just as mating mechanisms are built to accept cues of fertility—available information like low waist-to-hip ratio (Lassek and Gaulin 2008) that partially indexes information that is otherwise inaccessible—social mechanisms should use observable cues in a partner's behavior like loyalty or blindness to outside options to index the otherwise inaccessible information of a partner's association value.

This work informs the kind of mechanisms we should expect to underlie our moral intuitions of when to be nice. Humans and our recent ancestors have been

intensely social for millions of years, so these adaptive problems have been long-standing. Solving a problem like social foraging with a robust intuition may be a timeworn solution, one that doesn't suffer from failures of explicit reasoning to anticipate future benefits. But while fitness maximization may be the ultimate explanation for our moral intuitions, it does not minimize their sincerity or authenticity. Just as a mother's love and concern for her child, a genuine and passioned response if there ever was one, is the result of mechanisms designed to maximize reproductive fitness, there is every reason to expect that our affiliation to our friends, our feelings of genuine concern for others, our intuitions about who deserves help and when similarly result from mechanisms designed to maximize reproductive fitness via social behavior. Some moral phenomena may be by-products of these inward-facing mechanisms, like our mechanisms of parental care can spill over onto our pets. But what about the expression of these emotions, motivations, and decisions? What problems do they solve?

Outward-Facing Mechanisms

Taking the above mechanisms as a given immediately suggests a reciprocal set of adaptive problems to be solved: how do you best position yourself to be preferred or chosen by others? When others in your environment are distributing resources, allocating aid, and forming relationships nonrandomly with respect to the characteristics of the recipients, selection should act to increase the prevalence of designs that preferentially capture these benefits. Just as preferences in peahens select for plumage in peacocks, selection pressures for social foraging result in selection pressures on social display. The instantiation of the mechanisms that embody these solutions can take many forms. As above, there is little reason to predict a priori that the solutions should necessarily route through explicit reasoning. Just as babies don't smile at their parents because they consciously consider the benefits of smiling, the mechanisms that instantiate our outward-facing responses to social selection need not be proximately Machiavellian. In fact, we should expect them to not have this design. To the extent that a benefit was provided by someone who explicitly saw something in it for themselves, the beneficiary should not attribute the gift to an underlying disposition on the part of the actor to value the recipient (Tooby and Cosmides 1996; Tsang 2006). A "friend" who only helps you out when it is in their own best interest is not much of a friend. In contrast, it is precisely those who are (or appear) insensitive to their own proximate payoffs that should be the most trustworthy and dependable cooperative partners. Imagine asking your friend a favor only to find them ponder at length all of the possible consequences they would face. What would you think of them? A heuristic solution to this pressure of impression management is to simply cooperate yourself without considering alternative—potentially more appealing—options (Hoffman et al. 2015).

A growing body of work suggests that this dynamic is likely to extend beyond the case of cues indicating *whether* a partner considered outside options before

cooperating to the more general set of cues indicating how much a partner values you at all (Delton 2010; Krasnow et al. 2016; Petersen et al. 2012; Sell et al. 2009). One interesting place this design is turning up is in the psychology of charitable giving. Charity is widely viewed as morally good and intuitively about increasing the well-being of the recipient. But, if that were the concern of the mechanism generating our charitable impulses, we would probably do charity a lot differently than we actually do. Many have begun to point out that most of our charity is incredibly ineffective: We don't pick causes that present the biggest problems, we don't fund solutions that provide the biggest benefits, and in large part we don't seem to care (Money for good: Revealing the voice of the donor in philanthropic giving 2015). Why is this? An intriguing possibility is that our minds are actually designed to prefer giving to less efficient charities because of what they can signal about how much we value others. I needn't value a child very highly to spend a dime to feed her for a year; even if I cared for her very little, I would still prefer to give up the dime. But I must value her highly to spend a dime to give her just a grain of rice; for how little she benefited, I must value her highly to justify giving up the money. If our psychology of charitable giving is the product of mechanisms designed for this outward-facing target of value signaling, then we should in fact predict different designs than were the targets merely inward-facing: rather than giving benefits efficiently at low personal cost to provide large charitable benefits, a psychology designed to signal how much it values others should look for (but not be seen looking for) opportunities to pay large costs to provide comparably inefficient charitable benefits and have a chance to be seen doing so.

Moral Sentiments Regarding When to Be Mean

Inward-Facing Mechanisms

Often our moral concerns fuel anger, outrage, or indignation toward those who violate our moral code. Sometimes these emotions result in behaviors that harm these individuals, ostracizing them from benefits they would otherwise have access to or inflicting costs on them through punishment or more violent aggression. Many theories have been proposed to account for the evolution or expression of these kinds of motivations and behaviors. But, as above, I argue here that theories have a better chance of being right when we properly construe the target of the adaptation as either inward or outward facing. Some adaptive problems can be solved by reaching out and changing another individual's behavior; these adaptive problems select for outward-facing solutions. It is likely that our punitive responses often result from such an outward-facing mechanism. But, does it always? If an adaptive problem does not have this form, researchers looking for outward-facing solutions would be looking in the wrong place.

One adaptive problem that punishment can solve is the mere prevention of future bad actions. By ostracizing, incapacitating, or killing a bad actor, the punisher and those she cares about are no longer susceptible to the bad action (Duntley and Buss

2011). Especially in the case of killing, these responses don't require the targeted individual to change their mind about anything for their bad behavior to be prevented; the decision is unilaterally made by the punisher. But, taking this option also precludes enjoying any of the benefits that would have otherwise obtained if the punished person was still around. Optimally negotiating this trade-off involves design for reducing the motivation for harsh, incapacitating, or corporal punishment given cues that these forgone benefits would be substantial—that is, that the perpetrator has high association value. And the mind indeed shows this design, favoring rehabilitative sanctions more for high association value perpetrators and punitive sanctions more for low association value perpetrators when deciding on criminal sentencing (Petersen et al. 2012; Wilson and Rule 2015). This function can be accomplished merely by moderating the sanction a person metes out, though. The outward expressions of offense—including facial, postural, and vocal expressions—are big noisy signals that are superfluous to this function. If these are adaptations, their adaptive target is likely of the outward-facing variety.

Outward-Facing Mechanisms

The evolutionary function of punishment on most theories is to change another organism's behavior (Clutton-Brock and Parker 1995). For example, conflicts of interests abound in life and can sometimes be adjudicated by force. This situation can be modeled as an asymmetric war of attrition, where (1) two parties make costly bids for a contested resource, (2) both parties pay the cost of the lower bid (i.e., fight until one gives up), and (3) the higher bidder wins the resource (Hammerstein and Parker 1982). In this scenario, each party is incentivized to bid just up to their private valuation of the resource; any more would be entailing sure losses and any less would potentially leave value on the table. Imagine fighting with your sister over what to watch on television. You each have your own preferred show, but only one can be watched, and you can annoy each other into giving in. If you don't care very much about your show, it would be silly to put up too much of a fight as you would waste more in fighting than you cared about the show in the first place. And, if you don't put up enough of a fight, you might end up missing your show when you didn't have to. A costly strategy would be to actually keep fighting with your sister until it's not worth it anymore, just in case she backs down first. But, if you can predict being outbid and losing the fight, you can save your effort and avoid fights you would otherwise lose. You are likely to be outbid to the extent that she either (a) values the resource more than you do, or (b) faces lower costs of aggression than you do, or both. As such, mechanisms that outwardly express our valuation and formidability would be selected to cost-effectively deter aggressive conflicts with others.

Many aspects of the anger response in humans and other animals can be understood as components of this signaling architecture (Sell et al. 2010, 2012, 2014). Humans and other animals posture before fights to size up the competition to predict

if fighting would be worthwhile. During these prefight rituals, the potential combatants don't merely stand passively; they modify their visual and auditory appearance to seem bigger, stronger, and meaner than they usually do. By engaging in this signaling, individuals can reach into the minds of observers and manipulate their mental contents in ways that advantage the signaling individual, potentially earning the contested resource and more deferential treatment in the future.

This kind of outward-facing mechanism can be used in larger social contexts as well. Just as signaling to someone who offended against you can deter them from doing so in the future, signaling to someone who offended against others in your presence can signal that you would not tolerate such treatment yourself. The third-party punishment paradigm—where one participant can punish another for acting poorly toward someone else—has been widely used to model moral condemnation. Recent work has revealed that at least some of the third-party punishment we observe in experiments results from this kind of deterrence mechanism (Krasnow et al. 2016). Moreover, punishing on behalf of others has been found to signal cooperative value more broadly, such as that the punisher herself could be trusted to not act badly (Jordan et al. 2016). Third-party moral condemnation seems at least in part to result from two outward-facing mechanisms for regulating the behavior of others.

As these examples illustrate, the components of our anger, punitive, or condemnation psychologies that are geared toward outward-facing targets—like signaling to others—were under reliably different selective filters than those components that are merely inward facing. Outward-facing mechanisms of signaling, for example, are expected to be under arms-race dynamics (Dawkins and Krebs 1979). The value of a signal depends on the population of signals it competes with. If everyone but you exaggerates their formidability, by neglecting to exaggerate, you appear weaker by comparison. The same process should apply to our expressions of outrage or condemnation; if everyone but you exaggerates their outrage to some moral violation, by neglecting to exaggerate, you appear relatively less trustworthy, more exploitable, etc., than you otherwise could. In contrast to behaviors that result from merely inward-facing mechanisms, those with outward-facing components are expected to be prone to these dynamics.

An Evolutionarily Informed Study of Moral Psychology

Applying the lens of evolutionary psychology to the study of morality offers several unique insights. Most basically, analyzing the ancestral human ecology for morality-relevant selection pressures can help generate hypotheses of adaptations in moral psychology—design features in the mechanistic basis of our moral intuitions, motivations, and decision-making. Here I have argued that it is profitable to distinguish those selection pressures that can be solved by merely inward-facing mechanisms targeted at directing the organism's own behavior from outward-facing mechanisms targeted at changing the behavior of others. One reason this distinction may prove important is that outward-facing mechanisms (e.g., broadcasting cooperative

disposition by charitable giving or public moral outrage) are expected to be under selection to obscure their ecological rationality (e.g., obscuring "ulterior" motives) even from those attempting to study them from an objective perspective. Using our intuition as a scientific instrument and source of hypotheses is therefore likely to systematically mischaracterize the adaptive design of our moral psychology and especially those involving outward-facing mechanisms. An evolutionary perspective helps clarify why studying morality is such a difficult task and also helps guide our efforts so that we are at least looking in the right place for the answers.

References

Andreoni, J. (1990). Impure altruism and donations to public goods: A theory of warm-glow giving. *The Economic Journal, 100*(401), 464–477. doi: 10.2307/223413.

Axelrod, R., & Hamilton, W. D. (1981). The evolution of cooperation. *Science, 211*, 1390–1396.

Barclay, P. (2013). Strategies for cooperation in biological markets, especially for humans. *Evolution and Human Behavior, 34*(3), 164–175. doi:10.1016/j.evolhumbehav.2013.02.002.

Buss, D. M., & Schmitt, D. P. (1993). Sexual strategies theory: An evolutionary perspective on human mating. *Psychological Review, 100*(2), 204–232. doi:10.1037/0033-295X.100.2.204.

Clutton-Brock, T. H., & Parker, G. A. (1995). Punishment in animal societies. *Nature, 373*, 209–216.

Dawkins, R., & Krebs, J. R. (1979). Arms races between and within species. *Proceedings of the Royal Society of London B: Biological Sciences, 205*(1161), 489–511. doi: 10.1098/rspb.1979.0081.

Decety, J., Jackson, P. L., Sommerville, J. A., Chaminade, T., & Meltzoff, A. N. (2004). The neural bases of cooperation and competition: An fMRI investigation. *NeuroImage, 23*(2), 744–751. doi: 10.1016/j.neuroimage.2004.05.025.

Delton, A. W. (2010). *A psychological calculus for welfare tradeoffs*. Santa Barbara: University of California.

Delton, A. W., Krasnow, M. M., Cosmides, L., & Tooby, J. (2011). The evolution of direct reciprocity under uncertainty can explain human generosity in one-shot encounters. *Proceedings of the National Academy of Sciences of the United States of America, 108*(32), 13335–13340.

Delton, A. W., & Robertson, T. E. (2012). The social cognition of social foraging: Partner selection by underlying valuation. *Evolution and Human Behavior, 33*(6), 715–725. doi: 10.1016/j.evolhumbehav.2012.05.007.

Delton, A. W., & Robertson, T. E. (2016). How the mind makes welfare tradeoffs: Evolution, computation, and emotion. *Current Opinion in Psychology, 7*, 12–16.

DeScioli, P., & Kurzban, R. (2009). Mysteries of morality. *Cognition, 112*(2), 281–299.

Duntley, J. D., & Buss, D. M. (2011). Homicide adaptations. *Aggression and Violent Behavior, 16*(5), 399–410. doi: 10.1016/j.avb.2011.04.016.

Fehr, E., & Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation. *The Quarterly Journal of Economics, 114*(3), 817–868.

Gilovich, T., Griffin, D., & Kahneman, D. (2002). *Heuristics and biases: The psychology of intuitive judgment*. Cambridge: Cambridge University Press.

Gregory, R. L. (1997). Visual illusions classified. *Trends in Cognitive Sciences, 1*(5), 190–194. doi: 10.1016/S1364-6613(97)01060-7.

Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review, 108*(4), 814–834. doi: 10.1037/0033-295X.108.4.814.

Hamilton, W. D. (1964). The genetical evolution of social behaviour. *Journal of Theoretical Biology, 7*, 1–52.

Hammerstein, P., & Parker, G. A. (1982). The asymmetric war of attrition. *Journal of Theoretical Biology, 96*(4), 647–682.

Haselton, M. G., & Buss, D. M. (2000). Error management theory: A new perspective on biases in cross-sex mind reading. *Journal of Personality and Social Psychology, 78*, 81–91.

Hoffman, M., Yoeli, E., & Nowak, M. A. (2015). Cooperate without looking: Why we care what people think and not just what they do. *Proceedings of the National Academy of Sciences, 112*(6), 1727–1732. doi: 10.1073/pnas.1417904112.

Johnson, D. D. P., Blumstein, D. T., Fowler, J. H., & Haselton, M. G. (2013). The evolution of error: Error management, cognitive constraints, and adaptive decision-making biases. *Trends in Ecology & Evolution, 28*(8), 474–481. doi: 10.1016/j.tree.2013.05.014.

Jordan, J. J., Hoffman, M., Bloom, P., & Rand, D. G. (2016). Third-party punishment as a costly signal of trustworthiness. *Nature, 530*(7591), 473–476. doi: 10.1038/nature16981.

Krasnow, M. M., Cosmides, L., Pedersen, E. J., & Tooby, J. (2012). What are punishment and reputation for? *PloS One, 7*(9), e45662. doi: 10.1371/journal.pone.0045662.

Krasnow, M. M., Delton, A. W., Cosmides, L., & Tooby, J. (2016). Looking under the hood of third-party punishment reveals design for personal benefit. *Psychological Science, 27*(3), 405–418.

Krasnow, M. M., Delton, A. W., Tooby, J., & Cosmides, L. (2013). Meeting now suggests we will meet again: Implications for debates on the evolution of cooperation. *Scientific Reports, 3*. doi: 10.1038/srep01747.

Lassek, W. D., & Gaulin, S. J. C. (2008). Waist-hip ratio and cognitive ability: Is gluteofemoral fat a privileged store of neurodevelopmental resources? *Evolution and Human Behavior, 29*, 26–34.

Money for good: Revealing the voice of the donor in philanthropic giving. (2015). Retrieved from http://static1.squarespace.com/static/55723b6be4b05ed81f077108/t/56957ee6df40f330ae018b81/1452637938035/$FG+2015_Final+Report_01122016.pdf.

Petersen, M. B., Sell, A., Tooby, J., & Cosmides, L. (2012). To punish or repair? Evolutionary psychology and lay intuitions about modern criminal justice. *Evolution and Human Behavior, 33*(6), 682–695. doi: 10.1016/j.evolhumbehav.2012.05.003.

Pietraszewski, D. (2016). How the mind sees coalitional and group conflict: The evolutionary invariances of N-person conflict dynamics. *Evolution and Human Behavior, 37*(6), 470–480. doi: 10.1016/j.evolhumbehav.2016.04.006.

Rand, D. G., Greene, J. D., & Nowak, M. A. (2012). Spontaneous giving and calculated greed. *Nature, 489*(7416), 427–430.

Rand, D. G., Peysakhovich, A., Kraft-Todd, G. T., Newman, G. E., Wurzbacher, O., Nowak, M. A., & Greene, J. D. (2014). Social heuristics shape intuitive cooperation. *Nature Communications, 5*, 3677. doi: 10.1038/ncomms4677.

Sell, A., Bryant, G. A., Cosmides, L., Tooby, J., Sznycer, D., Von Rueden, C., et al. (2010). Adaptations in humans for assessing physical strength from the voice. *Proceedings of the Royal Society of London B: Biological Sciences, 277*(1699), 3509–3518.

Sell, A., Cosmides, L., & Tooby, J. (2014). The human anger face evolved to enhance cues of strength. *Evolution and Human Behavior, 35*(5), 425–429.

Sell, A., Hone, L., & Pound, N. (2012). The importance of physical strength to human males. *Human Nature, 23*, 30–44. doi: 10.1007/s12110-012-9131-2.

Sell, A., Tooby, J., & Cosmides, L. (2009). Formidability and the logic of human anger. *Proceedings of the National Academy of Sciences, 106*, 15073–15078. doi: 10.1073/pnas.0904312106.

Tooby, J., & Cosmides, L. (1992). The psychological foundations of culture. In J. H. Barkow, L. Cosmides, & J. Tooby (Eds.), *The adapted mind: Evolutionary psychology and the generation of culture* (pp. 19–136). New York: Oxford University Press.

Tooby, J., & Cosmides, L. (1996). Friendship and the Banker's paradox: Other pathways to the evolution of adaptations for altruism. *Proceedings of the British Academy, 88*, 119–143.

Tooby, J., Cosmides, L., & Price, M. E. (2006). Cognitive adaptations for n-person exchange: The evolutionary roots of organizational behavior. *Managerial and Decision Economics, 27*, 103–129.

Trivers, R. L. (1971). The evolution of reciprocal altruism. *The Quarterly Review of Biology, 46*, 35–57.

Tsang, J.-A. (2006). The effects of helper intention on gratitude and indebtedness. *Motivation and Emotion, 30*(3), 198–204. doi: 10.1007/s11031-006-9031-z.

Whitson, J. A., & Galinsky, A. D. (2008). Lacking control increases illusory pattern perception. *Science, 322*(5898), 115–117. doi: 10.1126/science.1159845.

Wilson, J. P., & Rule, N. O. (2015). Facial trustworthiness predicts extreme criminal-sentencing outcomes. *Psychological Science, 26*(8), 1325–1331. doi: 10.1177/0956797615590992.

Moral Psychology: An Anthropological Perspective

Paolo Heywood

Introduction

It is in many ways the traditional task of anthropology to point out exceptions to rules. Provide us with a generalization about human behaviour, and we will describe to you a far corner of the world in which it does not hold. This has to a large extent been true of our dealings with moral psychology, as I detail below, and it will come as no surprise to many readers that relativism, or at least rhetorical invocations of it, has long been a staple of anthropological approaches to morality.

But as this chapter will describe, recent developments in the anthropological study of ethics have led many anthropologists interested in the subject to reconfigure their understandings of the universal and the particular in relation to morality. It remains a matter of empirical fact that people across the world think differently about what constitutes right and wrong, good and bad, virtue and vice, and anthropologists continue to document that variety. But it is equally a matter of empirical fact that people across the world do indeed think about such things: that they exercise judgement and reflection about courses of action, ways of attributing responsibility, consequences, behavioural norms, and the like. As I outline below, for us to be able to account seriously and scrupulously for the differences between how people think about ethics, many anthropologists have come to believe that we must possess a coherent vision of what ethics actually means and an explanation for how it is that people do all seem to think about how they ought to live, even though they do so differently.

The new anthropology of ethics, in other words, goes against the grain of a great deal of anthropological writing, in that it begins not with a claim about any particular set of people, but with a claim about people more generally. That general claim is about the ubiquity of ethics, in the sense of moral reasoning, rather than about

P. Heywood (✉)
Division of Social Anthropology, University of Cambridge, Cambridge, UK
e-mail: pph22@cam.ac.uk

B.G. Voyer, T. Tarantola (eds.), *Moral Psychology*,
DOI 10.1007/978-3-319-61849-4_4

ethics in the sense of any particular set of values: as James Laidlaw puts it, 'The claim on which the anthropology of ethics rests is not an evaluative claim that people are good: it is a descriptive claim that they are evaluative' (Laidlaw 2014a: 9). What we value, and perhaps even how we value, in other words, will of course differ across time and space; but the fact that we are evaluative will not.

So nor is that general claim a culturalist one, so to speak: it is not an argument about how societies, cultures, ideologies, or other such systems oblige or compel us to behave and to think in certain ways, which would be another staple anthropological position, as I show below. The idea that people evaluate and reflect upon their thoughts and behaviour is incompatible with that sort of position, though it is not, of course, incompatible with the idea that the way in which they evaluate and reflect will be informed by the contexts in which they find themselves.

Exactly how that 'informed by' works however is a problem that continues to be debated (e.g. Englund 2006; Heywood 2015; Humphrey 2007; Laidlaw 2002, 2014a; Robbins 2007, 2009; Yan 2011; Zigon 2007, 2009a, b). A question this chapter will seek to address then is not so much whether certain values or moral beliefs are universal or particular, but the meta question of how best to theorize our capacity to reflect on such values and beliefs—our freedom—as both universal and particular at the same time.

I will begin by explaining some of the problems with earlier approaches to morality and ethics in anthropology and the ways in which what has come to be known as the anthropology of ethics attempts to resolve those issues and then detail some of the distinctive approaches to ethics that have emerged in the last 20 years, before going on to address the problem I outlined above: how do anthropological approaches to ethics and moral psychology reconcile the premise upon which they are largely built—that moral reasoning and reflection are universal capacities—with the idea that those capacities must also in some way or another be contextually inflected?

Problems with Moral Psychology in Anthropology

For anthropologists interested in the subject, it is by now a truism to note (Edel and Edel 2000 [1959]; Faubion 2001a; Howell 1997; Laidlaw 2002; Parkin 1985; Pocock 1986; Wolfram 1982) that prior to the last two decades, and depending on one's point of view, social anthropology either had a great deal to say about morality and ethics or it had nothing to say about them at all. There are two interrelated reasons for this rather paradoxical problem, and the ways in which those reasons come to appear as problematical rather neatly sets the stage for what the examination of ethics and morality means to contemporary anthropology. These two reasons also correlate broadly—if inexactly—with the approaches that American cultural anthropology and British social anthropology have tended to take towards questions of ethics and morality.

Ethics as Social Norms

In its British form, social anthropology was significantly influenced by the sociology of Emile Durkheim, who identified 'society' as an entity existing above and beyond the level of the individuals who compose it; its sanctions, customs, rules, and codes were understood to be expressions of the collective will of those individuals and hence both compelling and desirable (e.g. Durkheim 1906 [1953]; cf. Laidlaw 2002: 312–315; Laidlaw 2014a: 26–33). In this formulation, in other words, 'morality' consists of the system of constraining obligations imposed upon people by their participation in a larger social group. Or, to put the same idea a different way, 'morality' is the term employed to designate behaviour, beliefs, or actions that adhere to or are in accord with social norms. Durkheim would have been most at home in the world of 'antisocial behaviour orders'.

How effective or not a social system or structure is in its ability to oblige people to abide by its prescriptions is one consequently significant and interesting question. But even as I have just expressed it, it betrays an equally significant and interesting weakness of this understanding of morality: it is—unsurprisingly, given its origins in Durkheimian sociology—fundamentally mechanistic in its view of human behaviour (Laidlaw 2002: 314; Laidlaw 2014a: 28–29). Explanations for people's ideas about what constitutes good or right thought or action are to be sought in the arrangements of society and its component parts, whether those ideas are in accord with social norms (in which case the arrangement is properly functional) or deviate from them (in which case the system is dysfunctional in some sense—as, famously, in Durkheim's study of suicide). As James Laidlaw, one of the anthropologists responsible for our renewed interest in ethics, notes, whilst essentially Kantian in his emphasis on duties and obligations, Durkheim departed from Kant in one very important sense: the task of moral reasoning disappears along with the individual's freedom to reflect on such duties and obligations, whose efficacy no longer depends on the practical will of the subject but on the proper functioning of society as a moral system (Laidlaw 2002: 314; see also Pocock 1986: 8).

So the view of ethics Durkheim bequeathed to social anthropology left us with two serious and related problems: on the one hand, no clear way in which to distinguish between 'moral' and 'social' behaviour, and on the other hand, no language with which to describe people's capacity for moral reasoning or ethical judgment. If morality is simply what 'society' tells you to do, and if whether you do it or not depends simply on whether 'society' is or is not properly put together, then your capacity to think through the value, consequences, or virtue of doing it or otherwise is more or less redundant as far as the analyst is concerned. This is what David Parkin meant when he argued that Durkheim 'so conflated the moral with the social that ethnographers could not isolate for analysis those contemplative moments of moral reflexivity that…so typify human activity and predicaments' (Parkin 1985: 4–5). More recently, it was precisely this blind spot that inspired a number of prominent anthropologists such as Laidlaw to call, more or less at the same time, for the discipline to cease equating the desirable with the normative, in what James Faubion

called an 'objectivist fallacy', of which 'the very definition of ethics as "codes of conduct" is already guilty' (2001a: 83–84).

It is worth also pointing out, with Laidlaw (e.g. 2010a: 370), that even those anthropologists who have preferred to treat instances of 'resistance' to social norms and to attach the adjective 'moral' to domains in which it appears that such resistance occurs are guilty of the same sin. In James Scott's explanation of acts of rebellion on the part of southeast Asian peasants as a kind of 'moral economy' (Scott 1977), or Maurice Bloch and Jonathon Parry's depiction of the 'morality' of various forms of long-term exchange practices (Parry and Bloch 1989), it is in collective opposition to particular social norms that morality is located. But of course 'collective' and 'social' are synonyms, not antonyms, and the only serious difference between this and a more traditionally Durkheimian view of morality is in the moral preferences of the anthropologist it reveals. Both leave, in the end, little or no room for sustained reflection on the part of our interlocutors outside of which particular set of norms they choose to adhere to.

Relativism

The question of whether or not there is anything distinctive about ethics and morality beyond the relationships people possess to social norms is one that American cultural anthropology has also tended to avoid, often by resort to the much-contested notion of relativism. The logical problem with relativism is obvious and has been so since at least Plato's *Theaetetus:* that it is self-refuting to the extent that it itself rests on an absolutist premise—the premise that moral standards only hold value relative to the cultural contexts in which they exist. If that premise is absolutely true, then relativism is self-refuting; if it is not, then it is uninteresting, as parochial as the purported universalisms of which it is critical. Despite this logical problem, however, relativism has long been—and still is in many quarters—a methodological orthodoxy in anthropology (e.g. Geertz 1984), if not in other disciplines. In cultural anthropology, which developed in a somewhat different direction to Durkheimian British social anthropology, this orthodoxy can be traced back to Franz Boas and his students (e.g. Benedict 1935; Herskovits 1972; Mead 1928). In response both to cultural evolutionism as a theoretical perspective—the idea that cultures 'progress' towards a teleological endpoint—and to what they perceived to be a parochial moral universalism in American culture more broadly, they argued that each culture had its own distinct set of customs and norms and that these could not be understood outside of their context. Thus assessing their validity against those of our own is a meaningless and mistaken project. The consequence of this position, if held to consistently, is that the idea of an anthropology of ethics is a fundamentally misguided one, because ethics and morality have no cross-cultural validity as analytic terms, and because where they are referred to as objects in distinct cultures, they are essentially reduced to the customs people live by, as in Durkheimian social anthropology. Here 'morality' equates to 'culture' with the added argument that since 'cultures' are relative so, supposedly, must 'morality' be.

There are a number of problems with this perspective, in addition to the fact that it again leaves us with no way of accounting for the ways in which people actually exercise their moral reason beyond doing what they are told to do by their 'culture'. The main issue is what has elsewhere been called 'the problem of units' (e.g. Holbraad and Pedersen 2009): to function coherently relativism implies the entities that are argued to be relative to one another ('cultures'); yet the idea that the world can be neatly divided into separate spheres that differ from one another in clear and predictable ways has, for fairly obvious reasons, long ceased to be an acceptable position in anthropology (e.g. Fabian 1983; Marcus and Fischer 1986). No 'culture' exists outside of history and their histories are necessarily intertwined. But without the premise that they can somehow be distinguished, relativism can only ever be rhetorical. Once you have conceded that cultures as bounded units do not exist, then relativism is always, in Bernard Williams' terms, too late (Williams 2005: 69; cf. Laidlaw 2014a: 37–38): relativism presupposes separate moral spheres that become 'relative' to one another at the moment they in fact relate; if such spheres do not exist however, it is precisely because such 'relations' between 'cultures' are an ubiquitous, constant process, rather than being the 'problem' that relativism purports to solve.

The other main respect in which the relativism usually espoused in anthropology can only ever really be rhetorical is that as a project, it is almost invariably advanced in the service of a critique of our own values, whatever they are alleged to be in any particular case. As Laidlaw notes, there is an obvious contradiction in arguing on the one hand that we cannot judge the moral standards of a society and on the other that such standards are much superior to our own (2010a: 372).

The Anthropology of Ethics

To sum up, mainstream Anglophone anthropology on either side of the Atlantic has until recently effectively barred itself from enquiring seriously into ideas and practices that we might think of as distinctively ethical or moral. That said, there have been notable exceptions to these problematical trends: in a survey of remarkable breadth, Edel and Edel, a philosopher and an anthropologist, chart the cross-cultural variations in a number of moral problems such as incest and in-group aggression, and attempt to shed light on them with discussions of contemporary ethical theories (Edel and Edel 2000 [1959]); Christoph von Fürer-Haimendorf's *Morals and merit* (Von Fürer-Haimendorff 1967), whilst somewhat evolutionist in its overall outlook, nevertheless provides a detailed ethnographic perspective on the central moral problems of a number of South Asian populations, ranging from hunter-gatherers to Brahmins; D. F. Pocock, writing against Westermarck (2000 [1932]), argued persuasively that the demonstrable existence of moralities which do not apply equally to all individuals (e.g. kin and strangers) is not evidence for the impossibility of universal moral judgements nor does it provide us with a licence to rank moralities on the basis of their capacity for extension, let alone to deny some the status of

morality altogether. Instead, he argued, defining the object of moral acts, the person to whom one has particular duties and responsibilities, is itself a matter requiring of 'moral reasoning' (Pocock 1986: 18), the content of which may vary but the quality of which may be subject to comparative analysis (as later anthropologists have done—see below). In addition, edited volumes by Howell on morality in spheres including Argentinian football and a small Northumberland village (Howell 1997) and Parkin on evil in Buddhist, Islamic, Christian, and non-religious contexts (Parkin 1985) added ethnographic and theoretical breadth to what nevertheless remained a still nascent subfield in anthropology.

With the turn of the millennium however came a burgeoning interest in people's capacity to make moral choices on the basis of considered reflection and judgement. Taking their cue from Aristotelian and other forms of virtue ethics (e.g. MacIntyre 1981) and from Foucault's later writings on technologies of the self (e.g. Foucault 1985, 1986), authors such as Laidlaw (1995, 2002, 2014a), Faubion (2001a, b, 2011), and Lambek (2000, 2010, 2015) all called for sustained enquiry into ethics as an autonomous field of anthropological analysis and into the practices by which individuals pursue virtuous ends and form themselves into moral subjects, and a number of authors have since taken up these themes; 15 years later, the anthropology of ethics has carved out a place for itself within the discipline, and its object of concern is a great deal clearer and more specific than when 'morality' was just another word for 'society'.

A key aspect of this shift, in both its virtue ethicist and Foucauldian forms, has been an attention to the concept of freedom (e.g. Laidlaw 2002, 2014a), which has both helped us respond to the problems identified above, as well as revolutionized the way we understand people's relationship to their thoughts and behaviour more broadly. It is what makes the anthropology of ethics more than simply another subdiscipline.

Foucault and Freedom

To introduce the subject, it is worth distinguishing contemporary understandings of freedom in anthropology from earlier treatments of what might look like similar notions: clearly not all anthropologists have understood the relationship between society and the individual in quite as corporatist a manner as Durkheim or Boas and their descendants. A significant amount of ink has been spilled in the latter half of the twentieth century in an attempt to resolve what is often called the 'structure-agency' problem. In contrast to the idea that what we think and do is largely determined by the social contexts in which we find ourselves, a number of theorists have drawn attention to the active roles that people play in shaping their own social contexts (in being 'agentive'), and still others have attempted by various means to collapse the distinction between structure and agency entirely (e.g. Giddens 1984; Bourdieu 1990).

The problem with the notion of agency, however, as Laidlaw argues, is that it is 'pre-emptively selective' (Laidlaw 2002: 315) in its treatment of what we might

otherwise call freedom. By its nature (as a pole in the structure-agency dyad), it refers not to any behaviour, choice, or set of reasoning, but *specifically* to forms of these which are in some manner or other transformative with respect to social structure, usually either in producing it or in altering it in some respect. In other words, it can only denote action that the analyst deems important in relation to structure. Since most people do not take up or discard ethical ideas on the basis of the effect such an action will have on 'structure', as a concept agency still fails to provide us with a language with which to account for the vast majority of moral reasoning our interlocutors undertake.

Variants of what is called 'practice theory' are the most prominent examples of attempts to collapse the structure-agency distinction altogether (Bourdieu 1990), yet these too seem rather to swing between the two poles than to dispense with them (Laidlaw 2010a: 373). The basic premise of practice theory is that the world is both constructed by and constructive of what we do in it (hence its status as a purported resolution of the structure-agency problem). The concept of *habitus* was popularized by Bourdieu (1990) as a notion that would purportedly combine the corporeal and cognitive and conscious and unconscious aspects of behaviour. *Habitus* is supposed to be both 'structured' by context and 'structuring' of that context. In that latter sense, it points towards what we might think of as freedom. Yet this very capacity to point to instances in which *habitus* is *either* structured *or* structuring (and indeed most frequently it seems to be the former) is indicative of the fact that the two senses are mutually exclusive.

So if freedom is not agency, or *habitus*, what exactly is it? To contemporary anthropologists of ethics, it is, in the language of Foucault, the practice of taking oneself as an object of work and reflection. Understanding Foucault is crucial to understanding what today's anthropologists mean by freedom and ethics: though familiar to many through his work on power and discipline, in his later writings on antiquity, Foucault developed the analytic of 'techniques of the self', operations and exercises by which people actively constitute themselves as subjects. Such techniques come in a range of different forms—from diary-keeping to dietetics—and those forms and the ends to which they are directed will be drawn from and proposed by the historical and cultural contexts in which people find themselves.

Crucially, Foucault's late interest in ethics and freedom did not entail a rejection or replacement of his early writing on power, in which intersubjective relationships produce, rather than constrain, subjects. In works such as *Discipline and Punish* and *The History of Sexuality Vol 1* (Foucault 1975, 1976), he famously argued that power is not a repressive imposition on already-existing subjects, but the very thing that makes subjects what they are. But by their nature, such relationships must involve subjects who are free (to varying degrees) to exercise the power that constitutes these relationships, for outside of them there is nothing. The oft-repeated Foucauldian claim that power is everywhere is equally a claim that freedom is everywhere as well. With one term ('subjectivation') Foucault denotes both the ways in which subjects are produced through their interactions and relations with others and through the work they perform upon themselves. Here, in other words, we have a genuine collapse of the structure-agency distinction.

This conception of freedom does away with two commonly held and related assumptions about what freedom must mean, as Laidlaw points out (Laidlaw 2002: 323): first, the notion that to act freely is to act in accordance with one's 'authentic' self, for no such self can exist outside of its broader context—freedom works through such contexts, not against them—and, second, relatedly, that to act freely must mean to act in the absence of constraint, for there can be no situation in which freedom (or power) is not in some manner reciprocal, because context and self are intertwined. Here, in other words, power and freedom are truly two sides of the same coin.

It is worth noting also that this conception of freedom does away with the 'problem' of relativism. It takes for granted both the ubiquity of power relations and thus also the ubiquity of freedom as an aspect of, rather than an opposition to, those power relations. But of course the nature of those relations and the manner in which that freedom is exercised are going to vary. People will always and everywhere be incited and persuaded to think and act in certain ways by the contexts in which they find themselves, as they will always and everywhere consider and reflect on such thought and action as well, but the subject so produced will vary in all of the myriad ways in which ethnographic research suggests subjects indeed do.

Virtue Ethics

Another strand of the contemporary anthropology of ethics comes to similar conclusions but does so by drawing on Aristotelian virtue ethics and, often, its most recent exposition in the work of philosopher Alasdair MacIntyre (1981). In contrast to both deontological and consequentialist approaches to morality, virtue ethics is intrinsically particularistic: where previous moral philosophies such as the former pair sought universal justifications for moral obligations (whether in reason, the laws of God, sentiment, or utility), virtue ethics takes from Aristotle the idea that bridging the distinction between fact and value, between human nature and the ways in which we ought to live, requires a teleological and thus empirically thorough account of what human nature means in any particular instance and, crucially, what it tends towards. Without an account not only of man 'as he is' but also of man 'as he should be', ethics (the means by which you get from one to the other) makes no sense. Furthermore, as ideas about what man should be vary not only with differing 'traditions' (a concept from MacIntyre intended to be much more fluid and historically informed than 'culture') but also with differing practices and narratives, so will the virtues people pursue and the means by which they pursue them. A virtue ethical approach to moral psychology, in other words, requires an 'ethnographic imagination' in order to understand people's behaviours and motivations not with reference to abstract rules and imperatives but to the stories they tell themselves about their lives and how they shape them. As in the case of Foucauldian ethics, this focus on practical reason (or *phronesis* in Aristotelian terms) requires us to account for reflective judgement in a much more complex manner than debates around structure and agency had previously allowed.

Ethics in Ethnography

A number of anthropologists have made productive use of these two frameworks for thinking about reflection, judgment, and freedom. Talal Asad and his students Charles Hirschkind and Sabah Mahmood have been influential in developing an approach to morality and ethics that combines insights from both. Asad (1986, 1993, 2003) puts together MacIntyre's concept of tradition with Foucault's work on disciplinary techniques to show how Islam and other so-called world religions can be understood as discursive combinations of both orthodoxy and practice, thus eliding the problem of whether norms or actual behaviour should take precedence in analysis. Hirschkind and Mahmood both develop Asad's work on Islam through studies of contemporary Cairo (Hirschkind 2001, 2006; Mahmood 2001, 2005), and both make arguments particularly relevant to anthropological debates around the exercise of freedom and the ways in which people make moral choices. Hirschkind describes the ways in which cassette tape sermon audition can be understood as a technology of the self yet inflects this Foucauldian argument with some of Bourdieu's ideas about the importance of the body to action: what cassette sermon audition develops in listeners is not merely a set of cognitive or intellectual virtues in the sense of instructing them in the tenets of Islam, but also a range of affective, embodied traits such as an 'open heart'.

This idea is taken further in Mahmood's study of women's participation in the Egyptian Piety Movement, which is a critique specifically targeted at 'Western' assumptions about freedom. Writing against feminist arguments about agency residing in opposition to norms (see above), Mahmood makes a persuasive and innovative argument for understanding the ways in which her interlocutors strive to inhabit and fully to embody the norms of the Piety Movement as exercises of freedom. Instead of dismissing their reasoning and behaviour as misguided instances of false consciousness, or trying to locate 'resistance' to it, Mahmood attempts to take them seriously as ethical practices. The women she describes have reasoned and clear understandings of the virtues they wish to develop and why they wish to do so. Like Hirschkind though, she emphasizes the embodied aspect of these practices: indeed, the key virtue these women wish to foster in themselves is an automatic, bodily submission to the will of God; in other words, the moral endpoint of their project is that as a project it should cease to be self-willed, becoming instead a corporeal, preconscious reflex. This makes her arguments about freedom somewhat paradoxical: whilst her laudable goal is to depict the moral reasoning of her interlocutors as the exercise of freedom and reflection that it clearly is, she also seems to wish to depict the endpoint of this exercise—the extinction of the will, the very capacity that makes it an example of moral reasoning—as a form of freedom too (Laidlaw 2014a: 268).

This confusion is perhaps a consequence of too closely conflating ideals and actual behaviour. A number of critics have noted that Mahmood and Asad and MacIntyre before her are too much concerned with finding in the latter the coherence of the former. Anand Pandian, for example, working in a Kallar community in South India (Pandian 2008, 2009), argues that MacIntyre's emphasis on the need for

a tradition to be consistent is misguided: Kallars were seen by both colonial authorities and their Tamil neighbours as a 'criminal' caste. In their ethical reasoning and narrative depictions of virtue, they draw on a range of sources from development discourse to classical poetry to articulate a moral vision that encompasses the 'civil' virtues they have long been encouraged to adopt, as well as their relationship to the 'savage' nature both alleged to reside within them and upon which they physically labour as cultivators. They do this without reconciling this range into a coherent whole. Similarly, Samuli Schielke, working, like Mahmood and Hirschkind, in Egypt (Schielke 2009), highlights the ways in which the path to virtue that Mahmood can sometimes depict as simple and direct can actually be deeply complicated. The young Muslim men he describes are necessarily 'ambivalent' in their commitment to Islamic ideals because the lives they lead present them with alternative goals to pursue, such as love or pleasure. This is not simply a question of doubt in the value of piety as a virtue, but, as Laidlaw points out (Laidlaw 2014a: 203–204), a value conflict in which goods that are in many ways irreconcilable place people in the position of having to reason through their ethical decisions.

Other work in anthropology on ethics has drawn more exclusively on Foucault's vision of ethics. Indeed, a striking contrast—in some ways—with Mahmood and Hirschkind's ethnographies are those of James Faubion (2001a, b, 2011), who, drawing on Foucault's work on antiquity, sees the pedagogical relationship as being in many ways the foundation of ethics—it is a microcosm of intersubjectivity and social context. Thus in antiquity the problem of the relationships between older men and younger boys was not that they might or might not be sexual in nature, but that for them to be ethical they must tend towards developing the freedom of the pupil from the teacher, rather than, as in Mahmood's case, extinguishing that freedom. Which is not to say that Faubion's vision of freedom returns us to a vision of the unconstrained individual, liberated from social constraint—this idea, for Faubion as for Laidlaw and for Foucault himself, is an impossible one, presupposing as it does an asocial individual, an entity entirely lacking in intersubjective relations. But for Faubion the opposite pole of that dichotomy—total domination—remains a possibility (Faubion 2014: 439), and in such a situation there can be no ethics in the form of moral reasoning for there is no freedom with which to reason.

The advantage of this position is that it begins to delineate the contours of what is meant by freedom in a manner more precise than we have seen so far. Hitherto we have examined some of the things that freedom cannot mean—such as liberation—but we have yet to look at a case in which freedom, or a capacity for moral reasoning, can be said not to exist. If indeed it is possible to isolate cases in which freedom does not exist, then it must be more than an ubiquitous, free-floating, ever-present aspect of social life.

But is it really possible to do so? Faubion, in fact, does not give us much in the way of actual cases of total domination. He makes use of Foucault's argument that 'a slave has no ethics' (Foucault 1997: 286; cf. Faubion 2001a: 95; Faubion 2014: 441), suggesting that though its historical accuracy may be a matter of debate, it is helpful as an ideal-typical case (Faubion 2001a: 95). These are somewhat strange words of praise though given that we have already dismissed the opposite pole of

our dichotomy—that of total autonomy—at least partly on the basis that it can only ever be an ideal type. Foucault may have been inaccurate in his depiction of ancient Greek ethics; but it in no way logically follows from this that he intended the statement to apply to anything other than that particular concrete case, as indeed suggested by the context of the discussion in which it appears. Indeed, it sounds more as if Foucault is discussing an ideal type of a particular historical period, rather than a concrete historical type, let alone a universal ideal type. But even though he may have been wrong about ancient Greece, that does not mean he was right—or expected to be so—about anywhere else.

My point in raising this issue is to illustrate just how difficult it can in fact be to think of freedom and our capacity for moral evaluation as something other than the opposite of constraint and in a zero-sum relationship to it. Despite the vast differences between their respective positions, both Mahmood's vision of the endpoint of her interlocutors' moral projects and Faubion's conjuring of the 'anethical' slave share the characteristic of being situations in which—supposedly—moral reasoning has been extinguished by a totally dominating structure or system of power. But to imagine such situations as anything other than thought experiments in the manner of moral philosophy, or as unrealizable orthodoxies, is to return us to a conception of ethics in which culture or context is something that limits freedom, rather than one in which they are simply aspects of the same processes of subjectivation.

Another problem anthropologists have encountered in theorizing the relationship between culture and freedom is that of how to understand situations in which cultures, institutions, societies, or ideological systems present us with multiple, rather than singular moral projects. As I noted above, this idea can in some ways be seen as a response to the overemphasis on coherence and orthodoxy in work such as that of Mahmood. Value pluralism, for example, has a long and distinguished history as a concept in moral philosophy and is often invoked in order to make the argument that the obligation to choose between competing sets of values is an ubiquitous feature of human life and thus likewise is our ability to do so. For the purposes of this chapter, I will confine myself to discussing two anthropological examples in which this multiplicity of norms is an important factor, examples that are in some respects contrasting and in others similar.

Jarrett Zigon conceptualizes 'morality' as having a number of distinct sources (social or cultural institutions, media discourse, etc.), which may conflict with one another (Zigon 2007, 2009a). By 'morality' he means sets of normative social values that people follow largely unthinkingly, akin to those an earlier generation of Durkheimian anthropologists would think of as exhausting the dimension of the moral (see above). Zigon, however, wishes to combine this normative sense of morality with what I have been referring to as 'ethics' (a distinction made use of by a number of anthropologists of ethics, following Foucault) in the sense of an evaluative capacity. He does so by arguing that such a capacity emerges in moments of what he calls 'moral breakdown', namely, situations in which what we take for granted (i.e. moral norms) cannot provide us with a straightforward answer to a particular problem or dilemma. In such situations our evaluative capacities become activated, and ethics becomes something upon which we must think and reflect in

order to decide how best to behave. We do so in order to return to the state of unthinking moral automatism that we departed from in the moment of breakdown, though the norms we return to following will nevertheless have been subtly altered by our ethical choices. Though Zigon draws on the work of Foucault in making this argument, he also makes use of phenomenological insights from Heidegger and other continental philosophers in order to describe how this works at the level of the individual and the ways in which our embodied existence in the world can come to appear strange to us at certain moments. Thus he describes the case of a woman in post-Soviet Russia (a classic situation of moral breakdown, according to Zigon) confronted with the problem of whether or not to pay a bribe to a train inspector and the ways in which she steps outside of both Christian and socialist norms in resolving it (Zigon 2007: 145).

Joel Robbins, on the other hand, makes an argument about the multiplicity of values at a societal scale and draws on Weber and anthropologist Louis Dumont. In his work on the Urapmin, a Melanesian people of Papua New Guinea who converted wholesale to Christianity in the wake of colonial contact (Robbins 2004), Robbins portrays a culture in a state of perpetual moral torment. Whilst the Urapmin have adopted the moral and religious values of Christianity such as submission to God's will, their social and economic life, rooted in precolonial models of exchange and reciprocity, requires them to act in contravention of these values on a regular basis—for example, by neglecting their obligations towards one person in order to build a relationship with another. Building on this ethnographic work, Robbins argues—in a manner not entirely dissimilar to Zigon—that anthropology can retain both the Durkheimian notion of morality as a set of rules we follow more or less unreflectively and by obligation and the idea that we possess the capacity to reflect and evaluate, by employing the notion of value spheres: for the Urapmin, for example, there are situations in which traditional, pre-Christian morality obligates you to do one thing; there are equally situations in which the Christian morality they have adopted obligates you to do another; it is when the two conflict that obligation gives way to choice and unreflective action gives way to evaluation and freedom (Robbins 2007).

As I have noted, Zigon and Robbins differ in some fairly significant ways, for instance, in the scale of their focus (societies or individuals) and on whether or not the situations of choice they both refer to can exist within or only between normative systems of morality (Robbins 2009; Zigon 2009b). Yet both take a particular perspective on culture and on our evaluative capacities, in an attempt to resolve the problem of their relation. In both cases our evaluative capacities are activated in situations in which culture as a constraining system of norms is not properly operative, either because of a 'breakdown' or because it provides us with competing options, and thus obliges us to choose between them. Again, in other words, we have an illustration of the difficulties involved in seeing moral psychology and culture as mutually imbricated: both of these examples are attempts to do so; yet both in the end situate our capacity for moral reasoning as existing outside of culture, when it is dysfunctional in some way, thus returning us to the Durkheimian position I outlined at the beginning of this chapter, and turning ethics into something that operates at structural distance from culture, and only at occasional moments

(cf. Heywood 2015). Both Foucauldian and virtue ethicist approaches to moral psychology insist both on the ubiquity and 'ordinariness' of ethical considerations: that they are built into the fabric of everyday life, rather than in contradistinction to it. This is a point heavily emphasized by anthropologists such as Michael Lambek and Veena Das (Das 2010; Lambek 2000) in their discussions of 'ordinary ethics', their point being precisely that our evaluative capacities are routinely and regularly at work, rather than simply at moments of 'breakdown'.

Conclusion

In many ways the problem I have been seeking to illustrate with this chapter is a version of the old nature/culture dichotomy that has troubled the relationship between anthropology and psychology for a considerable period of time. It is the question of the universal and the particular. As is usually the case when this problem is raised, the answer proffered has unsurprisingly been a combination of both: we all possess a capacity for moral reasoning, yet this capacity is of course inflected by the contexts in which it is activated. This is not a particularly surprising conclusion to reach, though as I have sought to illustrate, it has taken anthropology a surprisingly long time to get there. But what I have also sought to highlight are the difficulties involved in sustaining this insight. It is clearly insufficient simply to state that moral psychology and culture are mutually imbricated, as if the statement alone resolves the problem of their relationship. It does not, particularly when that claims conflicts with others made in the course of it being worked out. As I have suggested, if it is really the case that moral psychology and cultural context are not antinomies in a zero-sum relationship, then there should not be situations in which one is entirely determining of the other. We should not be able to say of someone either that they have no ethics, being entirely bound by their context, nor be able to identify a special moment in which they acquire the unconstrained freedom to choose between different sets of norms, the latter having suddenly lost their power to inflect that freedom.

A consequence of the insight that, in the words of a prominent cognitive anthropologist, 'there are no non-cultural bits of us, as there are no non-natural bits' (Bloch 2012: 76; cf. Laidlaw 2014b), is that there is room for serious and sustained cooperation between anthropology and psychology, and of course such cooperation has been ongoing for some time. But, as with solely anthropological applications of this idea, holding firmly to it has implications for what that cooperation ought to look like. If, for example, freedom and a capacity for moral reasoning are categories we wish to take seriously for all the reasons hitherto outlined, then cross-cultural studies of morality as 'determined' by evolutionary adaptation, for example, look likely to be less than helpful. The anthropology of ethics as it currently stands rests on the idea that people possess a capacity for moral reasoning, which they in turn use to reflect, act, and thus shape their moral worlds, and so anthropologists interested in ethics are unlikely to agree that such worlds are solely the product of adap-

tations to the environment. Likewise, and for similar reasons, most anthropologists of ethics will probably feel that the substitution of experimental methods for sustained ethnographic study would be a methodological impoverishment: the moral psychology they are interested in best displays itself in and through everyday life, rather than when elicited in particular research settings.

There is ample reason to think, though, that experimental data and participant observation may be helpfully combined, as they have been, for example, by Tanya Luhrmann in her work on evangelical Christians' relationships with God (e.g. Luhrmann 2012, 2013). Luhrmann has carried out both traditional anthropological fieldwork with a range of Christian churches in the United States, as well as working and writing with prominent psychologists such as Howard Nusbaum on experiments designed to demonstrate, amongst other things, that practicing certain forms of prayer may cause changes in cognitive processing that lead, for example, to an increased vividness of mental imagery and more unusual sensory experiences, including religious ones. Laidlaw, who has also cooperated extensively with cognitively inclined anthropologists, makes the same point about putting the two together when he notes that the relationship between cognitive science and anthropology must be a 'two-way street' (Laidlaw 2014b): the anthropological studies I have described here and their accompanying insights are products not only of the conceptual premise that nature and culture are not distinct and divergent 'causes' of behaviour but also of sustained participant observation and ethnographic research.

It may have taken anthropology some time to discover an interest in moral psychology; but the kind of 'thick description' that anthropological research produces is in many ways ideally suited to investigating the way moral psychology works in particular situations. We have seen in this chapter that some of the most persuasive contemporary accounts of ethics are not those rooted in deontological or other abstract models of moral reasoning but ones in which people's evaluative capacities are scaffolded by the narrative structures of their lives and experiences, by the decisions they have made in the past and their visions for the future. Anthropology's unique methodological approach—living together with people over sustained periods of time and immersing ourselves in their everyday existence as best we can—puts us in an excellent position to make contributions to broader social science studies of ethics.

References

Asad, T. (1986). *The idea of an anthropology of Islam*. Washington, DC: Georgetown University.

Asad, T. (1993). *Genealogies of religion: Discipline and reason of power in Christianity and Islam*. Baltimore: Johns Hopkins University Press.

Asad, T. (2003). *Formations of the secular: Christianity, Islam, modernity*. Stanford: Stanford University Press.

Benedict, R. (1935). *Patterns of culture*. London: Routledge.

Bloch, M. (2012). *Anthropology and the cognitive challenge*. Cambridge: Cambridge University Press.

Bourdieu, P. (1990). *The logic of practice*. Cambridge: Polity Press.
Das, V. (2010). Engaging the life of the other: Love and everyday life. In M. Lambek (Ed.), *Ordinary ethics: Anthropology, language, and action*. New York: Fordham University Press.
Durkheim, E. (1906 [1953]). The determination of moral facts. In Sociology and philosophy. London: Routledge.
Edel, A., & Edel, M. (2000 [1959]). *Anthropology and ethics*. New York: Transaction Publishers.
Englund, H. (2006). *Prisoners of freedom: Human rights and the African poor*. Berkeley: University of California Press.
Fabian, J. (1983). *Time and the other: How anthropology makes its object*. New York: Columbia University Press.
Faubion, J. (2001a). Toward an anthropology of ethics: Foucault and the pedagogies of autopoesis. *Representations, 74*, 83–104.
Faubion, J. (2001b). *The shadows and lights of Waco: Millennialism today*. Princeton: Princeton University Press.
Faubion, J. (2011). *An anthropology of ethics*. Cambridge: Cambridge University Press.
Faubion, J. (2014). Anthropologies of ethics: Where we've been, where we are, where we might go. *HAU: Journal of Ethnographic Theory, 4*, 437–442.
Foucault, M. (1975). *Discipline and punish: The birth of the prison*. London: Allen Lane.
Foucault, M. (1976). *The history of sexuality: Volume one, an introduction*. London: Allen Lane.
Foucault, M. (1985). *The use of pleasure: The history of sexuality* (Vol. 2). New York: Random House.
Foucault, M. (1986). *The care of the self: The history of sexuality* (Vol. 3). New York: Random House.
Foucault, M. (1997). The ethics of the concern for self as a practice of freedom. In P. Rabinow (Ed.), *Essential works of Michel Foucault: Ethics, subjectivity, and truth*. London: Allen Lane.
Geertz, C. (1984). Anti anti-relativism. *American Anthropologist, 86*, 263–278.
Giddens, A. (1984). *The constitution of society: Outline of the theory of structuration*. Cambridge: Polity Press.
Herskovits, M. (1972). *Cultural relativism: Perspectives on cultural pluralism*. New York: Random House.
Heywood, P. (2015). Freedom in the code: The anthropology of (double) morality. *Anthropological Theory, 15*, 200–217.
Hirschkind, C. (2001). The ethics of listening: Cassette-sermon audition in contemporary Cairo. *American Ethnologist, 28*, 623–649.
Hirschkind, C. (2006). *The ethical soundscape: Cassette sermons and Islamic counter-publics in Egypt*. New York: Columbia University Press.
Holbraad, M., & Pedersen, M. (2009). Planet M: The intense abstraction of Marilyn Strathern. *Anthropological Theory, 9*, 371–394.
Howell, S. (1997). *The ethnography of moralities*. London: Routledge.
Humphrey, C. (2007). Alternative freedoms. *Proceedings of the American Philosophical Society, 151*, 1–10.
Laidlaw, J. (1995). *Riches and renunciation: Religion, economy, and society amongst the Jains*. Oxford: Clarendon Press.
Laidlaw, J. (2002). For an anthropology of ethics and freedom. *The Journal of the Royal Anthropological Institute, 8*, 311–332.
Laidlaw, J. (2010a). *Social anthropology. In the Routledge companion to ethics*. London: Routledge.
Laidlaw, J. (Ed.). (2010b). *Ordinary ethics: Anthropology, language, and action*. New York: Fordham University Press.
Laidlaw, J. (2014a). *The subject of virtue: An anthropology of ethics and freedom*. Cambridge: Cambridge University Press.
Laidlaw, J. (2014b). Anthropology and cognitive science: A two-way street? *Anthropology of this Century, 9*. http://aotcpress.com/articles/anthropology-cognitive-science-twoway-street/.
Laidlaw, J. (2015). *The ethical condition: Essays on action, person, and value*. Chicago: University of Chicago Press.

Lambek, M. (2000). The anthropology of religion and the quarrel between poetry and philosophy. *Current Anthropology, 41*, 309–320.
Lambek, M. (2010). *Ordinary Ethics: Anthropology, Language and Action,* ed. Michael Lambek, Fordham University Press.
Lambek, M. (2015). *The Ethical Condition: Essays on Action, Person, and Value*, University of Chicago Press.
Luhrmann, T. (2012). *When God talks back: Understanding the American evangelical relationship with God*. New York: Knopf.
Luhrmann, T. (2013). 'Lord teach us to pray': Prayer affects cognitive processing. *Journal of Cognition and Culture, 13*, 159–177.
MacIntyre, A. (1981). *After virtue: A study in moral theory*. London: Duckworth.
Mahmood, S. (2001). Feminist theory, embodiment, and the docile agent: Some reflections on the Egyptian Islamic revival. *Cultural Anthropology, 16*, 202–235.
Mahmood, S. (2005). *The politics of piety: The Islamic revival and the feminist subject*. Princeton: Princeton University Press.
Marcus, G., & Fischer, M. (1986). *Anthropology as cultural critique: An experimental moment in the human sciences*. Berkeley: University of California Press.
Mead, M. (1928). *Coming of age in Samoa*. New York: Morrow.
Pandian, A. (2008). Tradition in fragments: Inherited forms and fractures in the ethics of South India. *American Ethnologist, 35*, 466–480.
Pandian, A. (2009). *Crooked stalks: Cultivating virtue in South India*. Durham, NC: Duke University Press.
Parkin, D. (1985). *The anthropology of evil*. Oxford: Blackwell.
Parry, J., & Bloch, M. (1989). *Money and the morality of exchange*. Cambridge: Cambridge University Press.
Pocock, D. (1986). The ethnography of morals. *International Journal of Moral and Social Studies, 1*, 3–20.
Robbins, J. (2004). *Becoming sinners: Christianity and moral torment in a Papua New Guinea society*. Berkeley: University of California Press.
Robbins, J. (2007). Between reproduction and freedom: Morality, value, and radical cultural change. *Ethnos, 72*, 293–314.
Robbins, J. (2009). Value, structure, and the range of possibilities: A response of Zigon. *Ethnos, 74*, 277–285.
Schielke, S. (2009). Being good in Ramadan: Ambivalence, fragmentation, and the moral self in the lives of young Egyptians. *Journal of the Royal Anthropological Institute*, S24–S40.
Scott, J. (1977). *The moral economy of the peasant: Rebellion and subsistence in South-East Asia*. New Haven, CT: Yale University Press.
Von Fürer-Haimendorff, C. (1967). *Morals and merit: A study of values and social controls in South Asian societies*. London: Weidenfeld and Nicholson.
Westermarck, E. (2000 [1932]). *Ethical relativity*. London: Routledge.
Williams, B. (2005). *In the beginning was the deed: Realism and moralism in political argument*. Princeton: Princeton University Press.
Wolfram, S. (1982). Anthropology and morality. *Journal of the Anthropological Society of Oxford, 13*, 262–274.
Yan, Y. (2011). How far can we move from Durkheim? Reflections on the new anthropology of morality. *Anthropology of this Century, 2*.http://aotcpress.com/articles/move-durkheim-reflections-anthropology-morality/.
Zigon, J. (2007). Moral breakdown and ethical demand: A theoretical framework for an anthropology of moralities. *Anthropological Theory, 7*, 131–150.
Zigon, J. (2009a). Within a range of possibilities: Morality and ethics in social life. *Ethnos, 74*, 251–276.
Zigon, J. (2009b). Phenomenological anthropology and morality: A reply to Robbins. *Ethnos, 74*, 286–288.

Cognitive and Neural Sciences: Investigating the Moral System

Tor Tarantola

How do we make moral decisions, and what factors affect how we make them? Cognitive science approaches these types of questions by attempting to describe and systematize the underlying processes that give rise to certain behaviors. If we think of the human mind as a kind of computer, the cognitive scientist wants to understand the types of inputs it accepts (such as auditory, visual, and other sensory information), how it processes these basic inputs to form useful pieces of information, and how that information generates behavior.

The computational neuroscientist David Marr (1982) proposed three levels of analysis when studying a cognitive system. At the top is the *computational level*—what problem is the mind trying to solve, and why? Below that is the *algorithmic level*—what computations does the mind perform in order to solve this problem? At the bottom is the *physical level*—how does the nervous system carry out these computations? The physical level is the domain of neuroscientists, who try to understand how the biology of neurons and neural systems make higher-level processes, such as moral decision-making, possible.

While the cognitive and neural sciences have made tremendous strides in understanding some of the more foundational capacities that humans exhibit—such as processing visual and auditory information, using language, and learning from feedback—its application to moral psychology is relatively new. In this chapter, I review some of the important research that has begun to form a cognitive and neural account of moral decision-making at all three levels of analysis: from the computational level (what moral behaviors do we carry out, and why?), to the algorithmic (how do certain factors affect our moral decisions?), to the physical (how do these processes play out in the brain?). I examine this work through the context of what I call the *moral system*.

T. Tarantola (✉)
Department of Psychology, University of Cambridge, Cambridge, UK
e-mail: tor.tarantola@gmail.com

B.G. Voyer, T. Tarantola (eds.), *Moral Psychology*,
DOI 10.1007/978-3-319-61849-4_5

The Moral System

The cognitive science of morality is especially challenging—and exciting—because it is fundamentally social. Unlike more basic processes, such as vision or hearing, moral decision-making by definition involves other people. By analogy, while the study of a single neuron is complex enough on its own, studying how populations of neurons interact in systems increases this complexity by orders of magnitude. Nevertheless, studying the behavior of a single neuron in isolation is meaningless if we don't consider how this implicates its function in the broader context of the nervous system. Similarly, understanding an individual person's moral decision-making processes is meaningless if we don't consider how these processes affect, and are affected by, broader social contexts.

An appropriately comprehensive cognitive science of morality, then, considers each element of a *moral system*. This system loosely comprises three principal levels: individuals, interactions between individuals, and groups (see Fig. 1).

At the top level, groups of individuals negotiate and define norms of proper behavior. This dynamic reaches beyond the moral context (e.g., setting fashion trends), but in a moral system, it establishes rules and expectations for how people should treat one another (of course, not always with unanimous agreement) and mechanisms for enforcing those rules (Horne 2001). For example, a society may establish a norm against inflicting unprovoked violence against one another. That same society may also establish a norm permitting abortion in certain circumstances, though with substantially less agreement among its members. Indeed, subsets of a society may form their own sets of norms, such as when a political party develops its platform.

At the middle level, we consider how subsets of individuals apply (or eschew) these norms when interacting with one another and how the nature of these interactions produces predictable outcomes. While these interactions can comprise different numbers and types of actors, the behaviors we're interested in fall into one of two broad categories: *adherence*, or whether a person respects a norm when interacting with others, and *enforcement*, or whether a person or group of people chooses to punish or otherwise compel another person to adhere to a norm. For example, let's say a rancher were to consider violating a norm by grazing her animals on communal land without contributing money toward the land's upkeep. How likely is she to make this decision, and how might the other ranchers respond? The dynamics of these interactions, and the aggregate outcomes for the people who enter into them, are what we study at this level.

At the bottom level, we consider how each person processes relevant inputs and generates the behaviors that compose the interactions at the middle level. We also consider how individuals perceive the content of the norms generated at the group level and what those norms demand in particular circumstances.

Each of these levels has traditionally been the focus of different empirical disciplines that come with their own tools and perspectives, such as sociology, political science, behavioral economics, experimental psychology, and cognitive neuroscience. But meaningfully understanding the moral system requires these disciplines to work

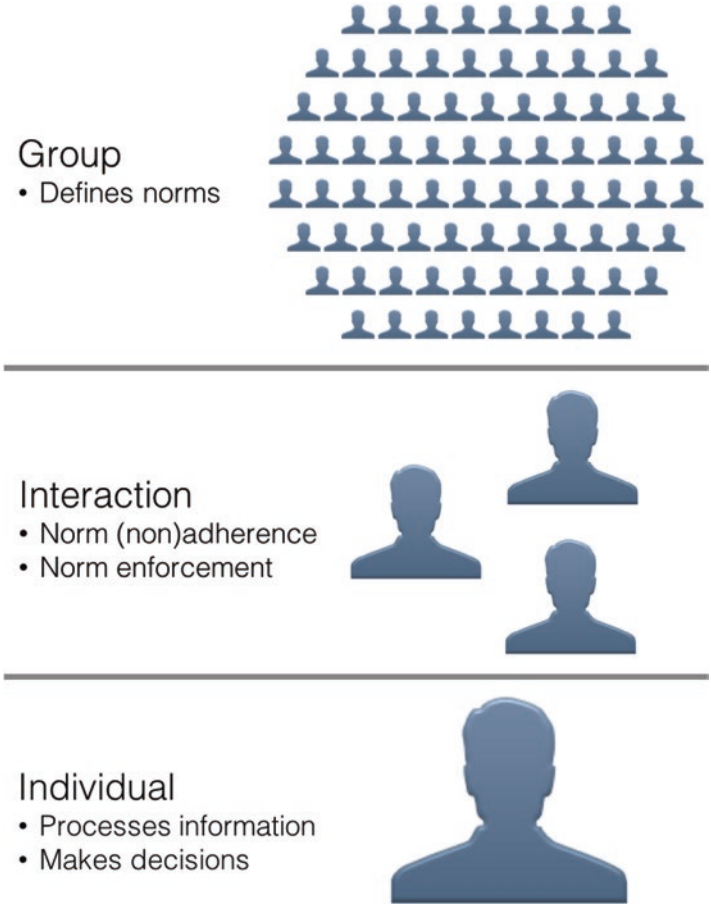

Fig. 1 The three levels of the moral system. Groups define norms, which are then applied (or eschewed) during interactions between individuals. These interactions might include norm adherence or nonadherence (e.g., cooperation or cheating) or enforcement (e.g., punishment). How individuals behave during these interactions is determined by how they process relevant information and make decisions

together (a principal motivation for this book). In the remainder of this chapter, I discuss a small sample of the work that has begun to illuminate each level of the moral system. I also suggest some areas for future research.

Defining Norms

It is widely believed by sociologists and other scholars that norms are *instrumental* (Hechter and Opp 2001)—that is, they exist to help groups of people maximize their collective welfare. Norms constrain individual behaviors that have the

potential to harm others. For example, we are prohibited from assaulting one another as a means of catharsis because the aggregate result of such behavior would be a decrease in overall welfare. Norms also encourage behaviors that provide a benefit to others at a low cost to the individual, such as calling an ambulance when someone is badly hurt.

The process by which norms emerge and evolve remains largely open to debate. What we do know is that norms vary in their prevalence across cultures (Henrich et al. 2006, 2010; Sober and Wilson 1998) and their level of consensus within cultures (Leung and Cohen 2011). If norms are rules designed to efficiently maximize a group's welfare in certain contexts, then it is conceivable that several different versions of a norm might achieve that goal equally well. Shaking hands when meeting a new person, for example, might work just as well as bowing. Such norms likely change stochastically over time as a result of new information, new ideas, and changes in environments that make existing norms less efficient (Ellickson 2001; Horne 2001). For example, as the dangers of second-hand smoke became more apparent, norms against smoking in public began to proliferate; the harms of inhaling second-hand smoke were greater, it came to be known, than the inconvenience of having to smoke outdoors (Ellickson 2001). Because the factors underlying the emergence of norms can vary, such norms may differ across groups due to differences in environments or artifacts of history.

Some norms, however, are ubiquitous and enduring (Robinson et al. 2007). Proscribing unprovoked violence, for example, is more common across cultures than shaking hands. In theory, the more common a norm, the more fundamental it is to collective welfare, the less dependent on the particulars of an environment, and the less susceptible to change. The basic constituents of these more common norms, such as altruistic tendencies or an aversion to harm experienced by others, may therefore have their foundations in our evolved biology (Robinson et al. 2007; Sober and Wilson 1998), which would guarantee more strongly our adherence to them.

Several sociologists have proposed theoretical models of how norms can change over time. For example, Horne (2001) proposes that norms emerge as a means of balancing individual interests against the interests of others. In cases where these interests conflict—that is, when a certain behavior accrues benefits to the actor at a cost to others, or vice versa—norms often emerge to regulate such behavior in a way that tends to maximize the group's welfare when broadly adopted. These norms are subject to change when the underlying costs and benefits change, for example, due to technological improvements or changes in the environment. The contents of particular norms are then transmitted informally through social networks, and sometimes formally through the enactment of laws. As anti-smoking norms propagated, for example, friends and family nudged each other to quit smoking, hosts asked their guests to smoke outside, and legislatures passed laws prohibiting smoking in public buildings (Ellickson 2001).

Fundamental to the successful transmission of norms is our reliance on information from other people when making judgments. In a foundational study, Asch (1951, 1956) placed unwitting participants in a room with several actors and asked them each to report which of the three lines drawn on a card was the longest. On

some trials, the actors—whom the lone participant thought were other participants—made obviously incorrect answers. This caused many participants to make the same incorrect answers, conforming to what they perceived as the group's judgment. Social information has also been shown to influence private value judgments. For example, Zaki et al. (2011) asked participants to rate the attractiveness of several faces and then told them how their peers rated the same faces. When asked to rate them a second time, participants made judgments that were more in line with their peers, despite these judgments being private.

This type of social conformity also affects how we glean the accepted norms in our immediate social environment, and consequently how we behave in relation to those norms. In the wake of the Second World War, some social psychologists turned their attention to a disturbing question: how could large groups of seemingly normal people carry out the kinds of atrocities committed by the Nazis? If average Germans were capable of it, does that mean anyone could be? In a famous series of experiments, Milgram (1963, 1965) asked participants to administer a series of increasingly powerful electric shocks to another person. Unbeknownst to the participants, the shocks were fake and the other person an actor. Despite the actor's screams and pleas to stop, a surprising number of participants followed the experimenter's instructions and continued administering what they believed were ever more powerful shocks. While the implications of these results are a continuing subject of debate, it seems that many participants adjusted their decision-making based on the moral balancing expressed by an authoritative person—that the scientific benefit of continuing the shocks outweighed the potential harm to the other person.

A few years later, in the Stanford prison experiment (Haney et al. 1972), 22 male participants were randomly assigned the roles of *guard* or *prisoner* in a makeshift prison in the basement of the Stanford psychology department. According to the researchers, these participants were "judged to be the most stable (physically and mentally), most mature, and least involved in anti-social behaviors" of the 75 men initially recruited (Haney et al. 1972, p. 7). Nevertheless, many guards quickly became abusive toward the prisoners. While physical violence was not allowed, "verbal affronts were used as one of the most frequent forms of interpersonal contact between guards and prisoners" (p. 20), and some guards "went far beyond their roles to engage in creative cruelty and harassment" (p. 21). The experiment was aborted after several prisoners experienced extreme emotional distress. A more recent variant of this experiment, the BBC Prison Study (Reicher and Haslam 2006), observed different patterns of behavior. Participants initially shunned their roles in favor of an egalitarian social order, but when they found this to be disorderly and unsustainable, they moved toward accepting a more hierarchical power structure in which guards would impose strict limits on prisoners' behavior. Interestingly, over the course of the experiment, participants' mean level of right-wing authoritarianism increased, as measured by a psychometric survey. This increase was driven by the participants who were initially lower in authoritarianism but whose attitudes conformed to those of the participants who advocated a more draconian regime (Reicher and Haslam 2006).

These experiments implicate several facets of human psychology, from personality to social identity to obedience. But they also demonstrate something fundamental about the dynamic nature of social norms. In each case, participants deviated from their typical behavior to conform to the new norms in their immediate environments—even exhibiting pathologically aggressive behavior in some instances (Haney et al. 1972). In the BBC Prison Study, the group norm itself seemed to shift from egalitarianism to authoritarianism once they believed the latter to be more effective—a shift that was also reflected in the individual attitudes of the participants (Reicher and Haslam 2006). The process by which these shifts take place, and how individuals affect and respond to them, is an area of research where much work remains to be done.[1]

Interactions

While the content and dynamics of social norms are complex, experimental psychology has recently begun investigating how individuals interact in situations that pit personal interests against the interests of others.

Underlying most forms of human interaction is the norm of *conditional cooperation* (Brandts and Schram 2001; Fehr and Fischbacher 2004a, b; Fischbacher et al. 2001; Keser and van Winden 2000). Humans are social and cooperative, meaning that we pursue goals through mutually beneficial interactions, rather than by operating in isolation. However, adherence to this norm is conditional on potential cooperation partners also adhering to it. In many cases, cooperative interactions give each individual the opportunity to cheat, gaining an increase in her own payoff by failing to meet an obligation to her cooperation partners. Returning to our earlier example, a rancher who grazes her cattle on communal land while not contributing to its upkeep unfairly free rides on the other ranchers. This raises a challenge—why would we choose to cooperate if we would be better off cheating? Why would the rancher pay her share when she could profit more from free riding?

The persistence of human cooperation undergirds the successful functioning of human civilization, yet there appear to be strong incentives for individuals *not* to cooperate. Evolutionary theorists have proposed different accounts of how prosocial instincts might be explained by natural selection. In other words, how might a tendency to cooperate—rather than cheat—lead to more successful survival and reproduction? Hamilton (1963, 1964a, b) offered a mathematical description of how individuals might behave selflessly toward their kin as a means of helping to propagate their genes. Making a significant sacrifice for one's child, in other words, would

[1] An important, related line of research concerns *self-construal*, or the extent to which individuals view themselves as independent from others in their society (Markus and Kitayama 1991). Different self-construals may lead to different levels of perceived *agency*, or the extent to which a person sees herself (or is seen by others) as having control over her actions, and how much of this control can be attributed to others in her social environment. For more, see Doris (2015), Voyer and Franks (2014), and Franks and Voyer (in press, reviewing Doris 2015).

benefit a person's genes if not her own survival. However, this theory does not fully explain the extent to which people behave prosocially toward strangers. Some theorists have attempted to explain this with theories of direct (Trivers 1971) and indirect reciprocity (Nowak and Sigmund 1998). Under indirect reciprocity, individuals who gain a reputation for acting prosocially toward others will benefit by attracting more cooperation partners, even if the beneficiaries of their behavior do not directly reciprocate. (See Krasnow, this volume, for a more thorough discussion.)

Evolutionary biologists often use computer simulations to test the viability of behavioral traits and whether they might reasonably stand up to the pressures of natural selection (e.g., Jordan et al. 2016; Nowak and Sigmund 1998). By creating simulated individuals and societies, and seeing how well those individuals are able to survive and reproduce with certain traits, they can test whether these traits might eventually become prevalent in the population. Such models have indicated that persistent cooperative behavior between individuals in groups can be maintained under certain conditions (Chalub et al. 2006; Fehr and Fischbacher 2004b; Nowak and Sigmund 1998; Ohtsuki and Iwasa 2006; Rand and Nowak 2013). First, each individual has a *reputation* for being a good or bad interaction partner. In general, we prefer to interact with people who have good reputations—and are less likely to cheat—and to avoid people with bad reputations, who are more likely to take advantage of us. In this way, people have incentives to cooperate in order to maintain good reputations and therefore have an easier time seeking cooperation partners in the future (Chalub et al. 2006; Nowak and Sigmund 1998; Ohtsuki and Iwasa 2006). Second, because otherwise well-intentioned people will sometimes behave badly, they should be inclined to apologize when making errors, and others should be inclined to forgive them (Ohtsuki and Iwasa 2006). This avoids the unnecessary shunning of people who may be good cooperation partners in general, but who may occasionally lapse in their moral judgments (Ohtsuki and Iwasa 2006). Lastly, a tendency to detect and punish cheaters is vital to maintaining cooperation, as it acts as an important disincentive (Fehr and Fischbacher 2004a; Fehr and Gächter 2002; Ohtsuki and Iwasa 2006; Ostrom et al. 1992).

Several studies have used economic games to investigate the effect that such punishment has on maintaining cooperation (Fehr and Fischbacher 2004a, b; Fehr and Gächter 2000, 2002; Ostrom et al. 1992; Yamagishi 1986, 1988). One such game—the *public goods game*—is an experiment in which participants are each given a sum of real money and then asked to contribute a portion of that money to a common pot. The pot is then multiplied by a certain number (between 1 and the number of participants in the group) and distributed to each participant. Each participant can see how much each other participant contributed to the pot. Importantly, each participant has an incentive to cheat (i.e., receive a portion of the contributions without contributing anything herself); but the entire group will benefit the most if everyone contributes the maximum amount. The challenge here is to figure out how to maximize cooperation, thereby maximizing the group's welfare. Several studies have shown that, when participants are given the ability to punish one another (reducing other participants' payments), contributions increase dramatically (Fehr and Gächter 2000, 2002; Ostrom et al. 1992; Yamagishi 1986, 1988; see Fig. 2).

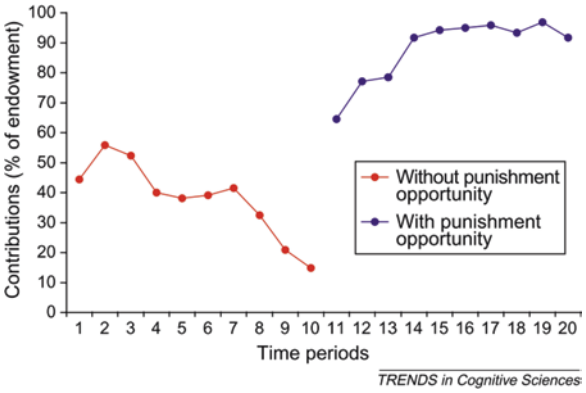

Fig. 2 In the first ten time periods of a public goods game, participants are not able to punish free riders, and the average contribution decreases precipitously. In the next ten time periods, after they are told they can punish one another, cooperation increases dramatically, leveling off near 100%. Reprinted from Trends in Cognitive Sciences, Vol. 8, Ernst Fehr and Urs Fischbacher, Social norms and human cooperation, Pages No. 185–190, Copyright (2004), with permission from Elsevier. This figure redraws data published in the American Economic Review, Vol. 90, Ernst Fehr and Simon Gächter, Cooperation and punishment in public goods experiments, Pages No. 980–994, Copyright (2000), with permission from the American Economic Association

Punishment can take two forms: second-party punishment, in which a person punishes a wrongdoer for causing her personal harm, and third-party punishment, in which a person punishes a wrongdoer for causing harm to someone else. Laboratory experiments have investigated both types of punishment and have found that many people are willing to absorb a cost to punish wrongdoers, even in the third-party context, when they are not themselves harmed by the wrongdoer's conduct (Fehr and Fischbacher 2004a).

In one common experimental paradigm known as the *ultimatum game* (Güth et al. 1982), one participant is given a fixed amount of money by the experimenter and allowed to share it with her partner, another participant with whom she has been randomly paired. The participant then makes an offer to her partner of an amount between 0 and 100% of the total sum, and the partner is asked whether she accepts it. If she accepts the offer, she gets to keep the amount that the first participant decided to share. But if she rejects the offer, neither participant receives anything. In many cases, if the first participant's offer is fairly low, her partner rejects the offer, despite that rejection being costly. In other words, the partner gives up the opportunity to keep the money in order to punish the first participant for making an unfairly stingy offer. This behavior, observed across cultures (Henrich et al. 2006), demonstrates the value that people place on fairness and the costs they are willing to absorb to carry out second-party punishment.

In a variation of the ultimatum game called the *dictator game* (Kahneman et al. 1986), the recipient has no power but must passively accept the offer made by the benefactor (for review and meta-analysis, see Engel 2011). Fehr and Fischbacher (2004a) ran a version of the dictator game in which they included a third participant to observe the interaction. This third participant was then allowed to spend a little of her own money to reduce the benefactor's payment if she believed the offer was unfair, even though she was personally unaffected. A surprising number of third parties—more than 60%, depending on the fairness of the offer—were willing to spend some of their own money to punish the benefactor on behalf of the recipient. This type of costly third-party punishment, often called *altruistic punishment*, can seem puzzling. Why would someone absorb a cost to themselves to punish someone who had caused them no harm?

One theory is that altruistic punishment signals trustworthiness to others. Jordan et al. (2016) have argued that trustworthy people—those who would rather cooperate than cheat—derive a greater benefit from punishing transgressors than it costs them to punish. Since trustworthy people benefit generally from deterring cheaters and promoting cooperation, the costs of punishment to them will be lower than for cheaters (Jordan et al. 2016; Rand and Nowak 2013). Because this cost is lower, trustworthy people might punish transgressors in order to differentiate themselves from non-trustworthy people, for whom punishing would be costlier (Jordan et al. 2016). Potential cooperative partners might then view the act of punishment as a signal of the punisher's trustworthiness, since they would only punish if the cost were low enough, and the cost would be low enough only if the person were trustworthy (Jordan et al. 2016). They argue that others are consequently more likely to cooperate with punishers than with non-punishers, providing benefits that further offset the cost of punishing.[2]

This body of work has begun to shed light on the complex dynamics of moral interactions between individuals and to demonstrate how these interactions might have evolved and sustained themselves. When considered together, the countless interactions of this type define how a society functions. Can we count on one another to be trustworthy? Will transgressors be reliably punished? As societies have grown larger, we often rely on our institutions to help maintain the conditions for cooperation (e.g., by enforcing contracts or punishing free riding). One of the principal challenges for our institutions is how best to structure interactions to encourage prosocial and deter antisocial behavior. Combining experimental evidence with simulation techniques can be helpful here. By predicting how people would respond to certain incentives and disincentives—and simulating the large-scale outcomes of these responses—researchers can help identify which regulations or punishments would maximize the general welfare.

[2] Interestingly, when third parties have the opportunity to spend some money to help the victim, they are less likely to punish the transgressor, and punishment serves as less of a signal (Jordan et al. 2016).

Individual Processing

The bottom level of the moral system is concerned with how individuals process, interpret, and use relevant inputs to make moral decisions. During each of the interactions described above, several individuals receive information and make decisions. What type of information is important? How is this information transformed into behavior? And how does this transformation take place in the brain?

Each time we make a moral decision—whether or not to cheat, or whether or not to punish, for example—several factors could potentially enter our calculation. How much would I benefit from cheating? How much would I be punished if I were discovered? How much might my actions harm others? This kind of calculation might seem cold, but some kind of calculation must underlie every type of decision, even moral ones. This is what David Marr (1982) called the *algorithmic level* of cognitive science. Figuring out the relevant inputs, and how those inputs are transformed into behaviors, is one important puzzle that the cognitive science of moral systems tries to solve (Buckholtz 2015; Crockett 2016; Cushman 2015; Hutcherson et al. 2015). This isn't to say that every person deliberatively weighs the costs and benefits of each decision before she makes it. A *cost* to cheating, for example, might not be material, but rather a feeling of guilt or shame that goes along with breaking a moral rule.[3] A *benefit* to punishing a transgressor might be a feeling that justice has been done, rather than a deliberative calculation about how it might enhance one's reputation. These decisions might feel more or less automatic, and our cognitive architecture might be structured in a way that predisposes us to cooperation fairly quickly (Hutcherson et al. 2015; Krajbich et al. 2015a; Rand et al. 2012). But there are still calculations being made, even implicitly, behind every moral decision. One exciting challenge in moral psychology is figuring out what those calculations are, and a powerful tool for doing this is the simple behavioral experiment. By carefully changing inputs and observing how these changes affect behavioral outputs, we can begin to sketch the contours of the hidden calculations that people perform.

Norm Adherence

One common type of experiment in moral psychology uses versions of the *trolley problem*, a type of thought experiment originating in philosophy (Foot 1967), to investigate how people decide whether to adhere to certain norms. A classic trolley problem, the *switch dilemma*, goes like this: You see a runaway train barreling down the track toward five innocent bystanders, all of whom would be struck and killed instantly. You can pull a railroad switch, which would set the train down a different

[3] The role of emotion in moral decision-making is a topic of ongoing debate. For more, see Huebner et al. (2009), Moll et al. (2005), Krasnow (this volume), and Nicoletti and Delehanty (this volume).

track, avoiding the five bystanders but killing a single bystander on the other track. Do you pull the switch (Foot 1967)?

This type of scenario is designed to pit a person's concern for consequences (saving an additional four lives) against her concern for moral rules (don't take actions that cause others harm). Several studies have used variants of this dilemma to tease apart which factors predict how people will choose whether to take an action. One common variant is the *footbridge dilemma*, in which participants are asked whether they would push a fat man off a footbridge to block the oncoming train from striking the five bystanders (Thomson 1985). The costs and benefits are putatively the same (one life in exchange for five), as is the moral rule being broken (don't cause others harm). Nevertheless, while a majority of people would pull the switch, most people would not push the man off the bridge (Greene et al. 2001). What accounts for this difference? Several psychological theories have been advanced, suggesting that a key difference may be that pushing is more direct and personal than pulling a switch or that pushing someone to his death is a more intentional harm than causing him to die as a side effect of diverting the train (for review, see Cushman and Souza 2013). One study found that participants' responses were affected by how vividly the action's harm was described as well as the number of lives that would be saved (Bartels 2008).

The ongoing debates over the computations underlying these types of moral decisions highlight a challenge of using dilemmas like the trolley problem to investigate them. As small stories, these dilemmas contain details laced with meanings that may differ from person to person. They also contain several details that differ between the stories themselves that make it difficult to precisely pin down which factors provide the inputs that are important in the computations that can generate different decisions. Aside from the numbers of victims and averted victims, these stories also tend not to include parametric variables—that is, variables that can be easily quantified and therefore manipulated numerically in order to measure their precise effects on behavior. This makes it difficult to discover the computations underlying moral decisions with much specificity.

Some recent work has begun using more computational methods to sketch the cognitive processes underlying decisions about when and how to adhere to norms. Many of these studies focus on how we value other people's interests in relation to our own. Some innovations in this area came from economists (e.g., Bolton and Ockenfels 2000; Fehr and Schmidt 1999) who aimed to explain moral decisions in terms of individuals' *utility*—the value that a person attaches to objects or outcomes, which can differ from person to person. A more cooperative or trustworthy person, for example, might attach greater utility to equity, and this utility might be greater than the utility gained from cheating (Fehr and Schmidt 1999). These models propose equations that explain how a person might balance one consideration against another when making a decision. This way of analyzing moral decision-making has the advantage of allowing a single person's decision to change as the relative magnitudes of opposing factors change. A person who might normally cooperate, for example, might cheat if the relative benefits were high enough. (More people would probably steal a candy bar if someone paid them a million

dollars to do it.) It also allows us to describe differences between people more precisely. Rather than just describing someone as untrustworthy, for example, we can specify quantitatively how much value they place on their own material gain relative to the potential harm caused to others by their actions. In a series of experiments, Crockett and colleagues measured how much money participants would be willing to forego to avoid painful shocks to themselves and others (Crockett et al. 2014, 2015, 2017). By varying the amount of money and the number of shocks over several trials, they were able to estimate a parameter to describe how each participant valued avoiding shocks to herself versus shocks to another person. (They found that most participants generously valued avoiding shocks to others more than avoiding shocks to themselves.)

Other recent work has used computational techniques from other areas of cognitive science. One particularly promising approach is the use of sequential sampling models of decision-making, such as the *drift diffusion model*, which can predict both choices and response times (Ratcliff and McKoon 2008; Ratcliff and Rouder 1998; Smith and Ratcliff 2004). The drift diffusion model imagines that each choice results from a single particle drifting toward one of two decision boundaries, with each boundary corresponding to one of two options (see Fig. 3). Once the particle reaches one of the boundaries, a response is made. The average speed at which the particle moves toward a boundary—the *drift rate*—is proportional to the relative strength of the evidence in favor of that option. The more obvious the choice, the higher the drift rate and the faster the response. This makes intuitive sense—when choosing between your favorite chocolate and a food you hate, you'll make the choice quickly. But when choosing between two of your favorite snacks, you'll probably take longer to decide.

Importantly, the drift diffusion model also assumes a certain amount of noise in the particle's drift, as illustrated by its jagged path in Fig. 3. While the particle tends to drift toward the boundary corresponding to the choice with greater evidence, sometimes this noise will push the particle off course, causing it to reach the other boundary instead. This noise means that responses are stochastic—people choose options with stronger evidence more often, but not always. The proportion of "correct" responses (choosing options with greater average evidence) to "incorrect" responses will increase as the drift rate increases. These types of models, while a useful way to visualize and think about decision processes, have also been shown to provide good descriptions of both choice and response time data, as well as neural activity (for review, see Forstmann et al. 2016). This added level of detail makes these models particularly powerful, allowing researchers to probe in greater detail how each relevant factor affects the neural and cognitive processes underlying each decision.

The drift diffusion model has recently been applied to moral decision-making (Hutcherson et al. 2015; Krajbich et al. 2015b). In one recent study (Hutcherson et al. 2015), participants' brains were scanned using functional magnetic resonance imaging (fMRI) as they made several decisions about how to share money with another anonymous participant. (fMRI indirectly measures activity in different parts of the brain by detecting changes in blood flow.) Participants could

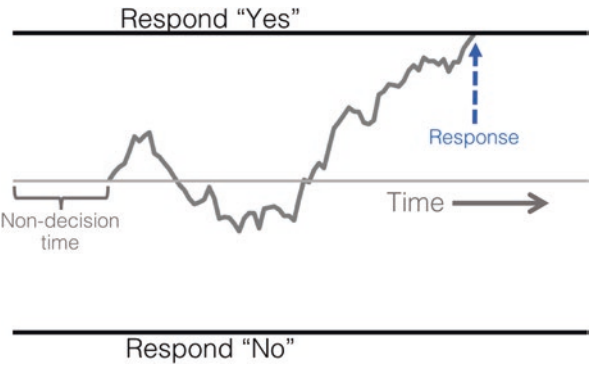

Fig. 3 The drift diffusion model. After a non-decision time to allow for stimulus processing, a decision particle moves toward one of two response thresholds at a rate proportional to the relative strength of evidence for that response. The path of the particle is subject to noise, indicated by its jagged path to the threshold. When the particle reaches the threshold, a response is made

accept more money in exchange for the other person receiving less, or vice versa, with the amounts varying from trial to trial. The authors tested a model in which the drift rate underlying each decision was determined by adding the potential change in the participant's earnings to the potential change in the other person's earnings. They allowed these "self" and "other" changes in earnings to be weighted differently—in other words, participants might weigh a gain to themselves more heavily than an equivalent loss to the other person. (In fact, this is what they found.) Their model provided a reasonably accurate description of participants' choices and response times. They also found that activity in different brain regions correlated with different components of the model. Specifically, activity in regions of the ventral striatum and the ventromedial prefrontal cortex (vmPFC) correlated with the amount to be received by the participant, while activity in the right temporoparietal junction (TPJ) and a smaller region of the vmPFC correlated with the total amount to be received by the other person (see Fig. 4). Activity in the TPJ has been suggested in many other studies to represent, among other things, other people's mental states (for review, see Abu-Akel and Shamay-Tsoory 2011). Several other studies of decision-making have suggested that the vmPFC may play a key role in representing the value of different options (for review, see Platt and Plassmann 2014). Hutcherson et al. (2015) suggest that the vmPFC may integrate concern for the self and concern for others—represented in distinct neural networks—into a single decision value that precipitates a choice.

These recent advances demonstrate the power of applying perspectives and methods from other fields to moral psychology. By introducing insights from non-moral decision-making, researchers have helped test the viability of different psychological theories (e.g., Krajbich et al. 2015a) and begun to develop a more nuanced picture of how people choose whether to adhere to social norms.

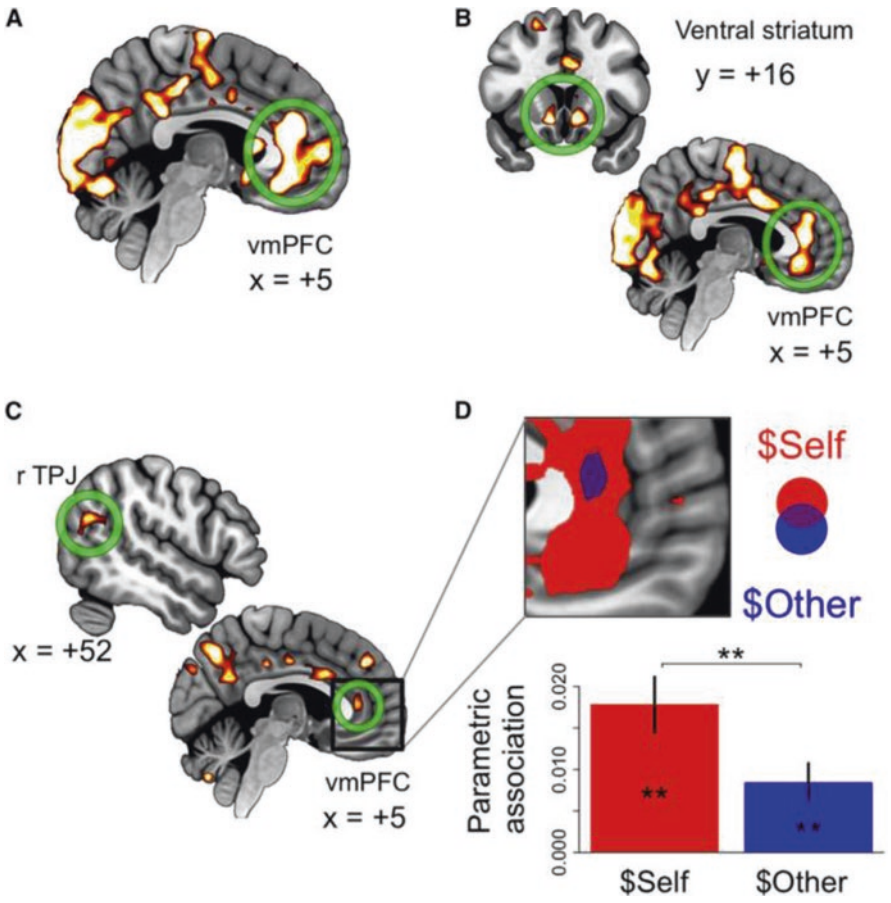

Fig. 4 (**a**) Activity in the vmPFC, at the time the choice was made, correlated with the participant's choice itself (on a four-point scale from *strong no* to *strong yes*). (**b**) Activity in the vmPFC also correlated with potential payments to the participant ("$Self") and (**c**) to the other person ("$Other"), but to different degrees (**d**). Reprinted from Neuron, Vol. 87, Cendri A. Hutcherson, Benjamin Bushong, and Antonio Rangel, A neurocomputational model of altruistic choice and its implications, Pages No. 451–462, Copyright (2015), with permission from Elsevier

Norm Enforcement

Another important category of moral behavior is *norm enforcement*—in particular, how individuals decide whether, and how much, to punish someone for violating a norm. In modern societies, punishment might seem to be an exclusively governmental function. After all, when someone is convicted of a crime, it's the government—not private citizens—who prosecutes and punishes. Even when private citizens bring lawsuits against one another, the court system is responsible for managing the process of adjudication and enforcing its outcomes. But the rules and principles

governing how punishment is administered are still decided by human beings, such as judges and lawmakers, and are often informed by popular intuitions.

What are the factors that inform how we think about punishment? Two significant factors are the *intent of the actor* (did she believe her action would bring about a harm, and did she want the harm to come about?) and the *consequences of her action* (how much did she cause the harm, and how bad did the harm turn out to be?) (Cushman 2008; Fincham and Jaspars 1983; Shaver 1985; Weiner 1995). The importance of these two factors can be seen in how criminal sentences are administered. The US federal sentencing guidelines, for example, suggest that a conviction for premeditated murder should get you life in prison. But if you attempt to commit the murder and fail—say, your gun jams—the guidelines recommend between about 11 and 14 years if you have no prior convictions. On the other hand, if you happen to kill someone unintentionally by driving recklessly, you should get between about 3 and 4 years (United States Sentencing Commission 2016). As these dramatic differences show, how much someone is punished is determined not just by her intent or the harm she caused but by an integration of the two.

In a large-scale survey study, Cushman (2008) set out to quantify how much intent and consequences matter when people make moral judgments about others' actions. In one experiment, using a set of hypothetical scenarios, he varied (1) whether the actor believed her action would cause harm, (2) whether she wanted to cause harm, and (3) whether her action actually caused harm. Participants were then asked how permissible each action was and how much the actor should be punished. When it came to permissibility, he found that 84% of the variation in responses for each scenario could be explained by intentional factors, with only 3% depending on the consequence (see Fig. 5). By contrast, punishment ratings were a combination of intentional factors (68%) and the consequence (20%).

Other work has used techniques from cognitive neuroscience to investigate how these different factors might be processed in the brain. Young and colleagues (Young et al. 2007, 2010; Young and Saxe 2009) have found evidence that the right TPJ plays an important role in representing information about an actor's intent. This region may work by suppressing the influence of other regions—such as the amygdala, which is sensitive to the severity of a harm—when considering how to punish unintentional harms (Treadway et al. 2014). These and similar studies have allowed researchers to begin sketching the neural circuitry underlying individual decisions about how to enforce others' adherence to norms (see Fig. 6; Buckholtz and Marois 2012a, b).

Research exploring the computations underlying third-party punishment, while in its relative infancy, has begun to provide some insights into which factors may be important and why. Information about an actor's mental state, for example, is clearly a significant factor in punishment intuitions—perhaps to facilitate forgiveness, which evolutionary models have found to be important to sustaining cooperation (Ohtsuki and Iwasa 2006). Better understanding the nuances of punishment intuitions, therefore, can yield new insights into the nature of cooperative interactions and the emergence of legal norms. Ultimately, these insights could help us improve the effectiveness of justice policy.

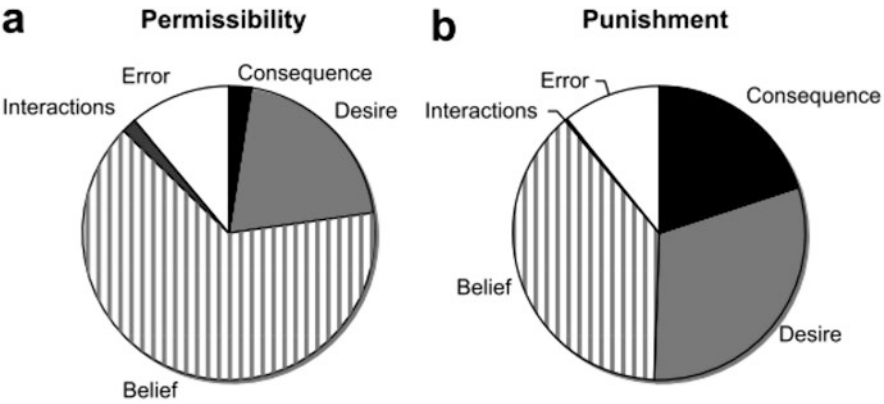

Fig. 5 Permissibility ratings (**a**) were chiefly determined by intentional factors (the actor's belief and desire), while punishment (**b**) combined intentional factors with the consequence of the action. Reprinted from Cognition, Vol. 108, Fiery Cushman, Crime and punishment: Distinguishing the roles of causal and intentional analyses in moral judgment, Pages No. 353–380, Copyright (2008), with permission from Elsevier

Moving Forward

This chapter surveyed only a fraction of the important research that has begun using the powerful tools of the cognitive and neural sciences to investigate the moral system. While much work has been done, much remains. Especially exciting is the potential for researchers to synthesize new knowledge at different levels of this system. How might different selfishness parameters (Hutcherson et al. 2015) influence the nature of interactions? How might these interactions, in the aggregate, affect the well-being of a society? How do intuitions about punishment influence legal norms, and how does the law influence our intuitions?

At the heart of cognitive science is a spirit of interdisciplinarity. Computational modeling can help move neuroimaging research from a neuroanatomical enterprise—locating where different factors are represented in the brain (see Machery and Doris, this volume, for a critique)—to a more nuanced investigation of brain systems (Behrens et al. 2009). Likewise, neuroimaging can help researchers distinguish between different potential models of behavior (see Li and Daw 2011, for an example from outside moral psychology). Simulation techniques, such as those used in evolutionary dynamics, can help test which cognitive processes might have evolved over time (e.g., Jordan et al. 2016). They can also help to quantify their broader social implications. For example, how might different punishment policies interact with observed distributions of selfishness parameters to affect the level of cheating in a society? By exploring and quantifying the implications of cognitive processes, in addition to the processes themselves, researchers can help inform public policy as well as science.

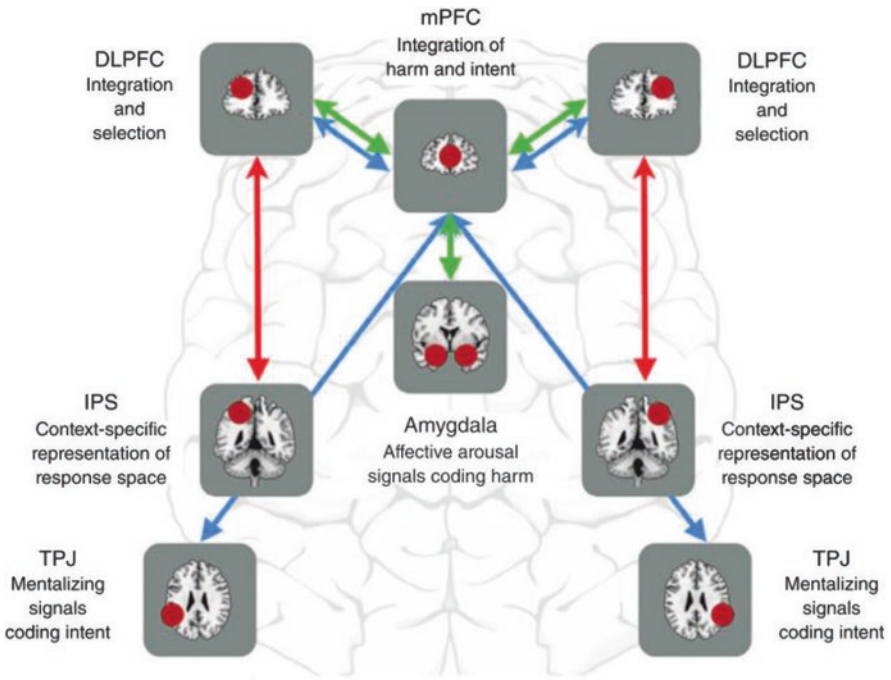

Fig. 6 Hypothesized model of the neural network underlying third-party punishment intuitions. Buckholtz and Marois (2012a) suggest that the medial prefrontal cortex (mPFC) may integrate the TPJ's representations of an actor's intent with the amygdala's emotional representation of the harm. The dorsolateral prefrontal cortex (DLPFC) may then precipitate a specific response after integrating the mPFC's representation with a representation of the range of potential responses from the intraparietal sulcus (IPS). While these types of models illustrate the major network nodes potentially involved in the representations of different information, in reality, each representation likely implicates a more complex and distributed network of neurons. Reprinted from Nature Neuroscience, Vol. 15, Joshua W. Buckholtz and René Marois, The roots of modern justice: Cognitive and neural foundations of social norms and their enforcement, Pages No. 655–661, Copyright (2012), with permission from Macmillan Publishers Ltd

As this volume attests, moral psychology is richly multifaceted and ripe for exploration by a diverse set of disciplines and methods. This is part of what makes it such an exciting enterprise—it operates at different levels of analysis, each of which interacts in complex ways. I tried here to delineate three levels of what I've called the *moral system*—this is less a theoretical proposal than a framework for beginning to define a research agenda for a complicated and vital field. Studying any biological phenomenon is incomplete without considering it at several levels of analysis. Studying cells would be meaningless without understanding the organisms they compose, and studying organisms would be meaningless without understanding their ecosystems. By combining insights and techniques from different disciplines to better understand moral psychology at all its levels, we can begin to sketch with greater detail the system that defines so much of human life.

References

Abu-Akel, A., & Shamay-Tsoory, S. (2011). Neuroanatomical and neurochemical bases of theory of mind. *Neuropsychologia, 49*, 2971–2984. doi:10.1016/j.neuropsychologia.2011.07.012.

Asch, S. E. (1951). Effects of group pressure upon the modification and distortion of judgments. In H. Guetzkow (Ed.), *Groups, leadership, and men* (pp. 222–236). Pittsburgh: Carnegie Press.

Asch, S. E. (1956). Studies of independence and conformity: I. A minority of one against a unanimous majority. *Psychological Monographs: General and Applied, 70*, 1–70.

Bartels, D. M. (2008). Principled moral sentiment and the flexibility of moral judgment and decision making. *Cognition, 108*, 381–417. doi:10.1016/j.cognition.2008.03.001.

Behrens, T. E. J., Hunt, L. T., & Rushworth, M. F. S. (2009). The computation of social behavior. *Science, 324*, 1160–1164.

Bolton, G., & Ockenfels, A. (2000). ERC: A theory of equity, reciprocity, and competition. *The American Economic Review, 90*, 166–193.

Brandts, J., & Schram, A. (2001). Cooperation and noise in public goods experiments: Applying the contribution function approach. *Journal of Public Economics, 79*, 399–427. doi:10.1016/S0047-2727(99)00120-6.

Buckholtz, J. W. (2015). Social norms, self-control, and the value of antisocial behavior. *Current Opinion in Behavioral Sciences, 3*, 122–129. doi:10.1016/j.cobeha.2015.03.004.

Buckholtz, J. W., & Marois, R. (2012a). The roots of modern justice: Cognitive and neural foundations of social norms and their enforcement. *Nature Neuroscience, 15*, 655–661. doi:10.1038/nn.3087.

Buckholtz, J. W., & Marois, R. (2012b). The roots of modern justice: Cognitive and neural foundations of social norms and their enforcement. *Nature Neuroscience, 15*, 655–661. doi:10.1038/nn.3087.

Chalub, F. A. C. C., Santos, F. C., & Pacheco, J. M. (2006). The evolution of norms. *Journal of Theoretical Biology, 241*, 233–240. doi:10.1016/j.jtbi.2005.11.028.

Crockett, M. (2016). How formal models can illuminate mechanisms of moral judgment and decision making. *Current Directions in Psychological Science, 25*, 85–90.

Crockett, M. J., Kurth-Nelson, Z., Siegel, J. Z., Dayan, P., & Dolan, R. J. (2014). Harm to others outweighs harm to self in moral decision making. *Proceedings of the National Academy of Sciences, 111*, 17320–17325. doi:10.1073/pnas.1408988111.

Crockett, M. J., Siegel, J. Z., Kurth-nelson, Z., Grosse-rueskamp, J. M., Dayan, P., Dolan, R. J., Crockett, M. J., Siegel, J. Z., Kurth-nelson, Z., Ousdal, O. T., Story, G., & Frieband, C. (2015). Dissociable effects of serotonin and dopamine on the valuation of harm in moral decision making. *Current Biology, 25*, 1852–1859.

Crockett, M. J., Siegel, J. Z., Kurth-Nelson, Z., Dayan, P., & Dolan, R. J. (2017). Moral transgressions corrupt neural representations of value. *Nature Neuroscience, 20*, 879–885. doi:10.1038/nn.4557.

Cushman, F. (2008). Crime and punishment: Distinguishing the roles of causal and intentional analyses in moral judgment. *Cognition, 108*, 353–380. doi:10.1016/j.cognition.2008.03.006.

Cushman, F. (2015). From moral concern to moral constraint. *Current Opinion in Behavioral Sciences, 3*, 58–62. doi:10.1016/j.cobeha.2015.01.006.

Cushman, F., & Souza, D. O. (2013). Action, outcome, and value: A dual-system framework for morality. *Personality and Social Psychology Review, 17*, 273–292. doi:10.1177/1088868313495594.

Doris, J. M. (2015). *Talking to our selves: Reflection, ignorance, and agency*. Oxford, UK: Oxford University Press.

Ellickson, R. C. (2001). The evolution of social norms: A perspective from the legal academy. In M. Hechter & K.-D. Opp (Eds.), *Social norms* (pp. 35–75). New York: Russell Sage Foundation.

Engel, C. (2011). Dictator games: A meta study. *Experimental Economics, 14*, 583–610. doi:10.1007/s10683-011-9283-7.

Fehr, E., & Fischbacher, U. (2004a). Third-party punishment and social norms. *Evolution and Human Behavior, 25*, 63–87. doi:10.1016/S1090-5138(04)00005-4.

Fehr, E., & Fischbacher, U. (2004b). Social norms and human cooperation. *Trends in Cognitive Sciences*. doi:10.1016/j.tics.2004.02.007.

Fehr, E., & Gächter, S. (2000). Cooperation and punishment in public goods experiments. *The American Economic Review, 90*, 980–994. doi:10.1257/aer.90.4.980.

Fehr, E., & Gächter, S. (2002). Altruistic punishment in humans. *Nature, 415*, 137–140. doi:10.1038/415137a.

Fehr, E., & Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation. *Quarterly Journal of Economics, 114*, 817–868. doi:10.1162/003355399556151.

Fincham, F. D., & Jaspars, J. M. (1983). A subjective probability approach to responsibility attribution. *The British Journal of Social Psychology, 22*, 145–161.

Fischbacher, U., Gächter, S., & Fehr, E. (2001). Are people conditionally cooperative? Evidence from a public goods experiment. *Economics Letters, 71*, 397–404. doi:10.1016/S0165-1765(01)00394-9.

Foot, P. (1967). The problem of abortion and the doctrine of double effect. *Oxford Review*, 5–15.

Forstmann, B. U., Ratcliff, R., & Wagenmakers, E.-J. (2016). Sequential sampling models in cognitive neuroscience: Advantages, applications, and extensions. *Annual Review of Psychology, 67*, 641–666. doi:10.1146/annurev-psych-122414-033645.

Franks, B., & Voyer, B.G. (in press). What does agency afford the self? A review of Talking to Our Selves: Reflection, ignorance, and agency. *The Behavioral and Brain Sciences*.

Greene, J. D., Sommerville, R. B., Nystrom, L. E., Darley, J. M., & Cohen, J. D. (2001). An fMRI investigation of emotional engagement in moral judgment. *Science, 293*, 2105–2108.

Güth, W., Schmittberger, R., & Schwarze, B. (1982). An experimental analysis of ultimatum bargaining. *Journal of Economic Behavior and Organization, 3*, 367–388. doi:10.1016/0167-2681(82)90011-7.

Hamilton, W. D. (1963). The evolution of altruistic behavior. *The American Naturalist, 97*, 354–356. doi:10.1086/497114.

Hamilton, W. D. (1964a). The genetical evolution of social behaviour. I. *Journal of Theoretical Biology, 7*, 1–16. doi:10.1016/0022-5193(64)90038-4.

Hamilton, W. D. (1964b). The genetical evolution of social behaviour. II. *Journal of Theoretical Biology, 7*, 17–52. doi:10.1016/0022-5193(64)90039-6.

Haney, C., Banks, C., & Zimbardo, P. (1972). Interpersonal dynamics in a simulated prison (NTIS No. AD-751 041). Stanford, CA.

Hechter, M., & Opp, K.-D. (2001). What have we learned about the emergence of social norms? In M. Hechter & K.-D. Opp (Eds.), *Social norms* (pp. 394–415). New York: Russell Sage Foundation.

Henrich, J., McElreath, R., Barr, A., Ensminger, J., Barrett, C., Bolyanatz, A., et al. (2006). Costly punishment across human societies. *Science, 312*, 1767–1770.

Henrich, J., Ensminger, J., McElreath, R., Barr, A., Barrett, C., Bolyanatz, A., et al. (2010). Markets, religion, community size, and the evolution of fairness and punishment. *Science, 327*, 1480–1484. doi:10.1126/science.1182238.

Horne, C. (2001). Sociological perspectives on the emergence of norms. In M. Hechter & K.-D. Opp (Eds.), *Social norms* (pp. 3–34). New York: Russell Sage Foundation.

Huebner, B., Dwyer, S., & Hauser, M. (2009). The role of emotion in moral psychology. *Trends in Cognitive Sciences, 13*, 1–6. doi:10.1016/j.tics.2008.09.006.

Hutcherson, C. A., Bushong, B., & Rangel, A. (2015). A neurocomputational model of altruistic choice and its implications. *Neuron, 87*, 451–462. doi:10.1016/j.neuron.2015.06.031.

Jordan, J., Hoffman, M., Bloom, P., & Rand, D. (2016). Third-party punishment as a costly signal of trustworthiness. *Nature, 530*, 473–476. doi:10.1038/nature16981.

Kahneman, D., Knetsch, J. L., & Thaler, R. H. (1986). Fairness and the assumptions of economics. *Journal of Business, 59*, S285–S300. doi:10.1086/296367.

Keser, C., & van Winden, F. (2000). Conditional cooperation and voluntary contributions to public goods. *The Scandinavian Journal of Economics, 102*, 23–39. doi:10.1111/1467-9442.00182.

Krajbich, I., Bartling, B., Hare, T., & Fehr, E. (2015a). Rethinking fast and slow based on a critique of reaction-time reverse inference. *Nature Communications, 6*, 7455.

Krajbich, I., Hare, T., Bartling, B., Morishima, Y., & Fehr, E. (2015b). A common mechanism underlying food choice and social decisions. *PLoS Computational Biology, 11*, e1004371. doi:10.1371/journal.pcbi.1004371.

Leung, A. K.-Y., & Cohen, D. (2011). Within- and between-culture variation: Individual differences and the cultural logics of honor, face, and dignity cultures. *Journal of Personality and Social Psychology, 100*, 507–526. doi:10.1037/a0022151.

Li, J., & Daw, N. D. (2011). Signals in human striatum are appropriate for policy update rather than value prediction. *The Journal of Neuroscience, 31*, 5504–5511. doi:10.1523/JNEUROSCI.6316-10.2011.

Markus, H. R., & Kitayama, S. (1991). Culture and the self: Implications for cognition, emotion, and motivation. *Psychological Review, 98*(2), 224–253. doi:10.1037/0033-295X.98.2.224.

Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. San Francisco: W.H. Freeman and Company.

Milgram, S. (1963). Behavioral study of obedience. *Journal of Abnormal and Social Psychology, 67*, 371–378. doi:10.1037/h0040525.

Milgram, S. (1965). Some conditions of obedience and disobedience to authority. *Human Relations, 18*, 57–76. doi:10.1177/001872676501800105.

Moll, J., Zahn, R., de Oliveira-Souza, R., Krueger, F., & Grafman, J. (2005). The neural basis of human moral cognition. *Nature Reviews. Neuroscience, 6*, 799–809. doi:10.1038/nrn1768.

Nowak, M. A., & Sigmund, K. (1998). Evolution of indirect reciprocity by image scoring. *Nature, 393*, 573–577. doi:10.1038/31225.

Ohtsuki, H., & Iwasa, Y. (2006). The leading eight: Social norms that can maintain cooperation by indirect reciprocity. *Journal of Theoretical Biology, 239*, 435–444. doi:10.1016/j.jtbi.2005.08.008.

Ostrom, E., Walker, J., & Gardner, R. (1992). Covenants with and without a sword: Self-governance is possible. *The American Political Science Review, 86*, 404–417. doi:10.2307/1964229.

Platt, M., & Plassmann, H. (2014). Multistage valuation signals and common neural currencies. In P. W. Glimcher & E. Fehr (Eds.), *Neuroeconomics: Decision making and the brain* (pp. 237–258). London: Academic Press.

Rand, D. G., & Nowak, M. A. (2013). Human cooperation. *Trends in Cognitive Sciences, 17*, 413–425. doi:10.1016/j.tics.2013.06.003.

Rand, D. G., Greene, J. D., & Nowak, M. A. (2012). Spontaneous giving and calculated greed. *Nature, 489*, 427–430.

Ratcliff, R., & McKoon, G. (2008). The diffusion decision model: Theory and data for two-choice decision tasks. *Neural Computation, 20*, 873–922. doi:10.1162/neco.2008.12-06-420.

Ratcliff, R., & Rouder, J. N. (1998). Modeling response times for two-choice decisions. *Psychological Science, 9*, 347–356. doi:10.1111/1467-9280.00067.

Reicher, S., & Haslam, S. A. (2006). Rethinking the psychology of tyranny: The BBC prison study. *The British Journal of Social Psychology, 45*, 1–40.

Robinson, P. H., Kurzban, R., & Jones, O. D. (2007). The origins of shared intuitions of justice. *Vanderbilt Law Review, 60*, 1633–1688.

Shaver, K. G. (1985). *The attribution of blame: Causality, responsibility, and blameworthiness, Springer series in social psychology*. New York: Springer.

Smith, P. L., & Ratcliff, R. (2004). Psychology and neurobiology of simple decisions. *Trends in Neurosciences, 27*, 161–168. doi:10.1016/j.tins.2004.01.006.

Sober, E., & Wilson, D. S. (1998). *Unto others: The evolution and psychology of unselfish behavior*. Cambridge, MA: Harvard University Press.

Thomson, J. J. (1985). The trolley problem. *The Yale Law Journal, 94*, 1395–1415. doi:10.2307/796133.

Treadway, M. T., Buckholtz, J. W., Martin, J. W., Jan, K., Asplund, C. L., Ginther, M. R., Jones, O. D., & Marois, R. (2014). Corticolimbic gating of emotion-driven punishment. *Nature Neuroscience*. doi:10.1038/nn.3781.

Trivers, R. L. (1971). The evolution of reciprocal altruism. *The Quarterly Review of Biology, 46*, 35–57. doi:10.1086/406755.

United States Sentencing Commission. (2016). *Guidelines manual*.
Voyer, B. G., & Franks, B. (2014). Toward a better understanding of self-construal theory: An agency view of the processes of self-construal. *Review of General Psychology, 18*, 101–114. doi:10.1037/gpr0000003.
Weiner, B. (1995). *Judgments of responsibility: A foundation for a theory of social conduct*. New York: Guilford Press.
Yamagishi, T. (1986). The provision of a sanctioning system as a public good. *Journal of Personality and Social Psychology, 51*, 110–116. doi:10.1037/0022-3514.51.1.110.
Yamagishi, T. (1988). Seriousness of social dilemmas and the provision of a sanctioning system. *Social Psychology Quarterly, 51*, 32–42. doi:10.2307/2786982.
Young, L., & Saxe, R. (2009). Innocent intentions: A correlation between forgiveness for accidental harm and neural activity. *Neuropsychologia, 47*, 2065–2072.
Young, L., Cushman, F., Hauser, M., & Saxe, R. (2007). The neural basis of the interaction between theory of mind and moral judgment. *Proceedings of the National Academy of Sciences, 104*, 8235–8240. doi:10.1073/pnas.0701408104.
Young, L., Camprodon, J. A., Hauser, M., Pascual-Leone, A., & Saxe, R. (2010). Disruption of the right temporoparietal junction with transcranial magnetic stimulation reduces the role of beliefs in moral judgments. *Proceedings of the National Academy of Sciences of the United States of America, 107*, 6753–6758. doi:10.1073/pnas.0914826107.
Zaki, J., Schirmer, J., & Mitchell, J. P. (2011). Social influence modulates the neural computation of value. *Psychological Science, 22*, 894–900.

(Im)Morality in Political Discourse?: The Effects of Moral Psychology in Politics

Nicholas P. Nicoletti and William K. Delehanty

Conceptualizing Moral Psychology in Political Science

In his January 2016 State of the Union Address, President Obama used the phrase "it's the right thing to do" four times. For example, he states, "And I will keep pushing for progress on the work that I believe still needs to be done. Fixing a broken immigration system. Protecting our kids from gun violence. Equal pay for equal work. Paid leave. Raising the minimum wage. All these things still matter to hardworking families. They're still *the right thing to do*. And I won't let up until they get done" (Obama 2016).[1] Later in the speech he remarks, "Providing two years of community college at no cost for every responsible student is one of the best ways to do that [reduce student loan payments], and I'm going to keep fighting to get that started this year. *It's the right thing to do*" (Obama 2016). He goes on, "When we help African countries feed their people and care for the sick—*it's the right thing to do*, and it prevents the next pandemic from reaching our shores" (Obama 2016). By framing his arguments in this way, he is sending a strong signal to the public that these initiatives are informed by moral commitments, and he is attempting to moralize his stance on the issues. Obama's use of moral language to justify policy goals is not specific to him; other political elites do the same. The study of morality in political science and political psychology attempts to explain how morality affects the political process and the behavior of individuals, including the public and political elites.

[1] Italics added for emphasis in all quotations from President Obama.

N.P. Nicoletti (✉) • W.K. Delehanty
Department of Social Sciences, Missouri Southern State University (MSSU),
3950 E. Newman Road, Joplin, MO 64801, USA
e-mail: nicoletti-n@mssu.edu; delehanty-w@mssu.edu

B.G. Voyer, T. Tarantola (eds.), *Moral Psychology*,
DOI 10.1007/978-3-319-61849-4_6

The vast majority of political science literature on morality has focused on morality policy and its uniqueness (see discussion below). However, recent work within the political psychology and political behavior literatures has started to focus on the role of moral psychology in areas such as belief formation, candidate evaluation, and civic engagement. Much of this research is theoretically interesting because scholars argue that models of morality are inherent to individuals and derive from evolutionary traits passed down through generations. In other words, one's moral foundations are a function of genetic traits that affect individual attitudes and behavior. Such traits have been used to explain voter turnout (Fowler and Dawes 2008), political orientations such as party identification (Funk et al. 2013), and political behavior more generally (Alford and Hibbing 2004).[2] The theoretical foundations for the evolutionary basis of moral commitments are discussed in later sections of this chapter and elsewhere in the volume.

The research on moral psychology can be applied by emphasizing how morality is expressed through values. It is a well-known finding in political science that individuals do not have well-constructed attitude constraints in the form of an ideology (Campbell et al. 1960; Converse 1964). A large number of people are unable to define what it means to be liberal or conservative. Recent research has demonstrated that attitude structure along ideological lines may be strengthening. This means that individual political attitudes become more consistent across attitude objects (such as governmental policy and policymakers) with respect to underlying evaluations of those attitude objects. Attitude-object evaluation occurs by reference to an ideological orientation, rooted in values. Values in this context refer to how citizens view their political world. Thus, they provide a cognitive "lens" with which to assess abstract political concepts such as "government" itself or to appraise concrete governmental choices such as public policy. Political ideology creates consistent evaluative standards for attitude objects—to help individuals discern how they think about the political environment. Defined succinctly, a political ideology is a comprehensive political orientation that allows individuals to assess political phenomena by reference to assumptions about the proper role of government in society and the economy.

Lewis-Beck et al. (2008) find tentative evidence that citizens' responses "seem to be connected to more fundamental value orientations. An optimistic interpretation of these findings would be that citizens actually display a capacity for fairly sophisticated political reasoning..." (pp. 233–234). However, they also acknowledge that the evidence is not highly compelling and that, "there seems to be little in the way of broad ideology, of a type that would join together an individual's response to disparate issues" (p. 234). Political scientists use a fair amount of caution when discussing the existence of mass political ideology because an enormous amount of

[2] See Shultziner (2013) for the main points of disagreement with the methodology and the interpretation of the results in various studies that argue for the evolutionary/genetic approach to political behavior—specifically, a criticism of the commonly used twin studies.

evidence exists that a small subsection of national populations—around 21% in the United States and Great Britain and 34% in West Germany—use ideological concepts to evaluate political parties (Dalton 2002; Feldman 2003). However, evidence does suggest that values are important for political attitudes.

The values individuals hold are relatively few in number when compared to individual attitudes. These values can provide the basis for reducing the complexity of political judgments and creating consistency among attitudes, operating in a similar way to political ideology (Feldman 2003). For example, if an individual values the concept of equal opportunity,[3] they are more likely to support government policies designed to make certain choices (such as employment and education) accessible to the public. Valuing equal opportunity helps individuals assess different governmental policies, while remaining more consistent in the support they provide to a range of policies. Political attitudes may be structured by values, and these values exist within greater systems and form the underlying foundation for broader ideologies. Rokeach (1973) defined a value as "an enduring belief that a specific mode of conduct or end-state of existence is personally or socially preferable to an opposite or converse mode of conduct or end-state of existence" (p. 5). A value system is thus an enduring organization of beliefs concerning preferable modes of conduct or end states. In this way, morality and moral conviction may be an outgrowth of individual value systems, with some of them forming along ideological lines. As we move through the various branches of literature, it is important to remember that, while polarization is increasing in the United States and elsewhere, strongly formed ideological belief systems exist for a small group of actively engaged and informed citizens.

In the final section of this chapter, we will return to the idea of the democratic citizen and posit a link between the concepts of moral conviction, elite influence, and political discourse. We contend that elite frames and cues, or the ways in which elites emphasize certain components of events and in so doing provide information to individuals,[4] can mobilize large segments of people by using moral language,

[3] The concept of equality of opportunity is a political idea which is opposed to a strict immobile caste hierarchy but not necessarily social hierarchy in the general sense. The assumption is that there is some social hierarchy based on desirable and undesirable traits that one is born into. However, when a society values equality of opportunity, there exists some competitive process of social mobility, where all members of society are eligible to compete on equal terms. Thus, those that value equality of opportunity will be more likely to support open access to certain factors which increase one's ability to move along the social ladder (e.g., education).

[4] The literature on political frames and cues is extensive. Political frames involve the social construction of some phenomenon, which emphasizes a certain interpretation. For example, the word "welfare" carries a different frame than the phrase "aid to poor children." Framing issues in different ways can help elites and the media elicit specific perspectives from their audience. Political cues from elites provide cognitive heuristics, which simplifies the decision-making process. For example, a citizen might not know much about tax policy, but they know that their political party supports tax cuts. This small piece of information—or cue—helps the citizen to take a stance on taxes without conducting much research. Party identification is one of the most influential and important cues citizens have to help make decisions on candidates when possessing low information.

particularly those who are not well informed or politically active. Because the activation of moral conviction mobilizes individuals via intuitive moral judgments, and such judgments are seen as universal and sui generis for the individuals possessing them, this may have tragic consequences for political discourse in democracies where a presumption exists for rational, reasoned debate focused on negotiation and compromise in dealing with political conflict.

The next section will engage with the morality politics literature, while the following sections will shift from the realm of the polity to the realm of individual moral conviction and its behavioral consequences. The chapter will end by making the argument that, while moral conviction tends to increase political participation, it also increases political polarization and inhibits political discourse, a potentially tragic outcome for modern democratic societies.

Morality Policy and Politics

Scholars studying contemporary public policy have noted the prevalence of what have been termed morality policies. This policy type is typically defined by reference to conflicts over basic moral values or first principles (Mooney and Schuldt 2008; Mooney 1999; Mooney and Lee 1995; Mucciaroni 2011; Haider-Markel and Meier 1996; Haider-Markel 1999; Norrander and Wilcox 1999; Smith 1999). Actors advocate for their respective moral values and seek to have their moral values reflected in public policy (Meier 1999). Scholarship in the study of morality policy tends to focus on policies that engender moral conflict of the type mentioned above. Thus, scholars have focused their analyses on pornography policies (Smith 1999), U.S. state-level abortion policy (Mooney and Lee 1995; Norrander and Wilcox 1999), drug and alcohol policy (Meier 1994), gay and lesbian rights (Haider-Markel and Meier 1996), and the death penalty (Mooney 2000). In these studies, scholars note that morality policy advocacy begins with the use of frames, which employ moral concepts, such as equating certain behaviors or choices with sin (Haider-Markel and Meier 1996; Meier 1999). Given this definition of what constitutes morality policy, it is important to note that public policies are not *intrinsically* moral or nonmoral. Instead, morality policies depend on how they are framed (Schuldt 2008: 200). Put differently, morality policies are not defined by the substantive subject of the policy, but how actors attempt to portray the policy as focused on (im)moral behavior. Thus, defining morality policy this way allows scholars to focus on the *politics* surrounding the creation of morality policy.

Scholarship on morality policy consistently notes the unique political dynamics that define conflicts over the creation of morality policy. Generally, morality politics is defined by a lack of compromise in policy debate (Meier 1999; Mooney and Schuldt 2008); an inability for policy experts to take advantage of their expertise to moderate policy outcomes (Meier 1999); issues that are technically simple, highly salient, and easy (Carmines and Stimson 1980; Mooney and Schuldt 2008; Haider-Markel 1999); greater participation in the policy debate by

citizens and elected officials, and a corresponding decrease in the influence wielded by interest groups (and policy bureaucrats) engaged in the policy process (Haider-Markel 1999, but see Haider-Markel and Meier 1996); and responsiveness of elected officials' policy choices to (perceived) public opinion (Norrander and Wilcox 1999; Mooney and Schuldt 2008). Therefore, the politics generated by morality policy is argued to be qualitatively different from the politics generated by other nonmoral policies. Traditionally, such "nonmoral" policies would include regulations designed to enhance market competition, to maintain industrial safety of workers, and macroeconomic policy. These policies are not technically simple nor highly salient. In addition, widespread public participation in debates about these policies is typically not the norm. Because these policies historically do not involve basic moral value conflicts, their politics is different from policies defined as "moral." It is generally assumed that the unique politics of morality policy comes from the underlying moral value conflicts animating such policies (Mooney and Schuldt 2008; Meier 1999).

An additional interesting political characteristic of morality politics is the degree to which policymakers rarely, if ever, support the behaviors and choices morality policies are designed to address (Meier 1999: 683). Policymakers do not support sin politically for two reasons. First, there is reason to believe that policymakers overestimate the demand for regulating sin or immoral behavior. Since policymakers rely on public (and not private) information regarding individuals' attitudes and acceptance of sinful behavior, they have a tendency to believe that the public supports being "tough on sin" when in fact the public (given its private behavior) is less supportive of such policies (Meier 1999: 683). Second, the political environment and debate surrounding morality policy adoption is one defined by conflicts over basic moral values. In this context, compromise on the part of policymakers regarding sin gives the impression that the moral values at stake in the policy are not *basic* and that there can be trade-offs between, say, the costs of the proposed policy and moral values. Policymakers in this context could be perceived as weak or, by some policy advocates, as morally suspect. The consequences of these dynamics can be public policies that are more punitive and costly, without producing a corresponding decline or reduction in sinful behavior or choices given the costs of the policy. In other words, the political dynamics of morality politics can produce public policies that will be expensive failures (Meier 1999). This is so because instrumental considerations associated with the policy (such as its social costs or ability to address sinful behavior) are discounted given the (assumed) political necessity to address sin. The costs or effectiveness of the policy addressing sin becomes secondary and, in some cases, peripheral, leading to policies that can be very costly but produce consequences that do not correspond to the costs borne by society to achieve them. For example, the so-called War on Drugs in the United States fits this discussion well. Despite significant governmental resources devoted to addressing the sale, production, and consumption of illicit drugs, it is not clear that corresponding declines in these activities/behaviors have resulted or that the private behavior of citizens (such as demand for illicit drugs) has changed significantly (Meier 1999). The same was true for Prohibition and the Eighteenth Amendment. Rather than eliminate the

consumption of alcohol, the policy created a string of bootlegging, trafficking, and organized crime centered on the black market for alcohol and was repealed with the Twenty-First Amendment (Hall 2010). Both the War on Drugs and Prohibition are examples of expensive and costly failures. A 2012 *New York Times* report estimated that the War on Drugs cost the United States between $20 and $25 billion a year over the last decade (Porter 2012). Even with all that money spent, according to the same article referencing the Drug Enforcement Administration, a gram of pure cocaine is 74% cheaper than it was 30 years prior. According to historian Michael Lerner, Prohibition cost the federal government a total of $11 billion in lost tax revenue while costing over $300 million to enforce (Lerner 2008).[5] Even with the high cost and failure of these policies, elites may take on these issues because of the strategic value they have for mobilizing certain voters with high moral conviction over certain issues.

In some studies, scholars make clear that policy *affects* politics, drawing from early work by Lowi (1964, 1972). The classic example is where a policy designed to distribute resources (such as projects for policymaker constituencies) generates support among policymakers since they benefit from the political support that comes from those who receive the resources. Thus, the policy produces predictable political outcomes. In other words, politicians pursue policies that they know will benefit them politically, such as political pork secured for their constituencies. In this case, the policy pursued increased the politician's likelihood of reelection in the next election cycle—otherwise they may not have pursued the same policy pathway. However, in other studies, the policy-politics relationship is less clear. Indeed, one of the areas that scholars in morality policy continue to debate is how to *classify* morality policies (Mooney and Schuldt 2008; Mucciaroni 2011). This ongoing debate by scholars who study morality policy (and its politics) is important because there are some who suggest that morality policies can be fit into existing policy typologies if such policies can be usefully thought of as redistributive (Meier 1999; Haider-Markel 1999). The basic thesis is that morality policies attempt to redistribute values in society, and such policy consequences privilege some values over others (Meier 1999). The redistributive nature of morality policies helps to explain the unique politics of such policies, although the explanation focuses more on the high salience of the policies and a partisan source of policy conflict, which accounts for the higher participation in such policy debates by citizens and elected officials (Meier 1999). In other words, some scholars have suggested that morality policies and their politics are subsumed under already existing policy types.

The focus on the politics of morality policy allows scholars to emphasize the *framing* of morality policy issues as those which involve basic assertions of right and wrong or deontological moral reasoning (for a review, see Mucciaroni 2011; Ryan 2014). There is an assumption made by scholars that if policy issues are *framed as moral*, policy advocates can be successful in removing discussion about the instrumental considerations of a policy (Meier 1999). The removal of these

[5] Accounting for inflation and using 1930s dollars, this amounts to over four billion 2016 dollars.

considerations is relevant and important given the tendency of nonmoral policy debate to focus on the costs and benefits of policy, how effective the policy will be at addressing a social problem, and the trade-offs involved in creating policies, which may benefit some, but not all, members of a society. The result of focusing on instrumental considerations in policy adoption may be to moderate any policy adopted and to promote an incrementalism in policy change over time. If policy advocates frame morality policy in moral terms, it is much more difficult to consider instrumental factors in policy adoption and implementation. Once a policy is moral, traditional cost-benefit analyses of the policy become less relevant, as it is difficult in the policy process to assess the costs or benefits of immoral behavior. This can lead to punitive policies that underestimate the costs to implement them, while overestimating the benefits of them. The problematic consequences of framing policies as moral are also due to increased conflict and interest such policies generate in the public, reducing the ability of policymakers to compromise and evaluate the costs and benefits of policy. For example, the tough on crime frame in the 1980s and 1990s led to harsher prison sentences for drug possession, leading to overcrowded prisons.

The theoretical significance of framing to produce a "morality policy" is that the politics of morality policy is *strategic*, where actors employ frames in their policy advocacy to define the contours of policy to the benefit of their own goals and interests (Mucciaroni 2011). While these goals and interests may be nonmaterial (e.g. "values" or "principles"), scholars of morality policy recognize that strategic framing occurs in the context of moralized political conflict. If strategic framing occurs in policy debates through moral argumentation, scholars must be more diligent in addressing (a) the type of frame used by policy advocates and (b) the strategic goals served by utilizing certain frames (and not others). A theoretical and empirical focus on these components of morality politics enables morality policy scholars to better explain policy outcomes by reference to actors seeking to secure their own goals.

A final area of interest in the study of morality policy concerns what *is* a morality policy versus a nonmorality policy. Some scholars posit that so long as one policy advocate frames their advocacy in moral terms, the resulting policy debate (and policy) is moral. Thus, the absence of moralized frames by policy advocates would distinguish between moral and nonmoral policies (Haider-Markel and Meier 1996). Other scholars focus on the presence of moral value conflict or conflicts over first principles as the defining trait of what counts as a morality policy (Mooney 2001; Mucciaroni 2011). Thus, nonmorality policies are those that do not involve conflicts over moral values or first principles. The necessity to define what *is* and *is not* a morality policy is critical for specifying what morality scholars are attempting to explain. Some scholars have pointed out that the current definition of what *is* a morality policy focuses too much on *personal conduct*, thereby significantly limiting what counts as a morality policy (Mucciaroni 2011).[6] Conceivably, policies that

[6] The typical personal conduct could include use of sexually explicit materials such as pornography, the purchase or use of recreational illicit drugs, the decision to terminate a pregnancy (abortion), and the decision of whether an individual will end their life due to a terminal illness (so-called

focus on behavior and choices that are not personal conduct could be construed as involving conflicts over moral values, such as policy debates regarding the fairness of the U.S. tax code. Morality policy scholars' conceptual definitions of a morality policy tend to overlook how morality vs. nonmorality policy can be distinguished by reference to the use of deontological moral frames (morality policy) vs. instrumentally rational frames (nonmorality policy). It is of course possible that policy advocates could incorporate deontological moral frames and instrumentally rational frames in their advocacy, depending upon what strategic goals they seek to achieve. The explicit use of deontological moral frames defines the boundaries of what is considered moral vs. nonmoral policy. Thus, it is how policy advocates employ these frames that helps to distinguish morality and nonmorality policy, *not* whether policy advocates engage in debate regarding first principles (Mucciaroni 2011).

Scholarship on morality policy does not directly address moral psychology *as such* in attempting to explain the creation and adoption of morality policies. Indeed, most studies of morality policy use policy as the unit of analysis and do not attempt to address individual-level factors that may affect the adoption of morality policies (for an exception, see Haider-Markel 1999). Yet, it is clear that scholars studying morality policy recognize that actors involved in advocating for morality policies do begin such advocacy from a set of moral values or principles, which motivate them to be active in the public policy process. Scholars in morality policy have yet to integrate the insights from other work in political science that starts from a theoretical argument explicating why individuals may think politically in moral terms and how morality structures their opinions and political behavior (Ryan 2014). It may be the case that morality policy scholars, given their unit of analysis, have difficulty taking advantage of the work done on the psychology of morality. Yet, it is striking how the predictions generated by scholarly work on the psychology of morality in political science at the individual-level match (albeit at a different level of analysis) the empirical predictions and results of morality policy scholars. The next section will discuss the interplay between emotions, moral psychology, and individual attitudes, moving the discussion from policy to individual behavior.

Emotions and Morality: Attitudinal and Behavioral Consequences in Politics

Scholarship in psychology and political science has consistently argued, and empirically demonstrated, that emotions can have an effect on individual political attitudes and behavior (Skitka et al. 2005; Skitka and Wisneski 2011; Skitka and Bauman 2008; Peterson 2010; Brader 2006, 2012; Marcus et al. 2000; Marcus et al. 2011; MacKuen et al. 2007; Wisneski and Skitka 2017). Scholars studying the relationship between emotions and political attitudes and behavior traditionally have assumed that individuals' responses to stimuli are affective, and such responses are defined by their *valence*, whether individuals appraise stimuli as positive or negative

"right to die" decisions). This list is representative but certainly not exhaustive.

(Brader 2006). Examples of stimuli used to study affective responses include music (by varying the tone, pitch, or volume) and images of smiling, laughing versus crying, and screaming children (to induce affective responses such as joy versus sadness or fear). The literature connecting emotions and political behavior shows that the use of emotional audiovisual cues in conjunction with valenced scripts produces effects on individual political interest and participation (Brader 2006: 86).

For example, scholars have demonstrated that when individuals are shown campaign advertisements, where the message of the advertisement corresponds to the audiovisual cues of the advertisement, varying the audiovisual cues to induce joy or anger produces predictable political responses. These responses include those experiencing joy to show greater interest in the advertisement and the candidate depicted in the advertisement (Brader 2006). More tellingly, when anger is experienced, those viewing the advertisements report being more likely to vote for the candidate depicted as addressing (not causing) the source of the viewer's anger (Brader 2006). The behavioral and physiological responses of individuals to affective appraisals of the environment defines the degree of arousal induced by the emotional appraisal (Brader 2006). The focus on behavior as indicating degree of arousal provided an early attempt to link affective arousal to changes in individuals' reactions to environmental change. Early scholarship in psychology assumed that emotional responses by individuals to stimuli could be studied by reference to the valence of the response *and* the degree of arousal induced by the response. More recently, scholars in psychology and political science have focused on specific emotions, rather than attempting to classify emotions by reference to their valence and arousal. This scholarship argues that specific emotions (such as anger, anxiety, and enthusiasm) have distinct origins as well as unique political consequences for attitudes and behavior (Brader 2006; Marcus et al. 2000, 2011; MacKuen et al. 2007; Peterson 2010).[7] Thus, scholarship that is more recent tends to focus on how discrete emotions affect individual processing of political information and the consequences of this information processing for political attitudes and behavior. Such a focus is important, for it suggests that the *specific* emotions individuals experience condition how they view political issues, candidates, and their political environments. For example, Wisneski and Skitka (2017) have demonstrated that moral conviction over abortion increased when the emotion of disgust was elicited prior to attitude measurement; however, the emotion of disgust was policy-relevant (abortion-related images) and participants needed to be consciously aware of the emotional cue.

It is important to note that this line of scholarship has shown that individuals can experience emotional reactions to stimuli in their environment without cognitively appraising the stimuli inducing the emotional response, giving rise to a literature in psychology stressing how emotional reactions to stimuli are preconscious, and therefore operate to alter attitudes and behavior without individuals being consciously aware of such processes (for a review of this literature, see Brader 2006).

[7] The specific emotions listed have their origins in how individuals appraise their environment, and a number of scholars address their origins. Thus, anger originates from the affective appraisal that others are thwarting individuals' goals, anxiety originates from uncertainty regarding environmental change, and joy typically arises when individuals can pursue their goals in an environment with markedly less uncertainty or risk.

Individuals do experience emotional reactions, most notably in the form of feelings associated with them. However, experiencing emotions through feelings does not preclude the possibility that emotions trigger physiological responses not "felt" by the individual. Put differently, individuals can feel emotional reactions, but not feeling an emotional reaction to stimuli does not mean such a reaction has not occurred. Moreover, these emotional responses can help to reinforce the attitudes that initially produced the emotional reaction. Scholarship in psychology emphasizes the important connection between cognition and emotion, and there is a lively debate about how (and in what ways) cognition and emotion work together to produce individual attitudes and behavior (Zajonc 1980; Damasio 1994; Lazarus 1981). Thus, while work in political science emphasizes the automaticity associated with emotional reactions to environmental stimuli, this does not preclude the possibility of conscious, cognitive activity operating in conjunction with emotional reactions to stimuli.

This argument assumes a measure of automaticity to the operation of emotions in facilitating information processing by individuals in their environment. The automatic operation of affective appraisal of the environment is understood to be largely functional and rational in that emotions help to reduce the costs of processing information, and they induce a range of decision-making strategies on the part of individuals, given environmental stimuli and how individuals emotionally react to those stimuli (MacKuen et al. 2007). For example, the literature shows that when people experience joy, they are much less likely to reconsider prior political choices, including providing electoral support to candidates associated with the political party they identify with (Brader 2006: 118–126). The practical consequence is that prior political choices are not subject to reevaluation, even if new information in the environment challenges their prior choices.

Some scholarship in political science dealing with emotions and politics borrows theoretically from psychology to argue that emotions allow individuals to respond to information from their immediate environment and to alter behavior in light of that information *as it is appraised* via emotion and cognition (Brader 2006). Emotions work alongside and together with cognition to allow individuals to appraise the consequences of their environments for their own well-being and goals *and* to prepare for changes in behavior based upon changes in their immediate environments. Environmental information is more quickly processed via emotions, but appraisal of that information occurs simultaneously as affective responses are created. Without emotions, individuals would have to appraise all incoming stimuli information cognitively, severely reducing the capacity of individuals to respond quickly to environmental change. In other words, emotions are critical for information processing and appraisal in light of individuals' goals. However, recent research has shown that for certain emotions, such as disgust, changes in reported moral conviction over the issue of abortion required policy-specific cues and conscious awareness (Wisneski and Skitka 2017).

Emotions serve as a *motivational* cue or resource for activity (Skitka et al. 2005; Peterson 2010). Once individuals appraise their environment (either consciously or not), an emotional response helps to prepare the individual for a behavioral change (if necessary). Specific emotions generate different cognitive and/or physiological

states of arousal. For example, the emotion of enthusiasm tends to reduce the motivation to acquire new information and reduce the cognitive effort individuals expend to evaluate information (Brader 2006). Individuals are more confident in their judgments and discount the role new information may play in their current decisions (Brader 2012). Enthusiasm does generally motivate activity (in order to pursue one's goals), and in politics, this can translate into greater political participation via voting and other mechanisms of political activity (Brader 2012). Fear and anxiety have the effect of inducing greater attentiveness to information and can induce the desire to acquire more information, particularly as related to the source of fear or anxiety (Brader 2012). Fear and anxiety reduce risk-taking behavior in individuals experiencing these emotions, with a corresponding tendency to engage in effortful, cognitive processing of information to evaluate alternative courses of action. In politics, this translates into individuals deviating from "habitual" political choices based on predispositions, such as partisanship (MacKuen et al. 2007), and seeking political information, even when the *content* of it may be in conflict with their preexisting political predispositions such as partisanship and ideology. Fear and anxiety do not generally motivate political action per se, but rather induce greater awareness of the political environment, which can change political choices of individuals, as they are more inclined to reevaluate long-standing, habitual choices in politics (MacKuen et al. 2007). Finally, anger tends to reduce effortful, cognitive processing of information, invokes punitive reactions to the perceived source of anger, and tends to dramatically increase political activity, whether in electoral contexts or elsewhere, *relative to* enthusiasm and fear (Brader 2012). Anger operates in a similar way cognitively as enthusiasm in that individuals who are angry tend to rely on habitual reactions to stimuli, such as candidate appeals for voter support, the presentation of political information (candidate or issue focused) in the mass media, and arguments proposed by opponents and proponents of government policies. Politically, this means that anger reduces the tendency for individuals to deviate from long-standing political choices induced by political predispositions such as partisanship and to decrease the acquisition of political information that is in conflict with long-standing political commitments (Brader 2012).

The work done by scholars who analyze the effects of specific, discrete emotions on political attitudes and behavior can be used to understand how moral psychology affects the structure of attitudes and the political consequences of attitudes, which are defined by reference to their being moral. Psychologists have developed a line of research which posits the existence of morally convicted attitudes (Skitka et al. 2005; Skitka and Bauman 2008; Skitka 2010; Skitka and Wisneski 2011). These attitudes are strong in the sense that they are important to those individuals who hold them. In addition, these attitudes are often extreme and held with certainty. They can also play a central role in conditioning how individuals evaluate attitude objects (Skitka et al. 2005). Not all strong attitudes are morally convicted attitudes, but morally convicted attitudes share certain structural characteristics with other strong attitudes. This distinction is critical, for scholars argue that morally convicted attitudes are defined by a set of basic characteristics that differentiate them from other nonmoral but strong attitudes.

First, in some studies, morally convicted attitudes seem to be self-evident to those who hold them, and when pressed by researchers in experimental settings to explain why an attitude object is wrong or bad, many people have a difficult time articulating reasons for their judgments (Haidt 2001).[8] Haidt's model does allow for the possibility that emotions condition the intuitive moral judgments of individuals. Thus, it is possible that the emotional reaction to an attitude object as good or bad conditions how an individual comes to a moral judgment regarding that attitude object. The self-evident nature of moral judgment (or the difficulty individuals have in producing reasons for their judgments) derives from the speed with which emotional evaluation of stimuli occurs. This quick evaluation reduces the possibility that they can recall the basis for their initial moral judgment (Haidt 2001: 819–820). Overall, it may be the case that to explain self-evident moral judgments, focusing on emotional evaluation of attitude objects is needed. Secondly, morally convicted attitudes are seen as universal: the judgments made by individuals regarding what is good or bad are not culturally dependent. Individuals with these attitudes cannot imagine others (outside of their cultural or social contexts) disagreeing with their moral judgments, even if it is known that others *do* disagree. In other words, individuals with these attitudes have a difficult time understanding moral judgments different from their own precisely because they assume their judgments are true. While scholars have noted that moral judgments vary across societies and cultures in reference to *what* is subject to moral praise or blame, they suggest also that the use of moral judgments to evaluate behavior or beliefs of individuals in their societies occurs across human cultures and societies, presumably due to the role moral judgments play in maintaining social order (Skitka and Bauman 2008). Thus, the universal tendency of human societies to create and maintain moral judgments can be linked to the proposition that people experience morally convicted attitudes as universals that cannot be violated or doubted, with the caveat that the *focus* of these attitudes will vary considerably across social and cultural contexts. Finally, morally convicted attitudes generate emotional responses to stimuli that are qualitatively different from the emotional reactions generated by other strong attitudes (Skitka and Bauman 2008; Skitka et al. 2005).

[8] In Haidt's experiments, the purpose was to design a situation that elicits an intuitive response, but they are constructed in such a way where logical or rational arguments against the actions in the vignette are harder to generate. For example, his most cited vignette reads: "Julie and Mark are brother and sister. They are traveling together in France on summer vacation from college. One night they are staying alone in a cabin near the beach. They decide that it would be interesting and fun if they tried making love. At the very least it would be a new experience for each of them. Julie was already taking birth control pills, but Mark uses a condom too, just to be safe. They both enjoy making love, but they decide not to do it again. They keep that night as a special secret, which makes them feel even closer to each other. What do you think about that, was it OK for them to make love?" (Haidt 2001, p. 814). Participants expressed an immediate repugnance to this story, overwhelmingly saying it was morally wrong for brother and sister to make love. However, when pressed to give some reasoning, they struggled because the story makes it clear no harm came of the action. Eventually, participants say something along the lines of, "I don't know, I can't explain it, I just know it's wrong." However, Haidt does not use overtly political or morally charged policies such as abortion or the death penalty. He takes this as evidence that individuals have intuitive reactions and then post hoc rationalize their opposition.

It is important to note that the connection psychologists make between morally convicted attitudes and emotions parallels the relationships political scientists have uncovered linking emotion, political attitudes, and political behavior. It would seem that moral attitudes gain their significance for those who hold them because of the *emotional* response generated by them. This means that morally convicted attitudes are not, on their own, sufficient to induce changes in behavior. Instead, such attitudes need the emotional reaction created by them to sustain changes in individual behavior. Put differently, emotions provide the motivation for individuals to engage in activities that sustain their morally convicted attitudes, particularly in circumstances where those attitudes are questioned or threatened by others in their environment (Skitka et al. 2005). It can be argued that emotions help individuals to defend their moral judgments against threats to them in their environment, further substantiating the claim made by psychologists and political scientists that emotions help individuals to assess environmental information in light of their own goals and well-being. Morally convicted attitudes represent what an individual knows to be good or bad and moral or immoral. Thus, it is reasonable to argue that emotions motivate individuals to defend behavior related to their morally convicted attitudes (or impel corrections to behavior that violates their morally convicted attitudes) given the relationship between these attitudes and more basic conceptions of good and bad that inform them.

It is also the case that this argument helps to explain the often intractable political conflicts surrounding moral issues in politics. To the degree that individuals hold morally convicted attitudes regarding certain attitude objects (issues, candidates, or public policies), individuals are likely to experience significant emotional responses when governments *threaten* the underlying moral judgments of good and bad that inform these attitudes. These emotional responses can motivate political activity oriented toward the removal of the threat, and since emotional responses help individuals to maintain a sense of well-being given environmental change, it is much less likely that compromise and negotiation can mitigate or reduce these conflicts. If individuals feel threatened, compromise and negotiation do not immediately reduce the feeling of threat; they may actually delay resolution of the threat by making individuals think through and consider the source of the threat, how dangerous it is, and a proper response to the threat. A *threat* is a challenge to a person's well-being, and the immediacy of this challenge may thwart attempts to prolong individuals' responses to threats.

However, it is also critical to note that scholarship in political science emphasizes how emotional responses can be created or manipulated via framing to induce behavior (Brader 2006). Thus, while scholarship tends to focus on how such induced emotional responses can benefit some political actors at the expense of others, it is also possible that reframing political issues could be used to mitigate the emotionally laden conflict surrounding moral questions in politics (Haidt 2001). For example, by reframing the use of illicit drugs as one of personal choice and limits on governmental authority, the moral reaction to illicit drug use as an activity of personal conduct becomes a question of how far governments can extend their authority into the personal private choices of individuals (Meier 1999).

Finally, it is worth noting that the existing literature in psychology and political science begins with the assumption that emotional responses to environmental stimuli operate in tandem with cognitive information processing mechanisms to produce attitudes and affect behavior (Brader 2006). This assumption makes it extraordinarily difficult to discern if emotions give rise to attitudes or, rather, if attitudes evoke emotions. This distinction is relevant given work by psychologists who have found that cognitive information processing may occur *after* affective appraisal of the environment (Evans and Frankish 2009; Cushman et al. 2006; Haidt 2001). For example, Cushman et al. (2006) use an experiment to put both the conscious and intuitive reasoning hypotheses to empirical verification. In their study, Cushman et al. (2006: 1083) looked at the morality of three different harm principles:

1. *The action principle*: Harm caused by action is morally worse than equivalent harm caused by omission.
2. *The intention principle*: Harm intended as the means to a goal is morally worse than equivalent harm foreseen as the side effect of a goal.
3. *The contact principle*: Using physical contact to cause harm to a victim is morally worse than causing equivalent harm to a victim without using physical contact.

After reading a series of vignettes, which focused on trade-offs for life and death,[9] participants were asked to rate the protagonist's harmful action on a scale from 1 to 7, where 1 equaled "forbidden," 4 equaled "permissible," and 7 equaled "obligatory." The participants where then asked to justify their patterns of responses, similar to Kohlberg's (1969) studies. The authors hypothesized that intuitive responses would be accompanied by insufficient or failed justifications, whereas conscious responses would be those that articulated the principles used in the judgments—which also aligns with Haidt's (2001) methodology. The results were mixed as the authors found that in the case of the action principle, justifications were sufficient, while the results for the intention principle were much more conducive to the intuitionist model. The results for the contact principle were somewhat intermediary as subjects were typically able to articulate the relevant principle used but were relatively unwilling to endorse it as morally valid. As the authors note, the results demonstrate "that while conscious reasoning was available for a large majority of subjects, others are not available and appear to operate in intuitive processes" (Cushman et al. 2006: 1087). While there is still much debate surrounding the rationalist model of Kohlberg (1969) and the intuitionist model of Haidt (2001), there does seem to be evidence of both processes in the experimental setting.

If emotions give rise to attitudes, then scholarship focusing on intuitionist approaches to moral judgment provides a reasonable explanation for the formation

[9] One example from the study is the vignette for the action principle. The vignette for the action principle read as follows: "Evan" (action, intended harm, no contact)—Is it permissible for Evan to pull a lever that drops a man off a footbridge and in front of a moving boxcar in order to cause the man to fall and be hit by the boxcar, thereby slowing it and saving five people ahead on the tracks? (Cushman et al. 2006, p. 1083).

of attitudes by emphasizing the automatic, nonconscious ways in which environmental stimuli induce emotional responses leading to verbalized moral judgments and attitudes (Haidt 2001). If, however, attitudes evoke emotions, it would seem that rationalist models of moral judgment provide a reasonable account regarding the formation of attitudes by emphasizing how cognition processes information to produce judgments, with emotions motivating individuals to engage in activities designed to sustain and protect their moral judgments.

Theoretical Approaches to Moral Psychology and Political Behavior

There are a growing number of scholars—primarily psychologists and political psychologists—who take a different approach from the morality policy scholarship referenced above when discussing the role of morality in politics. Those studying moral judgment and reasoning use morality as an independent variable that drives decision-making over a variety of political activities and does not necessarily limit the role of morality to a subset of issues that tend to elicit different political reactions from the citizenry. Unlike the morality policy literature, the moral reasoning literature sees morality as a dominant individual trait that can affect such actions as political participation, political opinion on a variety of issues, and even the willingness to accept violent means to achieve a preferred end (Skitka 2010). It is also clear from this literature that moral conviction matters for more than a subset of political issues such as abortion, pornography, and gay marriage; moral reasoning at the individual level has been shown to influence opinion and behavior on a wide range of issues, including those issues traditionally thought of as economic issues, such as taxes and social security.

The Rationalist Model of Morality

There have also been developments in the theoretical framework used to study moral reasoning. Jonathan Haidt (2001) outlines the customary rationalist model of moral reasoning and then argues it should be replaced by a modern social intuitionist approach based on an evolutionary psychology perspective. The rationalist approach stresses a priori reasoning to understand and internalize truths about the world. The process focuses on stages of reasoning and reflection, where the individual weighs the issues before making a decision. Moral rationalism has been studied as a process of reasoning that involves deciding what to do in a certain circumstance.[10] By studying how children reasoned over moral dilemmas,

[10] Kohlberg (1984) used a storytelling technique where he outlined a moral dilemma and then asked participants to make a judgment regarding what the characters should have done. For example, in

Kohlberg (1984) concluded that individuals can go through three stages of moral reasoning: (1) conventional reasoning, where moral standards were shaped by formal rules and consequences; (2) conventional morality, where moral standards were internalized and group norms are primary; and (3) post-conventional morality, where morality evolves into a system of self-chosen principles. Not all individuals pass through all three stages, but those who do are making purposeful adjustments to their moral code. Emotional responses can play a role in these decisions, but they are not the direct cause of the final judgment (Kohlberg 1969; Piaget 1932; Turiel 1983). Also known as the cognitive constructivist model, individuals reason through their moral convictions in stages. There are occasions when one moral value is in conflict with another. Individuals must think through this dilemma to form a coherent, unified response about their own morality. Moreover, individuals continue to reason about their morality as they move through stages of development, with each stage replacing their previous stage's moral reasoning with a more complex system. Higher stages encompass a wide range of moral problems (value conflicts) and generate more solutions to these problems (Emler 2002). Generally, scholars have found that social order reasoning leads an individual to adopt a more conservative moral framework, while principled reasoning lends itself to a more liberal worldview (Emler 2002). In social order reasoning, individuals are concerned with the wider rules of society and focused on preserving and obeying these rules to uphold the law. Principled reasoning involves the development of one's own set of guiding principles, which may or may not fit with the law; for example, the belief in a principle of universal human rights that should be defended even if it goes against the norms and rules of society. These forms of reasoning are derived from Kohlberg's (1984) stages of moral development.[11]

A similar type of behavior is also described by the rational choice literature in economics and political science, where human behavior is described by the maximization of utility, which is defined in terms of individual preferences over potential outcomes. The primary assumption of this perspective is often called *homo economicus* or economic man.[12] Early conceptions of rational man assumed that actors

one story, a husband whose wife is terminally ill decides to steal a potentially lifesaving drug from a local chemist who is charging ten times more than the cost to make the drug. Even so, the husband makes this decision only after he exhausts all options to buy the drug, including borrowing money from family members.

[11] Specifically, social order reasoning is a subset of conventional morality (level 2), while principled reasoning is a subset of post-convention morality (level 3).

[12] It is important to note that rationality is a simplifying assumption used in formal modeling, such as game theory. Mathematically, the assumption simply means that an actor has a set of preferences that do not cycle. People's preferences generate a choice among alternatives such that one alternative is preferred to the others. More formally, when an actor is "rational" in the formal modeling literature, they have complete and transitive (acyclical) preferences. Completeness refers to the idea that, when given a set of preferences, the decision-maker knows the subset of available choices and knows which choices they prefer (or are indifferent between) for a given pairing. Transitivity requires that an actor have a preference order that does not cycle. For example, if $A > B$ and $B > C$, then $A > C$ must also hold. Rationality is essential for modeling purposeful behavior and should not be confused with reasonableness. An "unreasonable" action in one per-

analyzed all possible outcomes and then proceeded to choose the one that maximized their own well-being (utility). This literature has evolved considerably with the differentiation between substantive and bounded rationality. Simon (1978, 1985) argued that actors are limited by the information they have, their cognitive abilities, and the amount of time in which to make a decision (the so-called *cognitive miser* model). This process involves actors constructing a simplified model of the real-world situation, which cuts down on the complexity and simplifies the decision. The actor then behaves rationally with respect to this simplified model. The construction of the simplified model is guided by perceiving, thinking, and learning. In many situations, cognitive heuristics are used to make decisions in complex environments. This is often called *satisficing*, and scholars have identified a number of cognitive heuristics for political decision-making and behavior such as partisanship (Campbell et al. 1960), trusted elite cues (Mondak 1993), and interest groups (Lupia 1994). As noted earlier, elite cues help citizens make decisions when they possess low information. For example, a legislator who is a gun rights advocate may consider the National Rifle Association (NRA) an extremely trusted source. If the NRA has stated that they oppose a bill before the general assembly, this individual—without any additional information—may also oppose the bill because they trust the NRA's stance. In this example, the NRA's opposition to the bill is the relevant cue and operates as a cognitive heuristic (or shortcut) that the individual uses to make a decision without paying the cost of gathering additional information.

Social Context Models

Another theoretical framework focuses on a social constructivist approach where moral reasoning is not an individual process as suggested by Kohlberg (1969, 1984) but rather is socially constructed and communicated. Emler (2002) calls this the sociogenetic model. Choice of political perspective is driven by value preferences and social influences, where distinctive patterns of moral reasoning are associated with comparing, contrasting, and communicating value preferences with others. Social influences help to generate political predispositions and solidify individual moral frameworks over time. Moreover, both liberals and conservatives have access to each other's moral arguments and can accurately predict what the other side will argue in political discourse. Early social influences include parents, and children tend to adopt the political party of their parents later in life (Niemi and Jennings 1991). Other scholars have shown that an individual's social network and the people they talk to about politics are associated with voting behavior and party identification (Huckfeldt and Sprague 1995). Citizens are interdependent and construct their preferences out of a complex interplay among local information environments,

son's mind is not necessarily an "irrational" action because no qualitative restrictions are placed on the decision-maker's preferences; their "rationality" rests on the consistency of their decisions when faced with different choice sets.

institutional and organizational contexts, and individual goals and motivations (Taber 2003). Under this framework, individuals will differ regarding cognitive sophistication, interest in politics, attentiveness, political knowledge, opinion range, and political participation. The social environment (social class, neighborhood, religious affiliation, etc.) and individual characteristics (race, education, age, partisanship, etc.) interact to form an individual's moral framework.

Although this model exists independently from the rationalist and intuitive models, it can coexist with both. The underlying argument is that—rather than being purely rationalist or intuitive—attitudes, values, and moral judgments are further impacted by social networks and the interplay between individuals in the greater social context. As individuals and groups communicate with and interact with others in a social context and exchange communications, their views can change and evolve. Political socialization is a great example of how social communication helps to mold one's beliefs. It is well known in the political science literature that agents of socialization have a strong influence on one's political orientation, with the family being the most important, followed by schools, peer groups, and the media (Sears and Levy 2003). This argument also fits well with the social constructivist model of the policy process, where social constructions within society influence the selection of policy tools and can even help to legitimize policy choices (Schneider and Ingram 1993a, b).

The Intuitionist Model

The social intuitionist model argues that moral intuition comes first and directly affects moral judgment. Any type of reasoning happens post hoc in an attempt to either justify the judgment of the individual or influence the attitudes of others. The research in this vein shows that when an individual is presented with a given situation that arouses morally relevant considerations, it elicits an intuition and a rather immediate judgment. This individual will then begin to post hoc reason as to why they made the judgment they did, and this reasoning is usually an attempt to influence another individual's intuition, judgment, and subsequent reasoning (Haidt 2001). This intuition, in ways similar to satisficing, acts as a cognitive heuristic that simplifies moral reasoning in an individual's mind and allows them to make quick decisions about what is right and wrong, facilitating the pursuit of their political preferences. In other words, intuition allows individuals to make quick judgments about certain policy preferences without paying the costs of gathering large amounts of information, just as an endorsement from a trusted political elite helps individuals make judgments regarding candidates without amassing independent research.[13]

The question then becomes: From where do the intuitions come? Haidt (2012) and others (see Sidanius and Kurzban 2003, for a full treatment) argue that natural selection works both at the individual and the group level. Natural selection favors genes that will help individuals survive, but there is also a between-groups

[13] See Skitka et al. (in press) for a critique of the intuitionist model using both experimental and field study evidence.

multilevel selection process where group-level adaptations assist with group survival. It is a common misconception that the evolutionary approach is an endorsement of "nature" in the nature/nurture debate. Rather, natural selection and the evolutionary approach argue that evolution results in the gradual improvement of the functional fit between an organism and its environment. Every human behavior is both culturally learned and biological in the sense that all behaviors have biological and genetic causes but are expressed in culturally specific ways.

Adaptations that enhance rates of reproduction (e.g., finding food, attracting mates, etc.) will spread. Genes are "selfish" in the sense that they only "care" about the rate at which they replicate relative to other genes (Dawkins 1976). However, there is debate as to what level natural selection operates on. Theories of multilevel selection have recently been used to explain how adaptations within individuals in groups have assisted with between-group conflict and major evolutionary transitions (Wilson et al. 2008). The critique of multilevel selection is that "individuals within a group that carry mutations that cause them to benefit themselves at the expense of the group out-produce more cooperative members, leading ultimately to the replacement of cooperative types with selfish types" (Sidanius and Kurzban 2003: 148). However, there may still be benefits to the group from these mutations. Group members will have a fitness impact on one another. If groups consist of both "altruistic" and "selfish" types, the selfish types will have an advantage, but groups with more altruistic types (acting in ways that benefit the group) will leave more descendants, in the aggregate, than groups with more selfish types. If the benefit from these altruistic types is large enough in terms of reproductive advantage, the frequency of these types can increase from one generation to the next (Sober and Wilson 1998; Sidanius and Kurzban 2003). This means that certain adaptations may be focused on fitness gains through coordinated and cooperative activity.

Morality is thus a series of partially innate intuitions which evolved from adaptations that assisted with group survival. The causal driver behind modern day morality constructs is a hypersensitive agency detection mechanism (e.g., a sensitivity to strange noises in the dark) that helped human beings survive. Hypersensitivity to one's surroundings, and to things like facial detection, had the effect of creating more false positives than false negatives. Early humans, equipped with both the hypersensitive agency detector and the cognitive ability to talk and reason began to share their perceptions and also develop causal logic for why these things happened. The groups that used these stories to develop moral communities were the ones that survived. For example, religion developed as a moral community with binding mechanisms for the group that also elicited commitments and helped overcome collective action problems. Genes played an important role in the multilevel selection process. Different groups will develop different social norms to govern behavior, and these norms are then followed by everyone in the group. Some groups will have norms beneficial to the group as a whole, while others will have detrimental norms. Over time, the beneficial norms tend to spread because of the relative success of the cooperative norms. This cultural selection process is fundamentally driven by individuals who share the same values and norms (Boyd and Richerson 1985; Sidanius and Kurzban 2003). In fact, this theoretical argument may explain in-group vs. out-group conflict based on competition over finite resources.

Haidt (2012) integrates this theoretical framework into the political world by arguing that people do not develop their ideologies randomly but rather are predisposed to a certain belief structure based on their genes. This does not mean they are predestined to choose a particular ideology, only that they are primed by their genes in certain ways. He argues that liberals gain pleasure from novelty, variety, and diversity and are also less sensitive to signs of threat. Those who care about the preservation of tradition, have a high sensitivity to threats, and feel strong loyalty to groups tend to be predisposed toward conservatism.[14] Once an ideological team is chosen, individuals are bound to a belief system that competes against other belief systems and makes political discourse exceedingly difficult. This theoretical framework concludes that individual reactions to political stimuli are automatic and then post hoc rationalized. There are many implications of this model, with the most important being that moral conviction may be an important causal driver of political belief and behavior.

Moral Conviction and Political Behavior

Moral conviction is often couched in the attitude strength literature. Attitude strength is defined as the extent to which an attitude is durable and impactful. Strong attitudes are highly durable to external attack and can be a strong impetus for action (Krosnick and Petty 1995). Durability is broken up into two aspects: persistence and resistance. Persistence refers to the degree to which an attitude remains unchanged over time, also known as stability. Resistance specifies how an attitude holds up to an attack. Strong attitudes tend to be highly stable and resistant. There are also two aspects to impactfulness. The first is the extent to which an attitude is used to form a judgment, and the second is the effect an attitude has on guiding behavior (Krosnick and Petty 1995). Several more aspects of attitude strength have also been identified, such as extremity, importance, certainty, and accessibility (Skitka et al. 2005). Once formed, stronger attitudes tend to be highly enduring relative to weakly held attitudes. However, there is debate as to whether moral conviction is simply another aspect of attitude strength or something different and more powerful.

Skitka et al. (2005) consider the moral mandate hypothesis, which predicts that attitudes associated with moral conviction are either different from other kinds of attitudes or have a more powerful relationship to behavior than other nonmoral attitudes with similar strength. The moral mandate hypothesis has important implications for political behavior. For example, moral conviction may be a predictor of political activism, campaign donations, voter turnout, the choice to campaign for specific candidates, and even the ability (or inability) to have meaningful political discussions and compromise. Moral convictions are taken as experiences of fact

[14] The argument of Haidt (2012) is substantively related to work in political science on the relationship between personality traits or "types" and public opinion. For a review, see Mondak and Hibbing (2012).

rather than as matters of preference or taste that can easily differ among groups of people. This often leads to the expression of these attitudes in terms of moral certainty in the sense that something is just fundamentally wrong, such as cannibalism or abortion. These certainties lead to the justification for action. Possibly, then, the most important aspect of moral conviction is that these attitudes are highly affective and associated with emotion. This may be what makes these attitudes highly impactful and action oriented. Due to these differences, Skitka et al. (2005) argue that moral mandates cannot be reduced to structural features of attitudes. While it is true that strongly held attitudes will share structural similarities with moral convictions, these moral stances will be more extreme, certain, important, and central. Moral convictions will also be idiosyncratic, as individuals will have moral convictions over a wide variety of different issues based on their point of view. If moral convictions are different from other strong nonmoral attitudes, there should be empirical evidence demonstrating that moral convictions produce differences in political behavior. First, we will consider whether conservatives or liberals have an advantage in the realm of morality and analyze theories of conservative versus liberal ideological structures.

This point specifically applies to the mobilization of voters who are morally committed over a variety of issues. If conservatives are morally committed to a wider range of issues than liberals are, it is easier for elites and other activists to mobilize this set of voters by activating their moral frames. The argument is based upon a framework that identifies liberals (and many libertarians) as having a higher level of moral relativism, which makes them more tolerant of a variety of viewpoints based on the belief that ethical principles are not universal but social rather constructions. Under this framework, conservatives have an unconditional moral view and are less tolerant of viewpoints opposed to their own; because morality can be a strong driver of political attitudes and behavior, this means that conservatives can be more easily mobilized by moral arguments, possibly giving them an advantage in elections where moral issues are highlighted. For example, evangelical Christians may have helped Bush to win in 2000 and 2004 because they were more concerned about value issues at the polls. However, the conservative advantage hypothesis is still heavily debated, and research tends to suggest that liberals and conservatives are driven by moral commitments over a wide range of issues and policies (Ryan 2014).

A Conservative Advantage?

In the United States, the Republican Party and those espousing the conservative ideology have branded themselves as the party of moral values. This is partially a result of Ronald Regan's campaign, which mobilized religious voters by taking socially conservative stances on abortion, prayer in public schools, and family values. Social conservatives have been an impactful and highly mobilized group of voters. This group helped George W. Bush win election in 2000 and again in 2004. In the 2004

presidential election exit polls, 22% of respondents believed "moral values" were the most important issue, and 80% of these voters cast ballots for Bush (Hillygus and Shields 2005). Moreover, Republicans were more likely than Democrats to report that moral values were what mattered most in their candidate choice (Skitka et al. 2005).[15] Research also suggests that liberals are prone to endorse moral relativism, and conservatives have a tendency to be moral absolutists with the idea that "right" and "wrong" are invariable regardless of culture (Skitka et al. 2005; Hunter 1991; Layman 2001).

Haidt (2003) argues that conservative elites tap into a wide range of moral foundations that appeal to voters with strong moral convictions. He outlines six psychological systems within a broad theory of moral foundation. The six systems are care/harm, liberty/oppression, fairness/cheating, loyalty/betrayal, sanctity/degradation, and authority/subversion. Liberals tend to speak to the care/harm and liberty/oppression systems, which emphasize ideas of social justice, compassion for the poor, and a struggle for political equality. Liberals also emphasize the fairness/cheating foundation. Conservatives, however, tend to emphasize all six foundations but interpret several of them differently. For example, in terms of the liberty/oppression system, liberals focus more on the rights of vulnerable groups, while conservatives focus on traditional ideas of liberty and the right to be free of government intrusion. Conservatives care about the care/harm foundations, but liberals care more. Both ideological groups care about fairness/cheating, but conservatives care more about meritocracy and getting rewards equivalent to the work put in. In addition to the care/harm, liberty/oppression, and fairness/cheating foundations, conservatives also appeal to the other three. Haidt (2003) argues that liberals are ambivalent to these moral foundations, while conservatives embrace them. Under this moral foundations theory, conservatives have the advantage because they actively appeal to all moral systems rather than only a few, thus mobilizing a set of voters with a wide range of moral conviction.

Similar arguments have been used to explain why people of lower-socioeconomic status vote against their economic self-interest (Frank 2004). Using vignettes regarding sexual acts, Haidt and Hersh (2001) found that liberals had a narrower moral domain, while conservatives' domain was broad and multifaceted. In this study, liberals focused on the ethics of autonomy, where actions that did not harm other people were not generally found to be morally wrong. On the other hand, conservatives were more concerned with the ethics of the community, divinity, and group norms, which made them more resistant to unusual sexual acts that they felt were against their moral code. This study corroborates the idea that conservatives have a larger moral system that elites can use to mobilize large numbers of voters and thus possibly have an advantage.

[15] It is important to note that Hillygus and Shields (2005) find that the effect on the electoral choice of respondents' attitudes toward moral issues of abortion and gay marriage was inconsistent and much smaller than voters' evaluations of the Iraq War, terrorism, and the national economy. Other research by Holsti (2008) has shown similar findings regarding the Iraq War. Holsti (2008) argues that the Iraq War produced an additional rally 'round the flag effect, which significantly helped Bush win reelection.

However, there is an alternative argument known as the equal opportunity motivator hypothesis (Skitka and Bauman 2008). Under this hypothesis, both sides are similarly motivated by moral considerations but in different ways. Lakoff (2002) argues that conservative and liberal attitudes are rooted in a moral system based on the conception of the family. These different ideas of the family ultimately lead to differences in moral values. The conservative worldview is based on the strict father model, which is focused on the traditional nuclear family. Under this model, the father has the primary responsibility of supporting and protecting the family, setting the rules, and enforcing the rules. The mother's role is to take care of the children and the house and to uphold the father's authority. Children must respect and obey authority and obey their parents in order to build character, self-discipline, and self-reliance. Love and nurturance are important but never outweigh parental authority. Once children mature, they are on their own and must use their acquired self-discipline to survive and make their own destiny.

The liberal worldview is called the nurturant parent model and emphasizes love, nurturance, and empathy, which assist children in becoming responsible, self-disciplined, and self-reliant. Unlike the strict father model, nurturance, caring, and respect by the parents are what lead to well-raised children. Support and protection are a large part of nurturance on the part of the parents. Obedience comes out of love and respect, not fear of punishment. If parental authority is to be legitimate, there must be good communication about the rationale for rules. Questioning by children is seen as positive, but in the end parents are still responsible for making good decisions. The principal goal is for children to live fulfilled and happy lives and for them to have empathy for others so that they can make necessary social ties. Parents help their children develop their potential for achievement and enjoyment.

Both of these moral frameworks for the family are built from the same elements but in different orders with radically different processes. These worldviews are linked to politics, as people see the government as a parent and a nation as a family. Conservatives believe that the function of government requires citizens to be self-reliant, self-disciplined, and thus able to help themselves. Liberals believe that the government should help people in need and thus support social programs. Conservatives stress political actions such as protecting the nation from external threats and upholding the moral order. They also embrace self-reliance and minimal government intrusion into the pursuit of self-interest. Liberal political actions promote empathy, helping those who cannot help themselves, and promoting fulfillment in life. Lakoff's (2002) theoretical framework suggests that both conservatives and liberals are compelled by moral convictions but in different ways. Barker and Tinnick (2006) test Lakoff's hypotheses and find that people envision proper power relations between citizens and the government based on their understanding of proper power relations between children and parents. Those with the strongest feelings regarding proper child rearing—either nurturant or discipline oriented—were more consistently liberal or conservative in their political leanings. Research has also indicated that, across a wide variety of issue domains, liberals and conservatives were equally likely to have strong moral convictions (Skitka et al. 2005; Ryan 2014). Research has also found that, as political partisanship increases in strength, moral conviction over a variety of issues is more prevalent on both the left and the right (Ryan 2014).

Given the state of the current research, there is some indication that conservatives tend to overtly moralize certain issues, and there is some connection with religiosity (Ryan 2014), but both sides of the aisle show high levels of moral conviction when the variable is measured using questions that explicitly ask the respondent to self-report whether or not they hold a certain attitude as a part of their moral framework. This suggests that there is not a conservative advantage in the realm of moral reasoning and politics. The more interesting question is how moral conviction affects people in terms of political opinion and behavior. We suggest that moral conviction is a double-edged blade; it has the desirable tendency to increase political action but also to limit the ability of opposing sides to deliberate, compromise, and build social capital in a democratic system.

The Consequences of Moral Conviction

There is a growing literature using moral conviction as an independent variable to explain a host of political behaviors. To measure morally grounded attitudes, scholars have generally relied on some form of self-report question to gauge whether an attitude held is a moral conviction for the holder. The question takes the general form, "How much are your feelings about _____ connected to your core moral beliefs or convictions?" (Skitka et al. 2005; Skitka 2010; Ryan 2014).[16] The answer to this question is given on a five-point Likert-type scale with labels such as "not at all," "slightly," "moderately," "much," and "very much." In order to isolate moral conviction relative to other aspects of attitude strength, researchers must also use a series of self-report questions to measure attitude extremity, importance, and centrality (personal importance). Questions such as these are designed to tap into the visceral recognition that an attitude over some political question or policy is based on a moral belief. As noted above, scholars have argued that moral convictions are based on intuition, and thus asking open-ended questions about why respondents feel the way they do will most likely result in post hoc justification rather than an acknowledgment of intuitive moral conviction (Haidt 2012; Ryan 2014).

The common approach to these studies is to present the participant with policy options or to ask the participant to tell the researcher what they think is the most important issue facing the nation today. A wide variety of issues has been looked at using similar approaches. Researchers then measure a series of attitude strength dimensions such as extremity, certainty, importance, and personal relevance (Krosnick and Petty 1995) to control for attitude strength. The moral conviction question is used as the primary explanatory variable to differentiate between attitude

[16] Ryan (2014, p. 384) actually uses three different questions to measure the overall construct of moral conviction. He asks to what extent the respondent's opinion is "a reflection of your core moral beliefs and convictions," "deeply connected to your fundamental beliefs about right and wrong," and "based on a moral principle." He notes that these questions were highly related, with Cronbach's alphas ranging between 0.90 and 0.93.

strength over nonmoral issues and issues for which participants have high moral conviction. Respondents may exhibit high attitude strength and high moral conviction over the same issue, but there is significant variation. Moral conviction is then used to explain a series of political behaviors and opinions, controlling for the other elements of attitude strength. The empirical findings generated from these studies are compelling and in many cases disturbing for normative theories of democracy.

Using the approach where the researcher allowed participants to choose the most important problem facing the nation today, Skitka et al. (2005) found that people tended to spontaneously think of issues that they said connected to their moral conviction. The issues mentioned were also correlated with issues that scored high on the other attitude strength variables. The dependent variable in this study was social distance or the degree to which respondents answered, "I would be happy to have someone who did not share my views on (their defined most important issue)… as 'President of the United States,' 'as Governor of my state,' 'as a neighbor,' 'to come to work at the same place as I do,' 'as a roommate,' 'to marry into my family,' 'as someone I would personally date,' 'as my personal physician,' 'as a close personal friend,' 'as the teacher of my children,' and 'as my spiritual adviser'" (Skitka et al. 2005). The scores of these items were averaged to create a global index of social distance, with higher values reflecting greater social distance.[17] The results demonstrated that, even after controlling for a host of other variables including attitude strength, moral conviction explained unique variance in social distance. Those respondents who felt that an issue was connected to their moral conviction preferred more social distance from someone with a dissimilar attitude.[18] Results in subsequent studies in the same article found the same was true when respondents were given a set of issues (capital punishment, legalization of marijuana, and nuclear power), even when controlling for political orientation, measured as liberal and conservative ideology. In an additional study, participants were told they would be having a discussion with another person. They were asked to pull up a chair from a row of chairs against a wall while the experimenter went to find the other participant. In the room was a gender-neutral book bag with a "Pro-Child" or "Pro-Choice" pin attached to it. The distance between the participant's and the discussion person's chair was measured. The variance in the distance between the chairs could be explained by the participant's view on abortion and whether they held a moral conviction over the abortion issue. In the final study within Skitka et al. (2005), the researchers looked at group formation and discussion. Respondents were asked to have a group discussion over a set of controversial issues. They had to discuss whether abortion should remain legal, if the death penalty should be continued, and

[17] The authors also looked at how different social distance relationships varied in intimacy based on an additional student survey. Based on the results, they broke these relationships down into those more intimate or close to the respondent and those more distant. Unsurprisingly, relationships such as prospective relationships of marriage were more intimate than electing public officials.

[18] The researchers controlled for gender, age, and features of attitude strength such as extremity, importance, and certainty. Moreover, participants rejected those who did not share their beliefs, irrespective of whether the relationship was intimate or distant.

if standardized tests should be a graduation requirement. The task was for the group to come to some agreement on who should make the decision and the procedure for how the decision should be made. It was stressed that they were not supposed to actually make a decision. They would then vote by secret ballot to determine if they had come to a consensus or if they had reached deadlock—these were two of the ways the group could end their discussion. The discussion would also end if the time ran out. The members of the group were given no information about the attitudes of the other group members. The groups were observed, and measures of good will, cooperativeness, and group outcomes were recorded. The results demonstrated that groups "that discussed procedures to resolve conflicts about moral mandates were less likely to agree to a procedure than were groups that discussed procedures to resolve conflict about a non-moral mandate or a strong attitude…" (Skitka et al. 2005: 912). Interpersonal connections were more strained with heterogeneous groups than homogeneous groups; respondents reported feeling less positive, and third-party observers could detect group-level tension in heterogeneous groups.

This study is particularly important for how moral conviction relates to democratic cooperation, compromise, and community. Scholars such as Putnam (2000) have commented on how social trust and social capital are declining in the United States. An important part of building trust is the ability to interact with those who are dissimilar from oneself to build bridging social capital. Moral conviction impedes this process as people reject those with dissimilar attitudes, preferring to form relationships with those who have similar attitudes. Moreover, collaboration, decision-making, and compromise are strained over issues that present different moral stances. The moral mandate hypothesis has serious consequences for decision-making in democracy. The results indicate that moral conviction may produce what Robert Bellah et al. (1985) called social enclaves, where like-minded people form isolated political communities.

Using a similar study to the one described above, Ryan (2014) finds that moral conviction can present itself over a wider range of issues such as collective bargaining, social security, stem cell research, gay marriage, and the war in Afghanistan. While some issues—such as gay marriage—clearly elicit increased moral conviction, it was also true that among some respondents, moral conviction was exhibited over a variety of issues. Moreover, using a principal component analysis, he found that moral conviction is not just another element of attitude strength. While dimensions of attitude strength loaded together, moral conviction questions loaded heavily on the same factor. Morally convicted attitudes tended to be more important, relevant, and extreme. It captured a distinct aspect of an attitude. More importantly, perceiving an issue as morally important predicted specific feelings toward issue opponents. Moral conviction generated a general negative affect, anger, and disgust. Morality was also more likely to predict these negative feelings than the elements of attitude strength. In an additional study in the same article, Ryan (2014) uses the American National Election Study (ANES) Evaluations of Government and Society (EGSS) questions to understand the relationship between moral conviction and political behavior. He finds that moral conviction is associated with one-sided political assessments and increased political participation. Moreover, respondents identified a

moral conviction over a wide range of issues including traditionally economic issues, such as unemployment, healthcare, immigration, the budget, and the environment. In fact, healthcare was perceived as more moral than abortion. Both left-leaning and right-leaning respondents had high levels of moral conviction over the issues studied. Not only did moral conviction predict a more one-sided view of politics, but this was true even when controlling for religiosity. Interestingly, moral conviction was correlated with high levels of partisanship, which can explain why strong partisans have trouble compromising in today's political system.

Ryan's (2014) study provides more evidence that when moral conviction makes its way into political discourse, democracy may be threatened. Moral conviction leads to negative affect toward opposing viewpoints and can materialize over a wide range of issues. Moreover, moral conviction is an action-oriented dimension of attitude. Those most likely to participate in the political system may also be the most likely to collaborate in homogeneous groups, reject opposing viewpoints, and create social distance between themselves and political opponents. More importantly, this is true not just for traditionally moral issues but for a wide variety of issues, including economic issues. Findings like these have led scholars to decree that there is a dark side to moral conviction (Skitka and Wisneski 2011; Skitka and Mullen 2002a). For example, research has demonstrated that individuals become unconcerned with how moral mandates are achieved, as long as they are achieved. Commitments to procedural safeguards that protect democracy and civil society (such as securing the free speech rights of those who speak about controversial topics, the associational rights of those who form groups to defend socially deviant practices, and prohibitions against governmental searches of personal property or belongings with the absence of legal justification) can erode when people are pursuing a moral end (Skitka and Mullen 2002b). These scholars worry that extreme acts such as terrorism and the acceptance of weakening civil liberties may be a side effect of those with strong moral convictions (Skitka 2010).

Studies have shown that a strong moral conviction over an issue or set of issues inspires action. Skitka and Bauman (2008) find that moral conviction motivated voter turnout in the 2004 presidential election, controlling for a host of other variables such as attitude strength and partisanship, and that the effect was strong for people on both the left and the right. Research by Waldron et al. (1988) demonstrated that a powerful motivating factor, among several others, in the opposition to nuclear weapons during the Cold War was a moral obligation to act to try to prevent nuclear war.[19] The authors suggest that moral responsibility is a strong factor in a group's ability to overcome the classic collective action dilemma put forth by Olson (1965).

Motivation for action is an important necessity in any democracy. Normative democratic theorists often lament that political participation has sharply dropped in the United States since the late 1800s. Moral conviction seems to solve part of this problem with its action-oriented influence. Research has shown that a moral conviction

[19] Other motivating factors consisted of nervousness regarding the nuclear arms race, emotional reactions to the nuclear arms race (specifically anger), tendency toward political activism and campaigning, and social approval of activism by reference groups.

can induce action on a wide range of political behavior including campaigning, voting, and group activism. However, it also impedes compromise and exacerbates polarization over a host of issues. Political elites also realize that framing an issue as moral can mobilize voters. A major implication of moral foundations theory is that elite rhetoric framed in moral terms can mobilize voters, moralize a wide range of issues, and further polarize public discourse. Moralizing an issue by elite actors will cue those with moral beliefs to be more one-sided with political opinions and thus prevent moderate, compromising discourse in the American public.

Immorality in Political Discourse: Elite Rhetoric, Moral Conviction, and Political Discourse

Thus far, the theoretical and empirical literatures in moral psychology suggest that the presence of morality in politics has significant consequences for political attitudes and behavior. The literature indicates that those individuals who possess morally convicted attitudes are more likely to be politically active *and* tend to view compromise or negotiation in regard to the objects of their morally convicted attitudes as inadequate responses to those who do not share their moral judgments. Indeed, the literature suggests that the presence of morally convicted attitudes induces a lack of compromise and a motivational drive to reduce threats to moral judgments. Put simply, the literature paints a problematic portrait of citizens who possess morally convicted attitudes.

The behavior of the mass citizenry in republican democracy has been a widely researched topic in political philosophy and science. Madison (1788/2003) warned of majority faction, and the framers designed the Constitution to protect against the "tyranny of the majority." De Tocqueville (1835/2000) wrote in *Democracy in America* about the tendency for the American public to make decisions based on whims and passions rather than reason. In his classic book, *Public Opinion*, Lippmann (1927) argues that people form beliefs about the world based on simplified "stereotypes" that do not represent the "world outside." The complexity and obscurity of politics, along with time constraint, lack of interest, and lack of information, limit the public's ability to come to sound conclusions. Complicating things further are political elites and the symbols they use to unify public opinion. In *The Phantom Public*, Lippmann (1927) concludes that the American people are relatively uninformed, uninterested, and usually haphazard in their views. He argues that opinions tend to emerge during crisis and fade away shortly after. Dewey (1927) was a little more optimistic, arguing that the public can be consequential and informed when they are threatened by a negative externality brought on by legislation; however, even Dewy argues that modern society is filled with distractions that divert the attention of the public away from politics. Decades later, Downs (1957) would formalize the idea of rational ignorance, where individuals rationally forego the costs of gathering information because their vote has little impact on elections.

For modern political science, the literature conflicts with traditional normative models of a democratic citizen who is considered a political equal with respect to fellow citizens (Dahl 2000; Gilens 2012). Citizens in a democracy ought to be given the capacity to engage in self-government with others. The literature focused on moral psychology suggests that equal respect for all citizens does not materialize, particularly if citizens possess differing moral judgments. Moreover, the literature undermines the argument proposed by some democratic political theorists that democratic citizens engage in thoughtful, reasoned debate regarding questions facing all members of the political community (Held 2006). If decisions in a democracy represent the collective authority of the public, debate must include those subject to the decisions, along with a consideration of their interests and goals. However, the moral psychology literature shows that collective decisions (particularly those dealing with political issues perceived as "moral") will not likely be debated in a reasonable and inclusive way. The tendency would be for such issue debates to focus on the basic moral judgments of those involved, producing polarization, conflict, and a breakdown of deliberative norms such as mutual respect and reciprocity between citizens. Finally, literature suggests that modern democratic citizens lack basic information about political events and processes (Delli Carpini and Keeter 1996). If citizens lack basic information about political events and processes, they are less likely to know how their interests are affected by governmental policy and to hold government accountable when its actions violate such interests (Gilens 2012). Moreover, in a democracy, the absence of a well-informed public means governmental authority can be used in ways that undermine the capacity of citizens to discern the consequences of policymaking for broader, public interests vs. narrower, private ones.[20]

Political scientists have long known that modern democratic citizens lack information about basic political processes, the decisions of elected officials, and their broader political environment. Yet, this so-called democratic dilemma is partially resolved by showing how cues and cognitive heuristics can induce behavior that is rational given low information (Gilens 2012; Lupia 1994; Popkin 1991; Lupia and McCubbins 1998). Essentially, cues and cognitive heuristics help to reduce the information costs associated with forming and maintaining opinions on matters of policy as well as electoral preferences. The problem is that cues and cognitive heuristics are not always neutral in the information they provide to citizens. Indeed, the provision of elite cues has the potential to alter how the public views policy and political issues, while cognitive heuristics can distort how information is processed by citizens, particularly given political predispositions such as partisanship (Graber and Dunaway 2015). In conjunction with the political science literature on cues and heuristics, media scholars have shown that citizens' opinions regarding political objects can be affected by the presentation of information through the processes of priming and framing (Gilens 2012; Iyengar and Kinder 1987; Chong and Druckman 2007;

[20] See Gilens (2012, pp. 70–71). Gilens argues that, at least in American democracy, the government tends to be overly responsive to the interests of the affluent, particularly when this group's interests diverge from those of the less well-off.

Slothuus 2008). Priming can be defined as information sources focusing public attention on certain problems, by reference to information provided about that problem and not others (Graber and Dunaway 2015). Put differently, priming is the process whereby information sources indicate to citizens what is *important*. Scholars argue that the process of priming can help to set the public agenda, at least to the degree that citizens *and* government focus on certain problems and not others (Graber and Dunaway 2015; Iyengar and Kinder 1987). In contrast, issue framing can be defined as "the process by which a communication source... defines and constructs a political issue or public controversy" (Nelson et al. 1997: 567). Because most political issues are complex, frames can be employed to alter how citizens think about issues, by activating "considerations" in the minds of those who receive the relevant frame (Zaller 1992; Slothuus 2008). The literature on "framing effects" further indicates that frames are not simply neutral in the provision of information for citizens. Instead, frames usually contain "evaluative content" that seeks to direct the receivers of the frame to a particular interpretation of issues and/or political events. Thus, frames serve as sources of information, *and* they enable "sensemaking" on the part of citizens who are thinking about a particular issue or event (Slothuus 2008).

Given the elite model of public opinion, which explains citizen attitudes by linking elite opinion to voter preferences, moral framing may have a significant impact. As noted above, attention to and knowledge about politics tend to be quite low on average for the American public. Zaller (1992) defined political awareness as the extent to which an individual pays attention and understands the information he or she has encountered. We expect that politically aware individuals will react differently to political messages than the unaware, who may not react at all. But there is another piece of the puzzle: political predispositions, which include political values. Zaller defined political predispositions as stable individual-level traits that regulate the acceptance or rejection of political communications. Those with higher levels of political attention and knowledge will likely have stable predispositions. Zaller (1992) and Zaller and Feldman (1992) present what they call the Receive-Accept-Sample (RAS) model of public opinion, which is based on four axioms. The Receive Axiom states that individuals who are more politically aware will be more likely to receive a given political message. The Resistance Axiom states that people will tend to resist political arguments in conflict with their political predispositions and possess the contextual information necessary to perceive a relationship between the message and the predisposition, which also requires political attention. The Accessibility Axiom states that the more recent considerations "on the top of the head" will be the ones recalled from memory when survey questions are asked. Finally, the Response Axiom states that individuals answer survey questions by averaging across considerations immediately salient or accessible to them.

When the information environment contains a single message, it will be those with moderate levels of political awareness that will experience attitude change; those with low attentiveness are unlikely to receive the message, and those with high attentiveness are likely to have the necessary information to resist messages inconsistent with their predispositions. When there are two messages, due to discourse between political elites, mass opinion will polarize. For example, early in the

Vietnam War, there was a single pro-war message, and thus support for the war fits the pattern of moderate awareness favoring support for the war most intensely, while highly aware doves were able to resist the dominant pro-war message. However, once political discourse began, hawks and doves who were moderately to highly aware began to polarize along ideological lines. Public attitudes toward major issues are a response to the relative intensity of competing political communications. Zaller (1992) sums it up perfectly:

> "When elites unite on a mainstream issue, the public's response is relatively non-ideological, with the most aware members of the public reflecting elite consensus most strongly. When elites come to disagree along partisan or ideological lines, the public's response will become ideological as well, with the most politically aware members of the public responding most ideologically." (Zaller 1992: 210)

Such processes best implicate the strategic use of morality in politics by political elites, particularly if a democratic public lacks the informational resources to cognitively evaluate how morality can prime them to view certain public issues as important and frame citizen interpretation of issues in moral terms. To the degree that democratic publics lack information and rely on cues (as well as cognitive heuristics) in the formation of their political interests and choices, political elites can use cues and heuristics embedded within issue frames to moralize policy debate and the issue positions of citizens.[21] And, to the degree that moralized policy debate activates intuitive moral judgments on the part of citizens, they are more likely to be politically active due to the motivational benefits associated with emotional responses to moral conflict. The problem, of course, is that the increased political activism of those mobilized on the basis of moralized elite frames produces a polarized and conflict-ridden political environment, while potentially serving the narrow political goals and interests of elites.

Taken together, the intuitive moral conviction model and citizens' use of elite cues to form opinions may be dangerous for republican democracy. Given the participatory qualities of moral conviction, elites' framing of issues as moral can mobilize a significant portion of the citizenry. Moreover, because moral conviction tends to invoke one-sided and uncompromising political attitudes, political discourse over these issues becomes more difficult. The moral framing of single issues—such as abortion or immigration—can generate responses from the citizenry that limit compromise. Once elected, officials who moralize these issues are then bound by a powerful electoral connection, whereby elected officials must act in accordance with the interests of those who elected them (Mayhew 1974), which can prevent compromise within legislatures, leading to further polarization in Congress and among the public. Recent research has clearly demonstrated that American political polarization has increased significantly and that this has led to increased political participation in terms of voter turnout and campaign activity (Abramowitz 2010).[22]

[21] Scholars have noted that the *effects* of issue frames may vary across individuals as a function of the political information they possess and their level of interest in, and awareness of, politics. See Slothuus (2008, p. 10).

[22] Political polarization has been a function of the disintegration of what was known as the New

However, while political participation may be spurred by moral conviction, it also reduces the ability of individuals to have democratic discourse. In this light, morally convicted attitudes seem to be rather immoral with respect to the normative model of republican democracy.

Getting to Political Compromise

While the literature seems to suggest that morally charged debates allow for very little compromise, as each side seems to demonize the other, there is evidence that it is possible to break through the moral obstacle to achieve fruitful political discourse. Tetlock et al. (2000) and Tetlock (2003), in the context of moral trade-offs, have shown that political elites can avoid negative affects when suggesting policy stances that go against individuals' moral predispositions. Solid arguments that involve a reframing of moral trade-offs in terms of tragic outcomes can be quite effective. For example, they found that, although most people rejected the idea of buying and selling organs for medical transplant, 40% qualified their opinion if these transactions were the only way to save lives or if steps were taken to assist the poor in purchasing them and preventing them from selling organs when desperate for money. Compromise can be achieved by going through the right rhetorical motions and using the power of reframing. The problem is that it is often in elites' strategic benefit to moralize an issue rather than attempt to reframe issues in terms that the morally convicted will accept. This is because moralizing an issue generates political support with few immediate costs to the elite. In contrast, reframing issues is costlier, effortful, and does not have the immediate political benefits that moralizing has. Furthermore, reframing issues can create conflict with others in society who already have framed an issue in light of their interests. Finally, it may be the case that the new frame is not accepted as an alternative way in which to view an issue, particularly by those who are inclined to think of the issue in moral terms.

Johnathan Haidt (2012) has also given several recommendations that can get morally convicted groups to begin to understand each other and possibly compromise. First, interpersonal connections are a powerful tool. The idea is to forge relationships with dissimilar people by emphasizing what they have in common first

Deal coalition. Under this coalition, the Democratic Party was made up of white Southerners, northern white and ethnic voters, white working class voters, and eventually African American voters during the 1960s' Civil Rights Movement. This helped the Democratic Party dominate for many years. However, as the Civil Rights Movement became an important issue for the Democratic Party (particularly with the Johnson Administration and the Civil Rights Acts of 1964, the Voting Rights Act of 1965, and the Civil Rights Act of 1968), white conservative voters began to shift their party allegiances toward the Republican Party (Carmines and Stimpson 1989). It took several decades for this shift to be completed. This change at the elite level of the parties began to affect the loyalties of voters. By the 1990s, the political parties were much more ideologically homogeneous, which led to less compromise and more polarization in both Congress and the electorate. See Abramowitz (2010) for a full discussion.

and moving toward discussions of issues where they are opposed. This is very similar to Putnam's (1993, 2000) idea of social trust and social capital. Forging relationships between dissimilar people through group interaction can lead to less demonization and more trust and compromise. The other recommendation is to focus on cooperative rather than conflictual goals. Rather than focusing on goals that morally opposed groups cannot agree upon, these groups need to focus on goals they can both get behind and then proceed to treat discourse over outcomes as positive sum, rather than zero sum competition. Competition breeds animosity—especially when it is over moral values—while cooperation allows for much more compromise. By getting morally opposed groups to focus on goals they agree upon, and getting a discourse to evolve in a positive direction, it may be possible to build social trust that will lead to future compromise on morally driven and divisive issues. This is an important argument because research tends to show that the more like-minded groups isolate themselves from others, and the more civic culture declines, the harder it becomes to build the social capital necessary to make democracy work in the long term (Putnam 1993, 2000).

To conclude, we argue that compromise is possible under certain conditions, such as those mentioned above. However, the primary obstacle to overcome is the perverse incentive that elites have to frame issues in order to mobilize maximum support. Moral framing is a very effective way to gain support and mobilize voters. Moreover, while elites moralize issues to get elected, they also have policy preferences in line with their own moral convictions. This makes political discourse exceedingly more difficult, while simultaneously increasing political participation. The juxtaposition of destructive democratic discourse with an active, politically engaged public underscores the promises and pitfalls associated with moral conviction in democratic politics.

References

Abramowitz, A. I. (2010). *The disappearing center: Engaged citizens, polarization, and American democracy*. New Haven, CT: Yale University Press.

Alford, J. & M. Hibbing. (2004). The Origins of Politics: An Evolutionary Theory of Political Behavior. *Perspectives on Politics. 2*(4): 707–723.

Barker, D. C., & Tinnick, J. D. (2006). Competing visions of parental roles and ideological constraint. *American Political Science Review, 100*(02), 249–263.

Bellah, R. N., Madsen, R., Sullivan, W. M., Swidler, A., & Tipton, S. M. (1985). *Habits of the heart: Individualism and commitment in American life*. Berkeley, CA: University of California Press.

Boyd, R., & Richerson, P. J. (1985). *Culture and the evolutionary process*. Chicago: University of Chicago Press.

Brader, T. (2006). *Campaigning for hearts and minds: How emotional appeals in political ads work*. Chicago: University of Chicago Press.

Brader, T. (2012). The emotional foundations of democratic citizenship. In A. Berinsky (Ed.), *New directions in public opinion* (pp. 193–216). New York, NY: Routledge.

Campbell, A., Converse, P. E., Miller, W. E., & Stokes, D. E. (1960). *The American voter*. New York, NY: Wiley.

Carmines, E. G., & Stimson, J. A. (1980). Two faces of issue voting. *American Political Science Review, 74*, 78–91.
Carmines, E. G., & Stimson, J. A. (1989). *Issue evolution: Race and the transformation of American politics*. Princeton, NJ: Princeton University Press.
Chong, D., & Druckman, J. N. (2007). Framing theory. *Annual Review of Political Science, 10*, 103–126.
Converse, P. E. (1964). The nature of belief systems in mass publics. In D. E. Apter (Ed.), *Ideology and its discontent* (pp. 206–261). New York: Free Press.
Cushman, F., Young, L., & Hauser, M. (2006). The role of conscious reasoning and intuition in moral judgment: Testing three principles of harm. *Psychological Science, 17*(12), 1082–1089.
Dahl, R. A. (2000). *On democracy*. New Haven, CT: Yale University Press.
Dalton, R. J. (2002). *Citizen politics: Public opinion and political parties in advanced industrial democracies*. Washington, DC: CQ Press.
Damasio, A. R. (1994). *Descartes' error: Emotion, reason, and the human brain*. New York: G.P. Plenum.
Dawkins, R. (1976). *The selfish gene*. Oxford, UK: Oxford University Press.
Delli Carpini, M., & Keeter, S. (1996). *What Americans know about politics and why it matters*. New Haven, CT: Yale University Press.
De Tocqueville, A. (1835/2000). *Democracy in America*. Indianapolis, IN: Hackett Publishing.
Dewey, J. (1927). *The public and its problems*. Athens: Ohio University Press.
Downs, A. (1957). *An economic theory of democracy*. New York, NY: Addison Wesley.
Emler, N. (2002). Morality and political orientations: An analysis of their relationship. *European Review of Social Psychology, 13*(1), 259–291.
Evans, J., & Frankish, K. (Eds.). (2009). *In two minds: Dual processes and beyond*. New York, NY: Oxford University Press.
Feldman, S. (2003). Values, ideology, and the structure of political attitudes. In D. O. Sears, L. Huddy, & R. Jervis (Eds.), *Oxford handbook of political psychology* (pp. 477–508). New York, NY: Oxford University Press.
Fowler, J. H., & Dawes, C. T. (2008). Two genes predict voter turnout. *Journal of Politics, 70*(3), 579–594.
Frank, T. (2004). *What's the matter with kansas?: How conservatives won the heart of America*. New York, NY: Holt Paperbacks.
Funk, C. L., Smith, K. B., Alford, J. R., Hibbing, M. V., Eaton, N. R., Krueger, R. F., Eaves, L. J., & Hibbing, J. R. (2013). Genetic and environmental transmission of political orientations. *Political Psychology, 34*(6), 805–819.
Gilens, M. (2012). Two-thirds full? Citizen competence and democratic governance. In A. J. Berinsky (Ed.), *New directions in public opinion* (pp. 52–76). New York, NY: Routledge.
Graber, D. A., & Dunaway, J. (2015). *Mass media and american politics* (9th ed.). Thousand Oaks, CA: CQ Press.
Haider-Markel, D. P. (1999). Morality policy and individual-level political behavior: The case of legislative voting on lesbian and gay issues. *Policy Studies Journal, 27*(4), 735–749.
Haider-Markel, D. P., & Meier, K. J. (1996). The politics of gay and lesbian rights: Expanding the scope of the conflict. *The Journal of Politics, 58*(02), 332–349.
Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review, 108*(4), 814–834.
Haidt, J. (2003). The moral emotions. In R. J. Davidson, K. R. Scherer, & H. H. Goldsmith (Eds.), *Handbook of affective sciences* (pp. 852–870). Oxford: Oxford University Press.
Haidt, J. (2012). *The righteous mind: Why good people are divided by politics and religion*. New York, NY: Random House.
Haidt, J., & Hersh, M. (2001). Sexual morality: The cultures and emotions of conservatives and liberals. *Journal of Applied Social Psychology, 31*(1), 191–221.
Held, D. (2006). *Models of democracy*. Stanford, CA: Stanford University Press.
Hall, W. (2010). What are the policy lessons of National Alcohol Prohibition in the United States, 1920–1933? *Addiction, 105*(7), 1164–1173.

Hillygus, D. S., & Shields, T. G. (2005). Moral issues and voter decision making in the 2004 presidential election. *PS: Political Science and Politics, 38*(2), 201–209.
Holsti, O. R. (2008). *American Public Opinion on the Iraq War*. Ann Arbor, MI: University of Michigan Press.
Huckfeldt, J., & Sprague, J. (1995). *Citizens, politics, and social communication: Information and influence in an Election Campaign*. New York, NY: Cambridge University Press.
Hunter, J. D. (1991). *Culture wars: The struggle to control the family, art, education, law, and politics in America*. New York, NY: Basic Books.
Iyengar, S., & Kinder, D. (1987). *News that matters*. Chicago: University of Chicago Press.
Kohlberg, L. (1969). Stage and sequence: The cognitive-developmental approach to socialisation. In D. A. Goslin (Ed.), *Handbook of socialisation theory and research* (pp. 347–480). Chicago: Rand McNally.
Kohlberg, L. (1984). *The psychology of moral development volume two*. New York, NY: Harper and Row.
Krosnick, J. A., & Petty, R. E. (1995). Attitude strength: An overview. In R. E. Petty & J. A. Krosnick (Eds.), *Attitude strength: Antecedents and consequences* (pp. 1–24). Mahwah, NJ: Lawrence Erlbaum.
Lakoff, G. (2002). *Moral politics: How liberals and conservatives think*. Chicago: University of Chicago Press.
Layman, G. (2001). *The great divide: Religions and cultural conflict in American party politics*. New York, NY: Columbia University Press.
Lazarus, R. S. (1981). A cognitivist's reply to Zajonc on emotion and cognition. *American Psychologist, 36*, 222–223.
Lerner, M. A. (2008). *Dry Manhattan: Prohibition in New York City*. Cambridge, MA: Harvard University Press.
Lewis-Beck, M., Jacoby, W., Norpoth, H., & Weisberg, H. (2008). *The American voter revisited*. Ann Arbor, MI: University of Michigan Press.
Lippmann, W. (1927). *The phantom public*. New Brunswick, NJ: Transaction Publishers.
Lowi, T. J. (1964). American business, public policy, case studies, and political theory. *World Politics, 16*, 677–715.
Lowi, T. J. (1972). Four systems of policy, politics and choice. *Public Administration Review, 32*(4), 298–310.
Lupia, A. (1994). Shortcuts versus encyclopedias: Information and voting behavior in California insurance reform elections. *American Political Science Review, 88*(1), 63–76.
Lupia, A., & McCubbins, M. (1998). *The democratic dilemma: Can citizens learn what they need to know?* Cambridge: Cambridge University Press.
MacKuen, M., Marcus, G. E., Neuman, W. R., & Keele, L. (2007). The third way: The theory of affective intelligence and American democracy. In W. R. Neuman, G. E. Marcus, A. N. Crigler, & M. MacKuen (Eds.), *The affect effect* (pp. 124–151). Chicago: University of Chicago Press.
Madison, J. (2003). No. 10 the same subject continued. In C. Kesler & C. Rossiter (Eds.), *The federalist papers* (pp. 71–79). New York, NY: Singnet.
Marcus, G. E., Neuman, W. R., & MacKuen, M. B. (2000). *Affective intelligence and political judgment*. Chicago: University of Chicago Press.
Marcus, G. E., MacKuen, M., & Neuman, W. R. (2011). Parismony and complexity: Developing and testing theories of affective intelligence. *Political Psychology, 32*(2), 323–336.
Mayhew, D.R. (1974). Congress: The Electoral Connection. New Haven, CT: Yale University Press.
Meier, K. J. (1994). *The politics of sin*. Armonk, NY: M. E. Sharpe.
Meier, K. J. (1999). Drugs, sex, rock, and roll: A theory of morality politics. *Policy Studies Journal, 27*(4), 681–695.
Mondak, J. (1993). Source cues and policy approval: The cognitive dynamics of public support for the Reagan Agenda. *American Journal of Political Science, 37*, 186–212.
Mondak, J. J., & Hibbing, M. V. (2012). Personality and public opinion. In A. Berinsky (Ed.), *New directions in public opinion* (pp. 217–238). New York, NY: Routledge.

Mooney, C. Z. (1999). The politics of morality policy: Symposium editor's introduction. *Policy Studies Journal, 27*(4), 675–680.
Mooney, C. Z. (2000). The influence of values on consensus and contentious morality policy: U.S. Death Penalty Reform, 1956–1982. *The Journal of Politics, 61*, 223–239.
Mooney, C. Z. (2001). The public crash of private values. In C. Z. Mooney (Ed.), *The public crash of private values* (pp. 3–20). Chatham, NJ: Chatham House.
Mooney, C. Z., & Lee, M. H. (1995). Legislating morality in the American states: The case of pre-Roe abortion regulation reform. *American Journal of Political Science, 39*, 599–627.
Mooney, C. Z., & Schuldt, R. G. (2008). Does morality policy exist? Testing a basic assumption. *Policy Studies Journal, 36*(2), 199–216.
Mucciaroni, G. (2011). Are debates about "morality policy" really about morality? Framing opposition to gay and lesbian rights. *Policy Studies Journal, 39*(2), 187–216.
Nelson, T. E., Clawson, R. A., & Oxley, Z. M. (1997). Media framing of a civil liberties conflict and its effect on tolerance. *American Political Science Review, 91*, 567–583.
Niemi, R. G., & Jennings, M. K. (1991). Issues and inheritance in the formation of party identification. *American Journal of Political Science, 35*, 970–988.
Norrander, B., & Wilcox, C. (1999). Public opinion and policymaking in the states: The case of post-Roe abortion policy. *Policy Studies Journal, 27*(4), 707–722.
Obama, B.H. (2016). *Inaugural address of President Barak Obama*. Retrieved January 12, 2016, from https://www.whitehouse.gov/sotu.
Olson, M. C. (1965). *The logic of collective action*. Cambridge, MA: Harvard University Press.
Peterson, M. B. (2010). Distinct emotions, distinct domains: Anger, anxiety and perceptions of intentionality. *The Journal of Politics, 72*(2), 357–365.
Piaget, J. (1932). *The moral judgment of the child* (M. Gabain, Trans.). New York, NY: Free Press.
Popkin, S. L. (1991). *The reasoning voter: Communication and persuasion in presidential campaigns*. Chicago: University of Chicago Press.
Porter, E. (2012, July 3). *Numbers tell the failure in the drug war*. P. B1.
Putnam, R. D. (1993). *Making democracy work: Civic traditions in modern Italy*. Princeton, NJ: Princeton University Press.
Putnam, R. D. (2000). *Bowling alone: Collapse and revival of American community*. New York, NY: Simon and Schuster.
Rokeach, M. (1973). *The nature of human values*. New York, NY: Free Press.
Ryan, T. J. (2014). Reconsidering moral issues in politics. *The Journal of Politics, 76*(2), 380–397.
Schneider, A., & Ingram, H. (1993a). Social construction of target populations. In P. A. Sabatier & C. M. Weible (Eds.), *Theories of the policy process* (3rd ed., pp. 105–151). Boulder, CO: Westview Press.
Schneider, A., & Ingram, H. (1993b). Social construction of target populations: Implications for politics and policy. *The American Political Science Review, 87*(2), 334–347.
Sears, D. O., & Levy, S. (2003). Childhood and adult political development. In D. O. Sears, L. Huddy, & R. Jervis (Eds.), *Oxford handbook of political psychology* (pp. 61–109). New York, NY: Oxford University Press.
Shultziner, D. (2013). Genes and politics: A new explanation and evaluation of twin studies results and association studies in political science. *Political Analysis, 21*(3), 350–367.
Sidanius, J., & Kurzban, R. (2003). Evolutionary approaches in political psychology. In D. O. Sears, L. Huddy, & R. Jervis (Eds.), *Oxford handbook of political psychology* (pp. 146–181). Oxford: Oxford University Press.
Simon, H. A. (1978). Rationality as process and as product of thought. *American Economic Review, 68*, 1–16.
Simon, H. A. (1985). Human nature in politics: The dialogue of psychology and political science. *American Political Science Review, 79*, 293–304.
Skitka, L. J. (2010). The psychology of moral conviction. *Social and Personality Psychology Compass, 4*(4), 267–281.
Skitka, L. J., & Wisneski, D. C. (2011). Moral conviction and emotion. *Emotion Review, 3*(3), 328–330.

Skitka, L. J., Wisneski, D. C., & Brandt, M. J. (in press). Attitude moralization: Probably not intuitive or rooted in perceptions of harm. *ECurrent Directions in Psychological Science.*

Skitka, L. J., & Bauman, C. W. (2008). Moral conviction and political engagement. *Political Psychology, 29*(1), 29–54.

Skitka, L. J., & Mullen, E. (2002a). The dark side of moral conviction. *Analyses of Social Issues and Public Policy, 2*(1), 35–41.

Skitka, L. J., & Mullen, E. (2002b). Understanding judgments of fairness in a real-world political context: A test of the value protection model of justice reasoning. *Personality and Social Psychology Bulletin, 28*(10), 1419–1429.

Skitka, L. J., Bauman, C. W., & Sargis, E. G. (2005). Moral conviction: Another contributor to attitude strength or something more? *Journal of Personality and Social Psychology, 88*(6), 895–917.

Slothuus, R. (2008). More than weighting cognitive importance: A dual-process model of issue framing effects. *Political Psychology, 29*(1), 1–28.

Smith, K. B. (1999). Clean thoughts and dirty minds: The politics of porn. *Policy Studies Journal, 27*(4), 723–734.

Sober, E., & Wilson, D. S. (1998). *Unto others: The evolution and psychology of unselfish behavior*. Cambridge: Harvard University Press.

Taber, C. S. (2003). Information processing and public opinion. In D. O. Sears, L. Huddy, & R. Jervis (Eds.), *Oxford handbook of political psychology* (pp. 433–476). Oxford: Oxford University Press.

Tetlock, P. E., Kristel, O. V., Elson, S. B., Green, M. C., & Lerner, J. S. (2000). The psychology of the unthinkable: Taboo trade-offs, forbidden base rates, and heretical counterfactuals. *Journal of Personality and Social Psychology, 78*(5), 853–870.

Tetlock, P. E. (2003). Thinking the unthinkable: Sacred values and taboo cognitions. *Trends in Cognitive Sciences, 7*(7), 320–324.

Turiel, E. (1983). *The development of social knowledge*. New York, NY: Cambridge University Press.

Waldron, I., Baron, J., Frese, M., & Sabini, J. (1988). Activism against nuclear weapons build-up? Student participation in the 1984 primary campaigns. *Journal of Applied Social Psychology, 18*(10), 826–836.

Wisneski, D. C., and L. J. Skitka. (2017). Moralizing Through Moral Shock: Exploring Emotional Antecedents to Moral Conviction.*Personality and Social Psychology Bulletin, 43*(2), 139–150

Wilson, D. S., Van Vugt, M., & O'Gorman, R. (2008). Multilevel selection theory and major evolutionary transitions. *Psychological Science, 17*(1), 6–9.

Zaller, J. (1992). *The nature and origins of mass opinion*. Cambridge: Cambridge University Press.

Zaller, J., & Feldman, S. (1992). A simple theory of the survey response: Answering questions versus revealing preferences. *American Journal of Political Science, 36*(3), 579–616.

Zajonc, R. B. (1980). Feeling and thinking: Preferences need no inferences. *American Psychologist, 35*, 151–175.

An Open Letter to Our Students: Doing Interdisciplinary Moral Psychology

Edouard Machery and John M. Doris

For decades, psychologists and other scientists have been producing fascinating research on morality. More recently, increasing numbers of scholars have been exploiting this science in humanistic study of morality, while scientists have increasingly attended to theoretical resources from the humanities. As with any project worth doing, interdisciplinary moral psychology can be done badly or well; there's been much important work, but there have also been unfortunate errors, some of which we ourselves have committed. As teachers, we want to avoid transmitting our mistakes; we hope those coming after us can do better than we have.

With this hope as our muse, we'll here offer a rudimentary guidebook for moral psychologists attempting to inform their theorizing by reference to the social, behavioral, and neurological sciences. In the first instance, we write to students of the humanities—especially students in our home discipline of philosophy—seeking to develop empirically credible theories. But we also have in mind students of science doing literature surveys for review pieces or introductions to experimental work (exercises of scandalously underappreciated difficulty and importance) that also require savvy science writing. Finally, we expect our remarks will be useful to editors and referees assessing the quality of work for publication.

We'll identify important errors to avoid and say something about how to avoid them. As is usual when professors are professing, much of what we say will seem quite obvious to our audience. Faced with this obviousness, some readers will doubt that the mistakes we identify have actually been made. Our experience, dating to the

E. Machery (✉)
Center for Philosophy of Science, University of Pittsburgh,
817 CL, Pittsburgh, PA 15260, USA
e-mail: machery@pitt.edu

J.M. Doris
Philosophy-Neuroscience-Psychology Program & Philosophy Department, Washington University in St. Louis, St. Louis, MO, USA
e-mail: jdoris@artsci.wustl.edu

B.G. Voyer, T. Tarantola (eds.), *Moral Psychology*,
DOI 10.1007/978-3-319-61849-4_7

inception of interdisciplinary moral psychology as a more or less robust academic discipline, leaves us less sanguine; like politicians say, mistakes have been made, and many of them seem, in retrospect, painfully evident. Here, as is also common when professors profess, we're going to ask that you take our word for it. While naming names would have the considerable expository advantage of concrete examples, it would also be too reminiscent of the accusatory, *gotcha!*, rhetorical climate afflicting the analytic philosophy of our graduate years, which we join a growing number of our colleagues in endeavoring to avoid. (Perhaps that's our first bit of advice: *be polite!—at least until provoked*.)

In the martial arts, it's said that the beginning techniques of one discipline are the advanced techniques of another; those exploring foreign disciplines are bound to make the occasional slip as they reach beyond the theories and methods of their training. These mistakes, and our advice about avoiding them, aren't all equally obvious: some aspects of the endeavor, like reading the science with care, are both mandatory and relatively straightforward, while others, like developing sophistication in statistical technique, are difficult and, for many of us, will remain aspirational.

But don't let the perfect be the enemy of the competent. If you're like us, you'll make mistakes, but that's no reason not to make your best effort. Often enough, best efforts are good enough; most always, they're better than half-assed efforts. By identifying some likely missteps, we hope to save you some of them; if some of our advice is obvious, and some of it excessively demanding, perhaps some of it is also good advice. Different readers, we expect, will be at different points in the creative process, from beginners to seasoned veterans, and will therefore find our advice differentially accessible and differentially useful. Hopefully, you'll find enough useful to justify your read.

Before beginning, we offer apologies for another weakness of professing professors, the appearance of telling people what to do. We realize that nobody likes to be told what to do; while the constraints of grammar entail that advice is often formed in the imperative, we mean our advice as advice, not orders. It's a free country, and you're free—unless you're writing your dissertation with us!—to take or leave any or all of the advice we offer here. Still, we're optimistic enough to think that some of our advice is worth taking and that some of you will be able to do your jobs a bit better if you do.[1]

Advice (1): *Know your questions.* Your first question, manifestly, concerns what it is you want to know. Close on its heels comes the question of how you can come to know it: in particular for the present context, whether your question is amenable to scientific inquiry, or is *scientifically tractable*. It's tempting to begin by suggesting that scientific questions are empirical questions, but then you'd be left with the unfortunate business of demarcating the empirical, and doing so with theoretical perspicuity makes for Aegean labor, even if we (mostly) know the empirical when we see it.

[1] We dedicate this paragraph to our good friend Shaun Nichols.

Instead, you might start flat-footedly, with what scientists actually study; better yet, start by getting clear on the things scientists *don't* study. We may be unable to make the observations needed to address some empirical questions, such as some questions about evolutionary or cosmological history. More prosaically, logistical difficulties make some quite recognizably empirical questions resistant to scientific inquiry; for example, the dynamic manifestations of personality over the life course are observable, but the long-term longitudinal studies needed to assess them are extraordinarily difficult to execute (but see the 75-year-long Grant study reported in Vaillant 2012). Then there are questions that do not look empirical at all: how people attribute desert to one another may be an empirical question, but questions of who deserves what and when are matters of normative theory, where empirical investigation has little direct import.

There, we've said it: *normative*. As the name "moral psychology" suggests, our enterprise mingles empirical and normative elements or, as we generally prefer to say, descriptive and prescriptive elements. A good many philosophers in analytic philosophy, perhaps the majority, have viewed this intermingling with deep suspicion; at least, the issue continues to draw controversy (Bedke 2012; Copp 2007; Darwall et al. 1992; Doris, Machery and Stich 2017; Fitzpatrick 2014; Nagel 1980; Railton 2003). What we're going to treat as an established methodology has been the subject of heated debate; writing in the pages of a major newspaper, one philosopher accused some of our fellow moral psychologists of "hating philosophy." We're trying to teach peace here, so we won't insist you "know thy enemy." But you'd better take care to know your audience, and also those who have reasons for refusing to sit in your audience.

Historically, the most prominent philosophical reasons for this reluctance are dicta like the "fact/value distinction" and the "is/ought gap," which maintain, more or less, that descriptive statements do not entail prescriptive statements. However the world may be, nothing directly follows about how to feel, and what to do, about it: the fact that smoking causes illness is a reason to quit smoking only if one happens to value continued good health over the sublime pleasures of tobacco, and that's not an outlook it is mandatory to have (perhaps you're here for a good time, not for a long time).

Fair enough. But while there probably exists some sort of inferential barrier between the *is* and the *ought* (Russell 2010), there equally appears a close relation between them. If it's commonly acknowledged that *is* does not imply *ought*, it's also commonly acknowledged that "ought implies can": the fact that you haven't the ability to bring your deceased loved ones back to life seems inescapably relevant to the question of whether you are appropriately censured for failing to do so.

Doubtless, both commonplaces bear more complication than we've conveyed; properly characterizing the relation between the descriptive and the prescriptive remains a central reoccupation of ethical theory (for the record, our approach follows Railton's 2003 methodological naturalism). Yet strangely, the practice of philosophical ethics often proceeds in innocence of what is arguably its central theoretical conundrum: when not explicitly agonizing about normativity, philosophers have happily, and often quite unselfconsciously, adverted to descriptive considerations in their argumentation (Doris and Stich 2005: 113–15).

In the first instance, our diagnosis of this curious circumstance is cheerfully question begging. Philosophers proceed as though descriptive considerations are relevant

to ethical theorizing because descriptive questions are relevant to ethical theorizing: how people reason is relevant to accounts of deliberation, under what conditions human organisms thrive is relevant to theories of well-being, and the organization of human personality is relevant to understandings of moral character.

In the second instance, we suspect that this circumstance obtains—despite the widely recognized existence of an inferential barrier—because ethical theorizing, and ethical reflection more generally, is not much concerned with strict logical entailments; as we might put it, ethical thought is not frequently "deductively tight." Questions about the most reasonable thing to do, or the most defensible way to live, are seldom settled by straightforward inference and associated questions of validity, as seems to us evident in Rawls' extraordinarily influential method for ethics, "reflective equilibrium" (Rawls 1971; Daniels 2013).

In short, the best theories in ethics and moral psychology are likely to contain various admixtures of fact and value. Yet the pervasive cross-pollination of the descriptive and prescriptive does not absolve you of responsibility for being maximally clear about what sorts of claims you're making, and when. However uncertain and fluid the descriptive/prescriptive distinction, defending descriptive claims like "the behavior of participants in experiments using the ultimatum game is best explained by reference to their intuitions about fairness" and prescriptive claims like "theories of justice should be sensitive to considerations of fairness" are different enterprises, requiring different techniques. At the same time, interesting theoretical claims in and around moral psychology, such as "people typically evince a robust ethical commitment to fairness, and the best theory of justice will be sensitive to such robust ethical commitments," will often be hybrids (cf. Robinson and Darley 1995). Of course, we don't think there's anything wrong with doing this; that's what we do. But in so doing, be clear on what structures in your theoretical edifice require which kind of support. (And be very clear on the connotations of the terms you use, especially across disciplines; psychologists, for example, are given to annoying philosophers by using "normative" in a nonnormative sense, to mean something like "typical.")

In pursuing an interdisciplinary moral psychology, we join something approaching a broad consensus—with the admonition that care be taken in figuring out what is actually going on in the disciplines you visit. The consensus, no doubt, papers over all manner of theoretical difficulty, but in this, we suspect, it is no different than any other intellectual consensus.

How exactly can you join this consensus and do some interdisciplinary work? Start by finding a theoretical question that interests you. Then, if you have ambitions toward getting paid for pursuing your interests, make sure it's also of interest to others. While we've expressed distaste for the pugilistic ambiance that characterized the philosophy of our professionally formative years, it remains the case that much academic research, most especially in philosophy, is agonistic, and asking about winners and losers is a very useful heuristic: if your hunches are right, what theorists should be comforted and which concerned? (As one of us likes to ask his students, a little hyperbolically, *whose ox gets gored?*)

With the dialectical space identified, one can then ask if any bones of contention are scientifically tractable. The best evidence of this is an *established* scientific literature addressing the questions in question. If there is no such literature, you might contemplate beginning one, as an experimental philosopher. We urge you to consider this prospect with extreme trepidation; there might be good reasons why there's not already a literature. (In any case, see our remarks on collaboration below.)

Supposing there are scientific materials ready at hand, you can get underway by drawing connections between the scientifically tractable questions in your rhetorical space and actual scientific attempts to tract them. Be very aware that these connections cannot be simply assumed; they must be *argued* for. And arguing for them is not trivial. Drawing these connections often involves extrapolating from empirical results: drawing conclusions that the original scientific research did not specifically test. There is nothing wrong with extrapolating, but it is inherently risky, must be done carefully, and must be acknowledged for what it is. For instance, one of us (Machery 2012a, in press) has expressed doubts about the universality of the distinction between moral and nonmoral norms. In support of this doubt, he notices that crosslinguistic research by the natural semantic metalanguage linguists shows that "moral" is not a linguistic universal: it is not found in every language (e.g., in Bengali). The conclusion that the distinction between moral and nonmoral norms is culturally specific does not logically follow from the linguistic evidence (people could draw the distinction even if they do not lexicalize it), but Machery brought together other pieces of evidence to argue that *together* they support the cultural specificity claim.

Remember, your one-discipline colleagues are likely to be dubious about your appeal to empirical literatures. They can read as well as you can; if they were convinced the science mattered to what they're doing, they already would be writing about it. Difficulty is more acute with the sorts of normative questions that often engage moral psychologists and ethicists; normative relevance is contestable (Doris 2002: 113), and you must be prepared to contest it. Indeed, establishing relevance is a big part of your job; it won't be enough to vindicate your theoretical hunches, but it will be enough to earn you a voice in the discussion that can't responsibly be ignored. Then, the fun begins.

Advice (2): *read the science*. Once again, that advice is obvious does not mean it is bad. Nor is obvious advice always followed. For example, one approach used by humanists discussing science is to borrow it from other humanists, taking their science on testimony. Sometimes, this is a suspicion that can't be proved; there's just the coincidence that Author B cites, near exactly, the same bits of the same material that Author A did. Or sometimes Author B mischaracterizes the content of her sources, exactly as Author A did, as do Authors C, D, etc., resulting in a trail of bungled citations. Other times, there are tells, as when Author B uses the trusty, "see sources cited in Author A." Occasionally, Author B is quite upfront, stating explicitly that he will simply stipulate that Author A has testified accurately on the science at issue.

Don't readily accept testimony. The testifier might have gotten things wrong: perhaps they lack the needed skills to fairly interpret a bit of science, or perhaps they are simply careless. However skilled and attentive, they surely have a theoretical agenda: after all, the point of importing material from one discipline into another is

to force theoretical movement in the importing discipline. If not, why bother? Even if the testimony is accurate, you risk charges of intellectual irresponsibility that undermine your rhetorical credibility: when the matter is controversial, the reporter who doesn't check facts is a sloppy reporter. If the testimony is inaccurate, matters are worse, and you risk being both irresponsible and wrong.

Too great a willingness to accept testimony is how myths get repeated from paper to paper—especially if the myths are good stories. One of our favorites is the famous case of Phineas Gage that has been discussed extensively in moral psychology (e.g., Damasio 1994; Greene 2013). If you trust pop psychology books, scientific articles, and even psychology textbooks, the poor fellow became an unreliable, unruly, violent thug after a three-foot long iron rod went through the front of his brain and out the top of his head. The reality may be less exciting than this fantastic tale. In fact, recent historical investigation indicates that Gage immigrated to Chile subsequent to his injury and became a trusted, respected coach driver (e.g., Macmillan 2002; Griggs 2015).

None of this is to say one should never rely on testimony about another discipline. The interpretations of difficult and controversial literatures by those steeped in them are not valuable, they are invaluable; limitations of time and expertise mean that outsiders must more than occasionally trust the interpretations of insiders. Trust, but, when possible, verify. And to verify, see for yourself. Insider testimony must be subject to critical scrutiny, and that scrutiny requires detailed firsthand experience with the literatures.

Typically, humanists enter empirical literatures via standard textbooks, whether introductions or upper level surveys. There's nothing wrong with this, to make a start. But like philosophers, scientists write with theoretical agendas. To take an example dear to our hearts, many philosophers who have commented on the person-situation debate that bedeviled social and personality psychology for some 30 years—and spawned the situationism-virtue ethics debate in philosophy—got their introduction to the issues through Ross and Nisbett's (1991) excellent *The Person and the Situation*. But Ross and Nisbett aren't neutral reporters; they are among the most prominent social psychologists critiquing personality psychology. It is therefore unsurprising that philosophers whose understanding of the psychology is indebted to them (e.g., Doris 1998, 2002; Harman 1999, 2000) are tempted to skeptical verdicts on philosophical notions of character.

Start with a book written by a personality theorist, such as Funder's (2012) *The Personality Puzzle*, and you may end in a different place. Be careful what book you browse, or what website you click, or your career may be structured by an arbitrary event! (Back when people still went to libraries, Doris stumbled on *The Person and the Situation* while looking for another book and noted that Nisbett was on the faculty at Michigan, where he was doing his degree; Nisbett graciously agreed to join Doris's committee, and a dissertation was born.)

To ameliorate the problem of testimony, start by reading lots of testimony from lots of perspectives. Figure out what the experts think about their field by reading their commentary on their field—and talk to experts about what they believe but can't commit to print; often horse sense may be more illuminating than journal

articles. But also go read what they read, or wrote, to make them believe it, which brings us to **advice (3): *read the original studies*.** Oftentimes, these studies will not be fully interpretable without the guidance of experts, and it's back to the surveys and texts. But if you put the time in, very often, you'll find instructive discrepancies between what is said about the studies and what is shown in the studies themselves. By triangulating across perspectives on a study, you can come to a sensible conclusion about it.

Still, there's only so much time, and you can't verify everything equally. That means one has to perform a kind of triage and determine what gets scrutinized and how much. The first element of this triage is internal to your work. Very likely, some of your empirical claims will be central, and some more peripheral, to your argument and theory; it's prudent to devote more time to support a major claim in a central argument than to the illustration of an offhanded aside. Nothing wrong with including an interesting tangent to your main argument—no one loves an interesting tangent more than us. But don't think the standard of scientific literacy for a good sauce is sufficient for the meat of your work. As everywhere, judgment is required: you must distinguish the meat from the sauce and devote your energies accordingly.

The second element of triage is equally plain: the more controversial the evidence, the more critical scrutiny is required. For example, there has lately been much contention over the extent to which celebrated priming studies—such as the Bargh lab's iconic finding that semantic primes invoking stereotypes of the elderly cause healthy young people to walk more slowly (Bargh et al. 1996)—are replicable by independent labs. In our view, RepliGate—the crisis in contemporary social psychology (and a few other areas in psychology) due to a surprisingly high number of replication failures—doesn't suggest that there's nothing to priming (we're pretty confident there is), but it does suggest that the consumer must take extra care to get a handle on the literature (for discussion, see Doris 2015: 44–49).

Advice (4): read the experimental parts of the articles (often called *Methods* and *Results*), not only the introduction and discussion. Introductions of articles in the behavioral sciences expose the competing theories and points of view about a given empirical issue, review the existing literature, and formulate the authors' hypotheses. Think about it as summarizing what the authors think we already know. The discussion summarizes the empirical results, elaborates on them in light of the authors' hypotheses, and discusses the studies' limitations. The methods and results sections are the most important parts of scientific articles: they present what scientists have done and what results they have obtained. Read them with care. The conclusions scientists claim to be entitled to are sometimes loosely connected with what the experiments really show. Sometimes, wrinkles are ironed out, generalizations asserted that go beyond the limited scope of the experiments, and grandiose conclusions asserted. Do not take scientists at their word; look at what they have done and shown, as presented in the methods and results sections. Of course, this requires a minimum of scientific and statistical expertise; we offer some guidelines below.

Also do not assume too quickly that scientists and philosophers always mean exactly the same thing when using the same words, and be sensitive to the different uses of the same language in different disciplines. For instance, saying that people

are utilitarian does not exactly mean the same thing in some parts of moral psychology and in ethics.

Advice (5): *consider* lots *of studies.* The one-result philosophy paper, where the central argument is based on a solitary scientific finding, is a familiar sort of embarrassment. Equally familiar is the cherry-picked paper, which only reports the findings most congenial to the author's view and neglects the others. Don't treat any piece of evidence or empirical result as if it could ever conclusively show anything. Scientists always view any piece of research as just one piece of the puzzle. As they know, other results are needed to confirm any piece of research, and these may turn out to undermine it. Conversely, don't assume that you have shown much if you have found some aspect of a scientific article that can be disputed or questioned. Scientists are well aware that every article has some flaw or other. Convergent lines of research compensate for the flaws any single article may have. For instance, in response to criticisms (Berker 2009), Greene (2014) reviews an impressively large and diverse body of findings in support of his dual-process model of moral judgment. Focusing on the flaws in any line of research is missing the forest for the trees.

The obvious advice then, the importance of which cannot be overestimated: consider the full range of studies, both congenial and contrary to your perspective, and identify the dominant trends.

We say "dominant" with malice aforethought, since dominant is the best one can hope for: it will be extremely rare, given the uncertainty of the empirical world and the dialectical character of science, for *every* finding in a literature to tell in the same direction. Even where robustly univocal trends emerge—smoking and cancer, say, or anthropogenic climate change—they may only emerge in the fullness of years, when we're needing more empirically credible theorizing right now. This doesn't mean one shouldn't attempt empirically informed theory—empirically *un*informed theory ain't likely to be better—but it does mean that one should theorize self-consciously. Know what seems solid, and what seems speculative, and be explicit about it. The best one can do, more often than not, is make a responsible wager on the state of the science.

Advice (6): *check the meta-analyses*, *carefully*. Meta-analyses (quantitative summaries of research literatures) help to identify the dominant trends within research literatures. By aggregating across a number of studies examining a given scientific question, they prevent unfortunate reliance on outlier studies, and they should help you avoid cherry-picking studies to make a theoretical point. But while useful, meta-analyses are not without problems. Necessarily, meta-analysts make many editorial decisions, and these decisions leave ample room for subjective, and potentially controversial, choices. First and foremost, they need to decide which studies to include, and different inclusion criteria—based on the assessed quality of the studies, their experimental designs, the hypotheses explicitly tested, etc.—can result in contradictory conclusions.

For example, the literature on stereotype threat—viz., the hypothesis that individuals from groups such as women and African Americans perform less well in tests when they are reminded of their identity (Spencer et al. 1999)—is one of the examples used by Gendler (2011) to argue that recent research on unconscious

influences on behavior reveals a conflict between our moral and epistemic commitments. But does this literature really establish the reality of stereotype threat? Nguyen and Ryan (2008) meta-analyzed the literature on stereotype threat. Using a permissive inclusion criterion, they found evidence for a small effect of stereotype threat for women and a larger effect for African Americans. By contrast, Stoet and Geary (2012) adopted a conservative inclusion criterion for assessing a more specific hypothesis: stereotype threat partly explains the gender gap in some areas of mathematics. They discounted most studies examining stereotype threat on various grounds, examining only studies that involved a mathematical test that had a control group of male participants, that did not recruit participants on the basis of their knowledge of gendered stereotypes, and that assigned participants to different conditions for manipulating stereotype cuing. Their conclusion was that the studies meeting their conservative criterion fail to support the stereotype threat hypothesis. It would be convenient if there were universally accepted criteria for deciding what studies to include in a meta-analysis, but this aspiration is utopian: judgment calls play an important, irreplaceable role.

And there are many such calls to be made. Meta-analysts can, but need not, include non-published studies in their meta-analysis. Meta-analysts can, but need not, weigh the included studies as a function of their perceived scientific quality. They can, but need not, give more weight to the studies with larger sample sizes. They must select a particular data analytic strategy to compute the aggregated effect size, with different strategies sometimes resulting in contradictory results. They can aggregate the data in many different ways and examine various variables possibly moderating the effect of interest. For instance, they can examine whether stereotype threat effects emerge when the test is given by a man or a woman, whether and how they depend on the nature of the cue reminding participants of their identity, whether and how they vary as a function of participants' self-identification with the test area (e.g., mathematics), and so on.

So, meta-analyses too can be tendentious! Unsurprisingly, then, they may fail to quench controversy. For instance, Baumeister's ego-depletion phenomenon, according to which "will" (a capacity for self-control) is a finite resource (e.g., Baumeister et al. 1998), has been widely discussed in the philosophy of action (e.g., Levy 2011), but its reality is now questioned (e.g., Job et al. 2010; Lurquin et al. 2016) in spite of a supportive meta-analysis (Hagger et al. 2010), because just how supportive this "supportive meta-analysis" is has been contested by alternative analyses (Carter and McCullough 2013, 2014).

What to do when meta-analyses or informal literature reviews are not available? **Advice (7):** ***don't rely on a single investigator***—even where they've produced many studies. Not surprising that a scientist tends to publish findings that support one another and jointly support their favored hypothesis. Hesitate to take such a cluster as independent evidence for a hypothesis; sometimes, it's more like reading something in multiple copies of the same newspaper. (Indeed, a single data set often funds many papers.) A bit less obvious is to avoid relying on a single lab, even where the author groups vary. If you attempt constructing an empirically grounded theory on the output of a single lab and that lab founders—due to difficulties in

replication, countervailing evidence from elsewhere, or (hopefully, less commonly) scientific misconduct—the empirical grounding of your theory has been eroded from under your feet. We realize that sometimes this will mean abstaining from exploiting exciting new results that would (if solid!) advance your theoretical agenda, but the risks will often outweigh the rewards, unless the result is supported by better established findings that are *very* closely aligned with the new work with respect to conceptual implications.

If possible, it's good to have evidence beyond a "lab family" consisting of a parent lab and offspring labs run by its students: again, no surprise if the findings of such clans trend in the same direction, since they will typically be generated by broadly similar technique and theory. Beware too of citation circles, clusters of researchers who, while not of the same immediate academic family, nonetheless advocate related, and mutually supportive, research programs—and cite one another accordingly. There's no easy formula here: one person's widely influential research program is another's incestuous citation circle.

And don't confuse either large citation numbers or ongoing patterns of citation with trustworthiness. You may be surprised how often scientists keep citing results that have failed to replicate or that have been seriously challenged, either because they just are not on top of the literature or because they have an ax to grind. It is quite dispiriting when thoroughly debunked studies are cited approvingly by scientists, who should know better, or by philosophers, who should be more careful and skeptical!

To wit, in an influential article, Boroditsky (2001) argued that East Asians conceive of time differently from Westerners because vertical metaphors of time are more common in East Asian cultures, while horizontal metaphors are more common in Western cultures. While her original article presented striking results, there have been ten failed attempts at replicating her original results (Chen 2007; January and Kako 2007; but see Boroditsky et al. 2011). Yet Boroditsky's early article continues to garner approving citations from both psychologists and philosophers.

Likewise, several influential moral psychology studies frequently exploited by philosophers have now hit hard times. Schnall and colleagues' (2008) famous study, according to which priming people with purity thoughts makes moral judgment less severe, has not always been replicated (Johnson et al. 2014, 2016; but see Huang 2014). The same is true for the Valdesolo and DeSteno (2006) study allegedly showing that participants are more likely to push the large person in the "footbridge case" after having watched a funny skit from the television how *Saturday Night Live* (Seyedsayamdost 2014; Duke and Bègue 2015). The same for Zhong's Lady Macbeth effect, according to which cleanliness leads to more severe judgments (Fayard et al. 2009 and Earp et al. 2014 on Zhong and Liljenquist's 2006; Seyedsayamdost 2014 on Zhong et al. 2010).

To be sure, that a study fails to replicate does not mean that the study was wrong; the replication study may itself be the failure. A fortiori, it does not mean that the author of the original study was doing shoddy work, to say nothing of engaging in "questionable research practices." But it does mean that one should not rely on that study until the dust has settled. To keep track of the dust, it is worth consulting web-

sites such as www.psychfiledrawer.org and osf.io/ezcuj/wiki/home/ and special issues (e.g., issue 3 of volume 45 of *Social Psychology*) reporting the recent results of replication attempts. You'd also be wise to tap "the grapevine" for news of what experimental paradigms are encountering replication trouble.

Not unrelated is **advice (8): *"investigate" the authors on whom you heavily rely*.** A theory built on the output of a lab discredited for scientific misconduct will have a limited shelf life—or it should. How to find such things out? As elsewhere, search engines help, and so too does consulting people working the field—the more the better; to avoid the pitfalls of illusory support, the theorist has to be something of a sociologist of science. You can't overestimate the importance of the gossip network. Psychologists, anthropologists, and cognitive neuroscientists talk, and not always in a friendly manner. At the bar, or during the poster session, they will tell you which scientist has a reputation for shoddy work and which can be trusted. Gossip may seem an ugly word, but however unseemly it may appear, the gossip network plays an essential role in science. Scientists need information to determine whom to trust, whom to submit to scrutiny, and whom to entirely distrust; reputation, built and destroyed through chitchat, is often the only source of information at hand. By all means, rely on more reputable sources as they appear; until then, keep your ear to the ground where you can.

As recommended above, it is *highly* advisable, whenever possible, to check controversial literatures for yourself. So, sometimes you will have to rely on your own judgment. Fortunately, there are a few useful tells that will allow you to identify suspect, if not outright poor, science.

Advice (9): *check the sample.* Psychologists and neuroscientists have too often relied on small sample sizes involving limited numbers of experimental participants (this tendency was a central pathogen in the RepliGate controversy). To see why, a bit of background in experimental design—in the form of a few more obvious observations—is needed. Most psychologists follow the methods introduced by R. A. Fisher in the first half of the twentieth century. In this methodological tradition, when a psychologist runs an experiment, she typically attempts to reject *the null hypothesis*—for example, the hypothesis that there is no difference between a control and an experimental condition or that two variables are not correlated—in order to accept the hypothesis of interest—e.g., that there is a difference between a control and an experimental condition or that two variables are correlated. Sample size is among the most important determinants of a test's *power*, i.e., roughly, the probability of obtaining a significant result if the null hypothesis happens to be false. A power of 0.50 means that the psychologist has one chance in two of rejecting the null hypothesis if the null hypothesis is false. So, when power is equal to 0.50 and the null hypothesis is false, you would do equally well flipping a coin as you would actually running the experiment! A power of 0.80 is often recommended in the behavioral sciences, but for more than 40 years, power in psychology has on average hovered around 0.50 (Sedlmeier and Gigerenzer 1989; Fraley and Vazire 2014).

A simple rule of thumb: the smaller the sample sizes in a literature, the lower the power, and the higher the rate of false positives. Let's see why. The *p*-value is, roughly, the probability of obtaining the data one has obtained (such as, roughly, the

difference between the means of the control and experimental conditions) or more extreme data (e.g., an even larger difference between these two means) if the null hypothesis is true.[2,3] For instance, a *p*-value of 0.03 in an experiment comparing an experimental and a control condition means, roughly, that the probability of obtaining the observed difference between the two means, or an even larger difference, if the null hypothesis is true, is equal to 3%. The significance level, often represented by α, determines the largest value the *p*-value can have for the null hypothesis to be rejected; by convention, it is typically set at 0.05 in psychology, meaning that the null hypothesis can only be rejected if the *p*-value is equal to or smaller than 0.05 (we then have *a significant result*).[4]

Suppose that the significance level is set at 0.05, as is usual. Then suppose that *n*% of null hypotheses in a given research literature happen to be true, and suppose that the power of tests in this research literature is equal to *m*/10 (e.g., 0.5 or 0.8). If a given literature contains 100 articles, we obtain the following proportions for different types of results.

	Significant result	Nonsignificant result
H_0 is false	$0.m \times (100 - n)$	$(100 - n) - 0.m \times (100 - n)$
H_0 is true	$0.05n$	$n - 0.05n$

The top left cell represents the number of *hits* in this 100-article literature: the null hypothesis is rejected when it is false. The lower left cell represents the number of *false positives*: the null hypothesis is rejected when it is true. The top right cell represents the number of *false negatives*: the null hypothesis is not rejected when it is false.

Psychology suffers from *a publication bias*: in general, only significant results are published, and negative results (experiments where the null hypothesis is not rejected) are shelved (a problem known as *the file drawer effect*). While a publication bias is found in all sciences, it appears to be more severe in psychology (Fanelli 2010). As a result, the rate of false positives in a given psychological literature will be $0.05n/(0.m \times (100 - n) + 0.05n)$. This rate increases with decreasing power ($0.m$). For instance, supposing 40% of null hypotheses tested in a given literature in psychology are true, the rate of false positives will be 3.6% if power is equal to 0.8 and 12% if power is equal to 0.5.

[2] What is the relation between *p*-values and power? If there is no effect to be detected (i.e., if the null hypothesis is true), then the *p*-value is uniformly distributed between 0 and 1 (i.e., one is equally likely to get a *p*-value between 0 and 0.2, between 0.2 and 0.4, etc.), independently of the power of the experiment. If the null hypothesis is false, the larger the power, the more likely it is that one will observe a small *p*-value (holding constant the effect size used to compute the power of the experiment).

[3] More precisely, it is the probability of obtaining a statistic (a function of the data such as *t* or *F*) or a larger one if the null hypothesis is true.

[4] In a recent article, Benjamin et al. (2017) have argued that the 0.05 significance level is insufficiently strict, and have recommended to decrease it to 0.005. They argue that findings at the 0.05 level provide too little evidence and that this lax significance level contributes to the current replication crisis in psychology and other sciences. One of us (EM) is a coauthor of this article and would like the 0.005 significance level to be widely accepted. For discussion, see however Amrhein and Greenland (2017); Lakens et al. (2017); McShane et al. (2017).

Studies with small sample sizes are also more likely to report an inflated effect size (e.g., the standardized difference between the means of two conditions): that is, the effect size reported by studies with a small sample size is likely to be larger than the true effect size. The reason is simple: everything else being equal, *p*-values decrease with larger sample sizes. When a sample size is small, only effect sizes that are large, including those that by chance are larger than the true effect size, can be significant.

So, beware if the sample size of a given study or if the typical sample size in a given scientific literature is small! What's a small sample size, you ask? Difficult question, since the answer will depend, among other things, on the kind of tests run. But, as a rule of thumb, if the experiment involves a between-subjects design (i.e., an experimental design in which each subject is involved in one and only one experimental condition), be wary if there are fewer than 50 subjects per condition. If the experiment involves a within-subjects design (i.e., an experimental design in which each subject is involved in all the experimental conditions), be wary if the total sample size is smaller than 30 subjects. In any case, the larger the sample size, the better.

Look also at the kind of findings reported by the article you're reading. If the article reports an effect, but only for, say, men over 30 with a conservative bent, be wary. The authors were probably hoping to report the effect for all participants but, being unable to find it, examined whether the effect held for various subgroups within their participants. In brief, you are probably looking at the result of a fishing expedition, the kind of result that is less likely to replicate because the authors are capitalizing on chance.

Low power and fishing expeditions are not the only causes of RepliGate. Selective reporting (running many studies and reporting the few that happen to yield a significant result), data peeking, sometimes called "optional stopping" (computing a *p*-value while collecting data and adding data until a significant result is obtained), and the exclusion of outliers on subjective grounds also increase the rate of false positives in published literature (Simmons et al. 2011). But the reader is often not in a position to find out whether such "p-hacking" occurred, while low power and fishing expeditions are relatively easy to spot.

Advice (10): ***size matters.*** Check the *p*-values reported in the article (i.e., those *p*-values that matter for the findings touted by the authors). If they are all near 0.05, be wary again. It is possible that the authors ran many more studies and reported only those that turned out to be significant, and all the results serendipitously appearing at the minimum publishable significance resulted not from a real effect but from the investigators capitalizing on chance by *p*-hacking.

Even where significance is come by honestly, it's only a small part of a complicated story. Encouraged by the quasi-magical aura surrounding statistical significance among psychologists themselves (understandably enough, since a significant result can be a make-or-break point in a scientist's career), philosophers often only pay attention to whether the study reports a significant result. Statistical significance is of course important, but there is more to a scientific report. In addition to finding out whether a result is statistically significant, it is often important to find out whether it is a small, a medium, or a large effect.

You probably want to know how to recognize small, medium, and large effects, don't you? Sometimes, this is an easy task: sometimes, the observed variable (what psychologists call "the dependent variable") has a meaningful metric (e.g., weight loss in nutrition studies). In other contexts, the dependent variable has no meaningful metric, and it is then more difficult to know what small and large effect sizes look like. Fortunately, psychologists (particularly, Jacob Cohen) have developed conventional benchmarks for assessing effect size: on this approach, a small effect size is an effect size that is smaller than the typical effect size in psychology (start with Cohen 1992). It is common to use an index called "Cohen's d" (often represented as *d*) to report the effect size. *d* reports the standardized difference between the means of two conditions (i.e., the difference between these two means divided by the standard deviation of the control condition). By convention, 0.2 is a small effect, 0.5 a medium effect, and 0.8 a large effect; 0.2 indicates that the two means differ by a fifth of the standard deviation and also that the mean of the experimental condition is at the 58th percentile of the distribution of the control condition—that is, if the data are normally distributed, 50% of the participants in the experimental condition have a higher score than 58% of the participants in the control condition. 0.5 indicates that the two means differ by half of the standard deviation. It also indicates that the mean of the experimental condition is at the 69th percentile of the distribution of the control condition. 0.8 indicates that the two means differ by four-fifths of the standard deviation. It also indicates that the mean of the experimental condition is at the 79th percentile of the distribution of the control condition. As noted above, keep also in mind that the effect sizes that are reported in studies with small sample size tend to be inflated, sometimes substantially so: the true effect size is likely to be smaller.

Another good reading habit is to treat *p*-values with care. As we have seen, a *p*-value reports, roughly, the probability of obtaining the data one has obtained or more extreme data if the null hypothesis—the hypothesis one is typically trying to reject—is true. A common mistake is to take *p*-values to be the probability that the null hypothesis is true, given the data obtained (what is commonly called *the posterior probability*). On this misinterpretation, a low *p*-value would then say that the null hypothesis is likely to be false. But that's just not what a *p*-value means, and it is easy to see why. The posterior probability of the null hypothesis depends on its *prior probability*—roughly, how likely to be true the null hypothesis was independently of the data obtained in the experiment. The lower the prior probability, the lower the posterior probability. The *p*-value says nothing about the prior probability and thus could not report a posterior probability. So, don't conclude that the null hypothesis is likely to be false and the hypothesis of interest likely to be true because the *p*-value is small.

Another mistake is to assume that a *p*-value determines the probability of replicating the significant finding: On this misinterpretation, a *p*-value of 0.05 indicates that the probability of replicating the finding is 0.95. But, again, that's just not what a *p*-value means. Let's suppose that the null hypothesis is indeed false and that one has correctly rejected it but that the power of both the original study and of its replication is equal to 0.5. Then, the probability of replicating the original result is not

0.95 but 0.5. So, don't assume that an experimental result will replicate because the *p*-value is small.

Now, consider the following exercise (based on Oakes 1986: 79–82):
Suppose you have a treatment that you suspect may alter performance on a certain task. You compare the means of your control and experimental groups (say 20 subjects in each sample). Further, suppose you obtain a *p*-value below the significance level ($p = 0.01$). Please mark each of the statements below as "true" or "false."

1. You have absolutely disproved the null hypothesis (that there is no difference between the population means).
2. You have found the probability of the null hypothesis being true.
3. You have absolutely proved your experimental hypothesis (that there is a difference between the population means).
4. You can deduce the probability of the experimental hypothesis being true.
5. You know, if you decided to reject the null hypothesis, the probability that you are making the wrong decision.
6. You have a reliable finding in the sense that if, hypothetically, the experiment were repeated a great number of times, you would obtain a significant result on 99% of occasions.

Take your time.

Now, if you have marked any of these statements as true, you have made a mistake. Statements 1–5 are related to the first misunderstanding pointed out earlier: They confuse the *p*-value, which says something about the data obtained (more precisely, it measures how likely it is to obtain these or more extreme data if the null hypothesis is true), with the probability of the null hypothesis, which measures how likely the null hypothesis is to be true. To repeat, *p*-values do not report the probability of the null hypothesis; a very low *p*-value does not mean that the null hypothesis is likely to be false. Statement 6 is related to the second misunderstanding pointed out earlier: it confuses the *p*-value with the probability of obtaining a significant result in a replication. Don't be too embarrassed if you made a mistake: mistakes were extremely common among the 70 psychologists Oakes examined.

Relatedly, you should not be too impressed by very low *p*-values (Look! $p = 0.001$!). *P*-values depend on several factors, including the precision of the measures used in an experiment and the sample size. A low *p*-value just means that it is extremely unlikely to obtain the data one has obtained or more extreme data if the null hypothesis is true; it does not mean that one has obtained a very large effect size—indeed, very low *p*-values are compatible with tiny effect sizes. Nor, to repeat, does a low *p*-value mean that the null hypothesis is likely to be false or that one is likely to replicate the finding.

While we're talking about *p*-values, how should you interpret nonsignificant results (when $p > 0.05$)? Of course, you can't just infer that the null hypothesis is true from a nonsignificant result, because nonsignificance may be due to low power. If the power of the test is equal to, e.g., 0.2, the probability of obtaining a

significant result if the null hypothesis is false is equal to 0.2. Also, do not rely on the following common, but nonetheless erroneous, rule of thumb: the higher the *p*-value, the more likely it is that the null hypothesis is true. This rule of thumb is mistaken because if the null hypothesis is true, a *p*-value between 0.05 and 0.1 is as likely as a *p*-value between 0.5 and 0.55 or between 0.9 and 0.95 (i.e., *p* is uniformly distributed).

Then how to interpret a negative result and how to accept the null hypothesis on the basis of a negative result? We recommend accepting the null hypothesis on the basis of a negative result if and only if the power of the test is high: as a rule of thumb, at least 0.8 (Machery 2012b). When power is not reported or when it is low, inferences that an independent variable has no influence on the dependent variable are illegitimate, and negative results should not be interpreted.

Advice (11): ***be alert to biases.*** Scientists are people. And they're people with theories. It's well documented that human beings are subject to "motivated cognition," the tendency to form congenial beliefs and resist forming uncongenial ones (Dunning 1999; Gilovich 1991; Kunda 1990). In the present context, this means you should expect investigators to be good at finding fault with arguments and evidence inconsistent with their own theories and less good at critically scrutinizing congenial argument and evidence. Don't be surprised if scientists' standards are sterner when assessing experimental results that would challenge their own research program.

Equally important, keep in mind that you are a person too and as likely to be a motivated cognizer as the scientists you scrutinize (as, of course, are we!). You also are, or will be, a person with a theory: perhaps you think this or that article is weak because it would undermine a favored conclusion of yours, were it correct. Similarly, maybe you think this or that study is "great stuff," in substantial measure because it gives you reason to hold a position you already hold.

Meta-analyses may help here, by compelling you to take into account many studies instead of just those fitting your preconceptions. It also pays to keep an eye on biases known to influence the interpretation of scientific findings. For example, readers seem to be unduly influenced by neuroscientific verbiage in explanations of human behavior—although the science on this matter itself suffers from replication issues (Weisberg et al. 2008; Farah and Hook 2013; Fernandez-Duque et al. 2015)! Don't get distracted by allusions to the lighting or firing of brain areas, and make sure that such allusions are really playing an explanatory role. While knowing about biases may not be enough to ameliorate them, a vivid appreciation of our cognitive infirmities may at least inspire us in our effort to "do the right thing" in our encounters with science (such as following the advice in this letter!).

Like other scientists and, we dare say, like philosophers themselves, psychologists and cognitive neuroscientists are not dispassionate seekers of the truth. (Really, what fun would that be?) Rather, they are often eager to promote their views and to be recognized for their contribution. Social goods like fame and material goods like salary and reduced teaching load accompany success. Unsurprisingly then, exaggeration and misrepresentation are not unheard of in science. Hype is a regrettably common gambit in the science game, especially when science is packaged for popu-

lar audiences. Admittedly, separating the wheat from the chaff is often difficult because scientists do not distinguish them in their own writings, possibly because they themselves are not always clearly drawing the distinction. So, caveat emptor—especially when reading the *Times* science pages.

Nor is science free of ideological biases. Concerned as they are with behavior, psychology and cognitive neuroscience are particularly likely to suffer from ideological biases, and, unsurprisingly, psychology has a checkered history with respect to race and gender. While psychology has undoubtedly furthered sexist and racist ideologies in the twentieth century, some psychologists have recently expressed concerns that biases have nowadays tilted toward the left and that this is hindering scientific progress (Duarte et al. 2015).

Consider what apparently happened to a recent article by Williams and Ceci (2015).[5] For years, Ceci and Williams have been accumulating actuarial data suggesting that the gender imbalance in STEM disciplines is not a result of biases against female applicants and scientists, a claim definitely at odds with the dominant opinion in this area of psychology; their 2015 *Proceedings of the National Academy of Sciences* article reports some experimental data in support of this contention. We're noncommittal about whether they are correct or not; what interests us here is the extraordinary review process their 2015 paper had to go through. Apparently, their article was reviewed by *seven* reviewers, while a 2012 study arguing for the existence of biases was only reviewed by two reviewers (the usual number, according to *PNAS*). Furthermore, their data set was examined by an external statistician, a most unusual step in the review process. We doubt that a more ideologically congenial article would have been submitted to anything like the same level of scrutiny. In any case, here is a motto: that some bit of science feels good does not make it right!

Conversely, one must also attend to cases where scientific argumentation is *not* theory driven. Psychologists and neuroscientists typically begin their articles with a justification of the hypotheses submitted to test. This justification alludes to extant theories, ongoing literatures, and the current body of evidence. Examine psychologists' and neuroscientists' justification closely. If the hypothesis seems ad hoc, unprincipled, or not supported by any theory in the existing literature, be wary. It is plausible that the psychologists had another (more principled) hypothesis in mind when they designed the experiment, but that, failing to confirm their prediction, they fabricated another prediction in order to salvage their experiment and to publish an article. If so, you are probably looking at the result of a fishing expedition, which may well not be supported by future replications. The hypothesis, formulated post hoc, is designed to fit maximally the data obtained in an experiment; thus, it overlooks the fact that every body of data is influenced by random factors and will thus not be identically reproduced in future replications. This kind of fishing expedition is like hypothesizing that a coin is unfair and yields six heads for four tails after having seen three heads out of five coin throws: such a hypothesis maximally fits the

[5] http://chronicle.com/article/Passions-Supplant-Reason-in/232989

observed data but overlooks that the data one has just observed are in part due to random factors.

Advice (12): take the long view. The latest may not be the best; it may not even be good. Until a trend is very well established, extreme caution is in order. Indeed, since even "established trends" may be destabilized, caution is *always* in order. Science can be faddish: scientists—especially graduate student scientists looking to "make their mark"—often board bandwagons, be it a particular theoretical approach, a particular research topic, a particular experimental paradigm, or a particular data analytic method. Of course, some current fashions turn out to constitute real scientific progress: the point is that it's hard to know in the moment.

Relatedly, be aware of what types of research are currently controversial. An important example is reverse inference, where the fact that participants have a particular mental state, such as experiencing a particular kind of emotion, is inferred from the fact that a particular part of their brain "lights up" in a neuroimaging study. Reverse inference is the object of a heated debate in cognitive neuroscience and in the philosophy of cognitive science (e.g., Poldrack 2006; Machery 2014; Glymour and Hanson 2016). In a nutshell, the issue is that when they reverse infer, cognitive neuroscientists often only take into account the probability of the activation of a given area (e.g., the insula) if participants entertain a particular mental state (e.g., disgust), i.e., the conditional probability P(insula activation | participants feel disgust), while overlooking the probability of the activation of this very area if participants entertain another mental state, P(insula activation | participants *do not* feel disgust). This is a bit like concluding that it has rained because the sidewalks are wet on the grounds that the sidewalks are likely to be wet if it rained while overlooking the fact that the sidewalks could be wet even if it had not rained (e.g., if they had been cleaned). At the same time, this technique has played an important role in recent moral psychology. Greene's classic neuroimaging study of trolley cases, which has been so important in giving rise to a cognitive neuroscience of morality (Greene et al. 2001), is a famous but controversial example of reverse inference. Whether or not reverse inference can be ultimately defended (as we believe), there is no doubt that it has very often been misused in cognitive neuroscience, and you'd be wise to look closely at articles relying on it.

Obviously, reverse inference is not the only form of controversial research practice. We have already mentioned priming: these striking studies, which attempt to manipulate participants' behavior by means of unconscious primes, are at the center of the RepliGate controversy, and we advise you to be critical of studies using this type of manipulation. Of course, this is easier said than done! One of us (EM) was inspired by a striking result in social psychology—hard-to-read materials prime people to be more thoughtful (Alter et al. 2007)—and used this priming manipulation in his own empirical research. Here is a bit of background. Sytsma and Machery (2010) argued that the lay concept of subjective experience does not correspond to the philosophical concept of phenomenal consciousness (roughly, the idea that there is something it feels like to have a perceptual experience). In response, Talbot (2012) proposed that Sytsma and Machery's vignettes had simply elicited "System 1" (roughly, fast, non-reflective) judgments that did not reflect people's genuine con-

cept of consciousness. To test Talbot's empirical conjecture, Sytsma and Machery (2012) attempted to show that their results did not change when care was taken to elicit slow, reflective ("System 2") judgments. One of their manipulations was inspired by a then much talked about psychology paper (Alter et al. 2007), which suggested that participants were more careful and reflective when they were presented with texts difficult to read (e.g., printed in a hard-to-read font). Sytsma and Machery reasoned that if Talbot were right, people would make different judgments when presented with easy-to-read vs. hard-to-read vignettes. Since they found no effect (a negative result) in a highly powered study, they concluded that Talbot's hypothesis was false. The twist is that we know now that Alter and colleagues' original manipulation does not succeed in eliciting slow, reflective judgments (Meyer et al. 2015): in fact, hard-to-read materials do not prime people to be more thoughtful, and Sytsma and Machery's negative result could not be taken to undermine the theory they were criticizing.

To here, we've been focusing on humanists as consumers. But at many institutions, academic humanists are subject to the same "publish or perish" incentives that help animate science, which means that, if you like to eat, you'll probably also be a producer, writing up what you've learned from the science. We do not have the space for a writing workshop here, although the importance of good writing cannot be overstated; indeed, it might be the single meritocratic property that most publications in good venues have in common, across a range of topics and disciplines (the role of extra-meritocratic considerations like pedigree is an unfortunate issue we will pass over in stony silence [but see Peters and Ceci 1982]).

Given limitations of space, we'll limit our advice on writing to a bit of **advice (13)** developed in the contexts of humanists writing on science: ***be persuasive***. Hopefully, this doesn't sound as sneaky as it might have before you started reading our letter: scientists are not, and indeed cannot be, dispassionate excavators of fact, and the same is no less true of humanists excavating science. Even for literature surveys, such as encyclopedia articles, with the most innocently pedagogic aims, the science writer has, at a minimum, one persuasive goal—to persuade the audience that she speaks authoritatively on the science. But many of you, we hope, will encounter science in the role of theoretical agitator, intent on persuading people that the theory you favor is compelling. Here, too, you have to legitimize your mantel of authority, by speaking with evident competence on the empirical evidence you discuss.

Initially, the best advice we can give is to follow the advice we've already given. Consume the science responsibly—nothing sneaky about that—and you're a long way toward home. But you still have to write it up, as convincingly as possible. Williams (1985: 39) observed, not without justice, that Aristotle's doctrine of the mean is "one of the most celebrated and least useful parts of his system," but here, it is just the guidance we need. Provide neither too much detail nor too little. Too much, and the reader will check out, and readers in states of catatonic boredom are not easily persuaded. Too little, and the reader will be unable to get their own sense of the science, which is required if they are to develop some sympathy with your interpretation: a reader constantly exclaiming "but wait!" or "what about…?!" is another sort of reader not easily persuaded.

How much detail is enough is a question of art, and the exact formula will vary with the contexts. This advice, we realize, is even less informative than the doctrine of the mean, but perhaps we can offer something of help. Some experiments are especially memorable or striking, and some lend themselves especially well to straightforward summary, and reporting these experiments will make for effective writing. Of course, there may be important experiments that are not easy to recount (the sophisticated designs in cognitive psychology, e.g., can make them difficult to handle). Here, you will have to decide whether the experiment's importance to your argument justifies trying the patience of your reader. This principle, of course, generalizes: the more central the experiment is to your purposes, the more advisable it is to spend expository space on it.

But you can't spend space on every deserving experiment, so what to do? This, maybe: denote some experiments your central "exhibits." Ideally, these experiments are memorable, expositorily tractable, not unduly controversial, and centrally implicated in your argument. Where an experiment is difficult to concisely explain, be sure your argument requires it. Then, situate your exhibits in the literature, explaining how they fit or (more rarely) fail to fit with established general trends.

This two-pronged strategy is crucial. Good writing is concrete and particular, and reporting general trends without vivid examples will convince no one. As cognitive science tells us (e.g., Bell and Loftus 1989), anecdotal evidence is weighted heavily, and the science writer who wishes to convince her readers would be foolish not to take advantage of this tendency. At the same time, as should by now be clear, a study or two should not convince a savvy consumer, so you must situate your exhibits within judicious summations of wider trends. With both pieces in place, one can craft, with luck, an argument that is both vivid and compelling, with the result that people may both remember your writing and be persuaded by it. If you can do that, we've done our job.

With respect to writing, we'll also offer this heartfelt **advice (14):** ***collaborate.*** Reading science and talking to scientists are indispensible, but better still is *working* with scientists: if you want to see how the sausages get made, best to make some sausages. You'll likely learn more from someone who's invested in a product on which her name will appear than you will from a (possibly offhand) response to an email or a (possibly tipsy) conversation at a party. So, if at all possible, try to coauthor some papers, and even if you don't coauthor, try to spend some time in a lab or at least sit in on a course or two.

For humanists, collaborating makes possible participating in projects, such as those requiring advanced statistics, that you couldn't do on your own. At the same time, you're possessed of expertises, such as in writing and theorizing, that make possible projects your scientist collaborators couldn't do on *their* own. There's charcuterie in the humanities too, and outsiders are at serious risk of errors, just as in the sciences. With the best sort of interdisciplinary collaboration, then, everybody benefits, and most importantly, in this short life we lead together, everybody gets to have a good time.[6]

[6] Since we're trying to be encouraging and cheerful, we've relegated this **advice (15)** to a footnote: ***when you collaborate, be* very explicit *about work responsibilities and authorial order,* from the**

Concluding advice (16): ***don't be discouraged, just put the work in.*** By now we hope you have an idea of how much sausage making is going on in the making of science, a circumstance that may not increase your appetite for the consumption of the product. In short, it ain't always pretty. Science isn't made by algorithms and computer programs. As we've said, it's made by people and is therefore sometimes driven by personal interests. Moreover, scientific inquiry involves countless pragmatic decisions that could just as easily have been made otherwise, and there are always many degrees of freedom in how data can be analyzed and findings reported, meaning that the process of inquiry is rife with contestable judgment calls.

While labs are home to human idiosyncrasies and biases, science as an institution suffers from systemic biases, such as the well-documented bias against publishing negative results, which limits science's capacity for auto-correction. And this is to say nothing about the extent to which the institutions societies charge with "the production of knowledge" may reflect the inequities and injustices of those societies, arguably to disastrous epistemic, as well as ethical, effect. All this noted, it is perhaps surprising that science is, in the aggregate, such a successful form of inquiry.

Science consumers, then, should keep the messy nature of science making in mind. Science is fallible: truths of the day will often turn out to be tomorrow's past mistakes, and scientific wisdom is almost always provisional. Our confidence in the cutting edge of science should thus always be guarded, and we should not be utterly disgusted when the scientific consensus we relied on falters. We take science as it is, we try our best to understand its dominant trends, and we acknowledge its fallible nature.

At this point, you may be worried that engaging with science is just too difficult for an outsider, and you can't possibly sort out the controversies in an intellectually responsible way. This, however, is *not* our take-home message. Many of the pieces of "obvious advice" in this epistle are commonsensical and easy to follow, and putting them into practice would improve substantially humanists'—and perhaps, scientists'—use of empirical literatures.[7]

---

outset. We've seen, and been party to, more than one unfortunate misunderstanding about contributions and credit, and they're no fun at all. While the occasional tiff is likely unavoidable, hammering out expected contributions and credit before serious work begins is a very useful preventative. Potential discomfort isn't a reason not to collaborate, but it is a reason to be very clear.

[7] A draft of this paper was presented to the Autumn 2015 meeting of the Moral Psychology Research Group. Many thanks to participants for comments, especially Fiery Cushman, Valerie Tiberius, Maria Merritt, Eddy Nahmias, and Shaun Nichols. We also would like to thank Wesley Buckwalter, David Danks, Benjamin Voyer, and Wayne Wu for comments and suggestions.

References

Alter, A. L., Oppenheimer, D. M., Epley, N., & Eyre, R. N. (2007). Overcoming intuition: Metacognitive difficulty activates analytic reasoning. *Journal of Experimental Psychology: General, 136*, 569–576.

Amrhein, V., & Greenland, S. (2017). Remove, rather than redefine, statistical significance. Nature Human Behaviour.

Bargh, J. A., Chen, M., & Burrows, L. (1996). Automaticity of social behavior: Direct effects of trait construct and stereotype activation on action. *Journal of Personality and Social Psychology, 71*, 230–244.

Baumeister, R. F., Bratslavsky, E., Muraven, M., & Tice, D. M. (1998). Ego depletion: Is the active self a limited resource? *Journal of Personality and Social Psychology, 74*, 1252.

Bedke, M. S. (2012). Against normative naturalism. *Australasian Journal of Philosophy, 90*, 111–129.

Bell, B. E., & Loftus, E. F. (1989). Trivial persuasion in the courtroom: The power of (a few) minor details. *Journal of Personality and Social Psychology, 56*, 669.

Benjamin, D. J., Berger, J. O., Johannesson, M., Nosek, B. A., Wagenmakers, E.--J., Berk, R., Bollen, K. A., Brembs, B., Brown, L., Camerer, C., Cesarini, D., Chambers, C. D., Clyde, M., Cook, T. D., De Boeck, P., Dienes, Z., Dreber, A., Easwaran, K., Efferson, C., Fehr, E., Fidler, F., Field, A. P., Forster, M., George, E. I., Gonzalez, R., Goodman, S., Green, E., Green, D. P., Greenwald, A., Hadfield, J. D., Hedges, L. V., Held, L., Ho, T.--H., Hoijtink, H., Jones, J. H., Hruschka, D. J., Imai, K., Imbens, G., Ioannidis, J. P. A., Jeon, M., Kirchler, M., Laibson, D., List, J., Little, R., Lupia, A., Machery, E., Maxwell, S. E., McCarthy, M., Moore, D., Morgan, S. L., Munafó, M., Nakagawa, S., Nyhan, B., Parker, T. H., Pericchi, L., Perugini, M., Rouder, J., Rousseau, J., Savalei, V., Schönbrodt, F. D., Sellke, T., Sinclair, B., Tingley, D., Van Zandt, T., Vazire, S., Watts, D. J., Winship, C., Wolpert, R. L., Xie, Y., Young, C., Zinman, J., & Johnson, V. E. (2017). Redefine Statistical Significance. Nature Human Behaviour. http://rdcu.be/wEgG.

Berker, S. (2009). The normative insignificance of neuroscience. *Philosophy & Public Affairs, 37*, 293–329.

Boroditsky, L. (2001). Does language shape thought? Mandarin and English speakers' conceptions of time. *Cognitive Psychology, 43*, 1–22.

Boroditsky, L., Fuhrman, O., & McCormick, K. (2011). Do English and Mandarin speakers think about time differently? *Cognition, 118*, 123–129.

Carter, E. C., & McCullough, M. E. (2013). Is ego depletion too incredible? Evidence for the overestimation of the depletion effect. *Behavioral and Brain Sciences, 36*, 683–684.

Carter, E. C., & McCullough, M. E. (2014). Publication bias and the limited strength model of self-control: Has the evidence for ego depletion been overestimated? *Frontiers in Psychology, 5*, 823.

Chen, J. Y. (2007). Do Chinese and English speakers think about time differently? Failure of replicating Boroditsky (2001). *Cognition, 104*, 427–436.

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*, 155–159.

Copp, D. (2007). *Morality in a natural world*. Cambridge, UK: Cambridge University Press.

Damasio, A. R. (1994). *Descartes' error: Emotion, rationality and the human brain*. New York: Putnam.

Daniels, N. (2013). Reflective equilibrium. In W. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Winter 2013 Edition). http://plato.stanford.edu/archives/win2013/entries/reflective-equilibrium/.

Darwall, S., Gibbard, A., & Railton, P. (1992). Toward fin de siecle ethics: Some trends. *The Philosophical Review*, 115–189.

Doris, J. M. (1998). Persons, situations, and virtue ethics. *Noûs, 32*, 504–530.

Doris, J. M. (2002). *Lack of character: Personality and moral behavior*. New York: Cambridge University Press.

Doris, J. M. (2015). *Talking to our selves: Reflection, ignorance, and agency*. Oxford: Oxford University Press.
Doris, J. M., & Stich, S. P. (2005). As a matter of fact: Empirical perspectives on ethics. In F. Jackson & M. Smith (Eds.), *The Oxford handbook of contemporary philosophy*. Oxford: Oxford University Press.
Doris, J. M., Machery, E., & Stich, S. (2017). Can psychologists tell us anything about morality? *The Philosophers' Magazine*, (77), 24–29.
Duarte, J. L., Crawford, J. T., Stern, C., Haidt, J., Jussim, L., & Tetlock, P. E. (2015). Political diversity will improve social psychological science. *Behavioral and Brain Sciences, 38*, 1–13.
Duke, A. A., & Bègue, L. (2015). The drunk utilitarian: Blood alcohol concentration predicts utilitarian responses in moral dilemmas. *Cognition, 134*, 121–127.
Dunning, D. (1999). A newer look: Motivated social cognition and the schematic representation of social concepts. *Psychological Inquiry, 10*, 1–11.
Earp, B. D., Everett, J. A. C., Madva, E. N., & Hamlin, J. K. (2014). Out, damned spot: Can the "Macbeth Effect" be replicated? *Basic and Applied Social Psychology, 36*, 91–98.
Fanelli, D. (2010). Positive. Results increase down the hierarchy of the sciences. *PloS One, 5*, e10068.
Farah, M. J., & Hook, C. J. (2013). The seductive allure of "seductive allure". *Perspectives on Psychological Science, 8*, 88–90.
Fayard, J. V., Bassi, A. K., Bernstein, D. M., & Roberts, B. W. (2009). Is cleanliness next to godliness? Dispelling old wives' tales: Failure to replicate Zhong and Liljenquist (2006). *Journal of Articles in Support of the Null Hypothesis, 6*, 21–30.
Fernandez-Duque, D., Evans, J., Christian, C., & Hodges, S. D. (2015). Superfluous neuroscience information makes explanations of psychological phenomena more appealing. *Journal of Cognitive Neuroscience, 27*, 926–944.
FitzPatrick, W. J. (2014). Skepticism about naturalizing normativity. *Res Philosophica, 91*(4), 559–588.
Fraley, R. C., & Vazire, S. (2014). The N-pact factor: Evaluating the quality of empirical journals with respect to sample size and statistical power. *PloS One, 9*, e109019.
Funder, D. C. (2012). *The personality puzzle* (6th ed.). New York: W.W. Norton.
Gendler, T. S. (2011). On the epistemic costs of implicit bias. *Philosophical Studies, 156*(1), 33–63.
Gilovich, T. (1991). *How we know what isn't so: The fallibility of human reason in everyday life*. New York: The Free Press.
Glymour, C., & Hanson, C. (2016). Reverse inference in neuropsychology. *The British Journal for the Philosophy of Science 67,* 1139–1153.
Greene, J. D. (2013). *Moral tribes: Emotion, reason and the gap between us and them*. London: Penguin Press.
Greene, J. D. (2014). Beyond point-and-shoot morality: Why cognitive (neuro)science matters for ethics. *Ethics, 124*, 695–726.
Greene, J. D., Sommerville, R. B., Nystrom, L. E., Darley, J. M., & Cohen, J. D. (2001). An fMRI investigation of emotional engagement in moral judgment. *Science, 293*, 2105–2108.
Griggs, R. (2015). Coverage of the Phineas Gage story in introductory psychology textbooks: Was gage no longer gage? *Teaching of Psychology, 42*(3), 195–202.
Hagger, M. S., Wood, C., Stiff, C., & Chatzisarantis, N. L. (2010). Ego depletion and the strength model of self-control: A meta-analysis. *Psychological Bulletin, 136*, 495–525.
Harman, G. (1999). Moral philosophy meets social psychology: Virtue ethics and the fundamental attribution error. *Proceedings of the Aristotelian Society, 99*, 315–331.
Harman, G. (2000). The nonexistence of character traits. *Proceedings of the Aristotelian Society, 100*, 223–226.
Huang, J. L. (2014). Does cleanliness influence moral judgments? Response effort moderates the effect of cleanliness priming on moral judgments. *Frontiers in Psychology, 5*, 1276.
January, D., & Kako, E. (2007). Re-evaluating evidence for linguistic relativity: Reply to Boroditsky (2001). *Cognition, 104*, 417–426.

Job, V., Dweck, C. S., & Walton, G. M. (2010). Ego depletion—Is it all in your head? Implicit theories about willpower affect self-regulation. *Psychological Science, 21*, 1686–1693.
Johnson, D. J., Cheung, F., & Donnellan, M. B. (2014). Does cleanliness influence moral judgments? *Social Psychology, 45*, 209–215.
Johnson, D. J., Wortman, J., Cheung, F., Hein, M., Lucas, R. E., Donnellan, M. B., et al. (2016). The effects of disgust on moral judgments: Testing moderators. *Social Psychological and Personality Science , 7(7),* 640–647.
Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin, 108*, 480–498.
Lakens, D., Adolfi, F. G., Albers, C. J., Anvari, F., Apps, M. A. J., Argamon, S. E., … Zwaan, R. A. (2017). Justify Your Alpha: A Response to "Redefine Statistical Significance". Retrieved from psyarxiv.com/9s3y6
Levy, N. (2011). Resisting 'weakness of the will'. *Philosophy and Phenomenological Research, 82*, 134–155.
Lurquin, J. H., Michaelson, L. E., Barker, J. E., Gustavson, D. E., von Bastian, C. C., et al. (2016). No evidence of the ego-depletion effect across task characteristics and individual differences: A pre-registered study. *PloS One, 11*, e0147770. doi:10.1371/journal.pone.0147770.
Machery, E. (2012a). Delineating the moral domain. *The Baltic International Yearbook of Cognition, Logic and Communication, 7*. doi:10.4148/biyclc.v7i0.1777.
Machery, E. (2012b). Power and negative results. *Philosophy of Science, 79*, 808–820.
Machery, E. (2014). In defense of reverse inference. *The British Journal for the Philosophy of Science, 65*, 251–267.
Machery, E. (in press). Morality: A historical invention. In K. Gray & J. Graham (Eds.), *The atlas of moral psychology*. New York: The Guilford Press.
Macmillan, M. (2002). *An odd kind of fame: Stories of Phineas Gage*. Cambridge, MA: MIT Press.
McShane, B. B., Gal, D., Gelman, A., Robert, C., & Tackett, J. L. (2017). Abandon Statistical Significance. arXiv preprint arXiv:1709.07588.
Meyer, A., Frederick, S., Burnham, T. C., Guevara Pinto, J. D., Boyer, T. W., Ball, L. J., ... & Schuldt, J. P. (2015). Disfluent fonts don't help people solve math problems. *Journal of Experimental Psychology: General, 144*(2), e16.
Nagel, T. (1980). Ethics as an autonomous theoretical subject. In G. S. Stent (Ed.), *Morality as a biological phenomenon: The presuppositions of sociobiological research* (pp. 198–205). Berkeley; Los Angeles: University of California Press.
Nguyen, H. H. D., & Ryan, A. M. (2008). Does stereotype threat affect test performance of minorities and women? A meta-analysis of experimental evidence. *Journal of Applied Psychology, 93*, 1314–1334.
Oakes, M. W. (1986). Statistical inference. Epidemiology Resources.
Peters, D. P., & Ceci, S. J. (1982). Peer-review practices of psychological journals: The fate of published articles, submitted again. *Behavioral and Brain Sciences, 5*(02), 187–195.
Poldrack, R. A. (2006). Can cognitive processes be inferred from neuroimaging data? *Trends in Cognitive Sciences, 10*, 59–63.
Railton, P. (2003). *Facts, values, and norms: Essays toward a morality of consequence*. Cambridge: Cambridge University Press.
Rawls, J. (1971). *A theory of justice*. Cambridge, MA: Harvard University Press.
Robinson, P. H., & Darley, J. M. (1995). *Justice, liability, and blame: Community views and the criminal law*. Boulder, CO: Westview Press.
Ross, L., & Nisbett, R. E. (1991). *The person and the situation: Perspectives of social psychology*. Philadelphia: Temple University Press.
Russell, G. (2010). In defence of Hume's law. In C. Pidgen (Ed.), *Hume, is and ought: New essays* (pp. 151–161). Hampshire: Palgrave MacMillan.
Schnall, S., Benton, J., & Harvey, S. (2008). With a clean conscience cleanliness reduces the severity of moral judgments. *Psychological Science, 19*, 1219–1222.
Sedlmeier, P., & Gigerenzer, G. (1989). Do studies of statistical power have an effect on the power of studies? *Psychological Bulletin, 105*, 309–316.

Seyedsayamdost, H. (2014). *Reproducibility of empirical findings: Experiments in philosophy and beyond*. Doctoral dissertation, The London School of Economics and Political Science (LSE).

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science, 22*, 1359–1366.

Spencer, S. J., Steele, C. M., & Quinn, D. M. (1999). Stereotype threat and women's math performance. *Journal of Experimental Social Psychology, 35*, 4–28.

Stoet, G., & Geary, D. C. (2012). Can stereotype threat explain the gender gap in mathematics performance and achievement? *Review of General Psychology, 16*, 93–102.

Sytsma, J., & Machery, E. (2010). Two conceptions of subjective experience. *Philosophical Studies, 151*, 299–327.

Sytsma, J., & Machery, E. (2012). On the relevance of folk intuitions: A commentary on Talbot. *Consciousness and Cognition, 21*, 654–660.

Talbot, B. (2012). The irrelevance of folk intuitions to the "hard problem" of consciousness. *Consciousness and Cognition, 21*, 644–650.

Vaillant, G. (2012). *The triumphs of experience*. Cambridge, MA: Belknap Press.

Valdesolo, P., & DeSteno, D. (2006). Manipulations of emotional context shape moral judgment. *Psychological Science, 17*, 476–477.

Weisberg, D. S., Keil, F. C., Goodstein, J., Rawson, E., & Gray, J. R. (2008). The seductive allure of neuroscience explanations. *Journal of Cognitive Neuroscience, 20*, 470–477.

Williams, B. A. O. (1985). *Ethics and the limits of philosophy*. Cambridge, MA: Harvard University Press.

Williams, W. M., & Ceci, S. J. (2015). National hiring experiments reveal 2: 1 faculty preference for women on STEM tenure track. *Proceedings of the National Academy of Sciences, 112*(17), 5360–5365.

Zhong, C. B., & Liljenquist, K. (2006). Washing away your sins: Threatened morality and physical cleansing. *Science, 313*, 1451–1452.

Zhong, C. B., Strejcek, B., & Sivanathan, N. (2010). A clean self can render harsh moral judgment. *Journal of Experimental Social Psychology, 46*, 859–862.

Current Perspectives in Moral Psychology

Frans de Waal, Hanno Sauer, Paolo Heywood, Verena Wieser, Edouard Machery, and John M. Doris

Abstract Moral psychology has undergone a renaissance in recent years. Methodological and theoretical advances promise new perspectives on old questions—and as academic disciplines become less siloed, the potential for cross-disciplinary collaboration becomes even greater. In this chapter, we ask leading scholars to offer their views on the future of moral psychology. Biologist and primatologist Frans de Waal, philosopher Hanno Sauer, social anthropologist Paolo Heywood, and marketing scholar Verena Wieser share their thoughts on recent developments and their implications. The chapter ends with a conversation between philosophers Edouard Machery and John M. Doris—two founders of modern moral psychology—about how the field has progressed in the academy.

—Benjamin G. Voyer & Tor Tarantola (Eds.)

The editors thank Jennifer Sheehy-Skeffington for suggesting questions to pose to the contributors.

F. de Waal (✉)
Living Links Center, Emory University, Atlanta, GA, USA
e-mail: dewaal@emory.edu

H. Sauer
Department of Philosophy and Religious Studies, Utrecht University, Utrecht, The Netherlands
e-mail: h.c.sauer@uu.nl

P. Heywood
Division of Social Anthropology, University of Cambridge, Cambridge, UK
e-mail: pph22@cam.ac.uk

V. Wieser
Department of Strategic Management, Marketing and Tourism,
University of Innsbruck School of Management, Innsbruck, Austria
e-mail: verena.wieser@uibk.ac.at

E. Machery
Center for Philosophy of Science, University of Pittsburgh, Pittsburgh, PA, USA
e-mail: machery@pitt.edu

J.M. Doris
Philosophy-Neuroscience-Psychology Program & Philosophy Department, Washington
University in St. Louis, St. Louis, MO, USA
e-mail: jdoris@artsci.wustl.edu

B.G. Voyer, T. Tarantola (eds.), *Moral Psychology*,
DOI 10.1007/978-3-319-61849-4_8

Frans de Waal

With recent developments in moral psychology and experimental philosophy, there's no going back for the way philosophy is done. Would you agree?

Since the Enlightenment, philosophy has taken over the top-down role in moral thinking from religion. Instead of religious dogma or scripture telling us how to behave, the philosophers provided us with principles, logic and reasoning underlying our moral decision-making. Rather than working with human psychology, or, as I would say, primate behaviour, many philosophers declared natural behavioural tendencies as largely irrelevant. It was all about the "ought," and not about the "is." Philosophers would come up with principles, such as utilitarianism, that deny the fundamental loyalties that mark every mammal. Every mammal values its own kin and offspring above everyone else, but utilitarianism asks us to value all human life equally and go by the numbers (the more happiness the better), which is not how mammalian psychology has been designed. I would love to see a moral philosophy that is more in tune with human tendencies and recognizes that these tendencies have an age-old history. I know very well the "naturalistic fallacy" argument, but think it is grossly overrated: driving a wedge between morality and biology has given us a view that is out of touch with human nature. Even David Hume (1739/1985: 335)—to which the naturalistic fallacy arguers often refer—recognized this, as he never said we should ignore human biology (in fact, he invoked it very much himself when he spoke of human sympathy) but only added that "a reason should be given" for how we argue from the facts of life to the values we strive for. Asking us to give a reason is not the same as saying it cannot be done.

The idea that moral principles can be born from very basic natural tendencies was brought home to me in the most forceful manner when we found signs of a sense of fairness in other primates. Not only do monkeys (and also dogs and corvids) protest against receiving less than a partner for the same task, chimpanzees show, just as humans, a tendency to equalize outcomes even if doing so is not to their immediate advantage. Although we believe that in the long run this equalizing tendency is advantageous (Brosnan and de Waal 2014), the fact is that the sense of fairness of chimpanzees is hard to distinguish from that in humans. This means that fairness, instead of a moral principle arrived at by means of reasoning or societal ideals, is an old tendency with evolutionary advantages. It obviously requires cognition (the parties need to be able to learn the advantages of equalized outcomes), but then, the cognition of chimpanzees and humans is more similar than different. It is reflected in how they solve the dilemma between wanting as many rewards as possible and wanting profitable cooperation. Philosophers need to start rethinking their field in the context of not only human psychology but also our species' evolutionary background.

Can moral psychology help answer moral or ethical questions?

I cannot answer this question for psychology, but for biology I think it is rather simple. Biology does not dictate any specific moral rules. These rules vary by human

culture and vary across time within a given culture, so cannot be given by biology. But biology has given us the basic capacities we need to build moral systems. First of all we are interested in others and in working with them, which is a prerequisite for morality. Then there are the capacities for empathy, the following of social rules, sense of fairness, tit-for-tat cooperation, social attachments and commitments and so on, all of which enter the moral equation and are older than our species.

Human morality is like language. We are all born with the capacity to develop it, using the moral building blocks and sentiments recognizable in the work of Edward Westermarck (1908) and David Hume (1739/1985), but how precisely we fill in the capacity is up to our environment and culture.

How important is it to be multidisciplinary when doing research in moral psychology? What are the main difficulties in achieving this?

The field of moral psychology could benefit from more exposure to studies on animal behaviour. After all, in the study of social animals, we are very used to social organization constrained by rules and regulations. The social hierarchy of the primates is one big system of regulation, which requires emotional control and inhibitions. Even if these rules and regulations are not justified by what we would call moral principles, the fact that animals cannot express themselves in unlimited ways, but face all sorts of social constraints, is obviously very similar to a moral system. We, humans, speak of "right" and "wrong," whereas in many animals life turns around what is "acceptable" and what is "unacceptable" behaviour. Punishment for the latter behaviour makes animals refrain from it. Here is a description from my book *The Bonobo and the Atheist* (de Waal 2013: 149), which treats these connections at length:

> At Tama Zoo, in Tokyo, I witnessed a surprising ritual. From the rooftop of a building, a caretaker spread handfuls of macadamia nuts among 15 chimpanzees in an outdoor area. The chimps rushed about collecting as many macadamias as they could in their mouths, hands, and feet. Then they sat down at separate locations in the enclosure, each with a neat little pile of nuts, all oriented toward a single place known as the "cracking station." One chimp walked up to the station, which consisted of a big rock and a smaller metal block attached to it with a chain. She then placed a nut on the rock's surface, lifted the metal block, and hammered until the nut gave up its kernel. This female worked with a juvenile by her side, whom she allowed to profit from her efforts. Having finished her pile, she then made room for the next chimp, who placed her nuts at her feet and started the same procedure. This was a daily ritual that always unfolded in the same orderly fashion until all nuts had been cracked. I was struck by the scene's peacefulness, but not fooled by it. When we see a disciplined society, there is often a social hierarchy behind it. This hierarchy, which determines who can eat or mate first, is ultimately rooted in violence. If one of the lower-ranking females and her offspring had tried to claim the cracking station before their turn, things would have gotten ugly. It is not just that these apes knew their place; they knew what to expect in case of a breach of rule. A social hierarchy is a giant system of inhibitions, which is no doubt what paved the way for human morality, which is also such a system. Impulse control is key.

There is very interesting work on emotional control, such as the marshmallow test conducted on apes and parrots, and these animals being as good at controlling their impulses as human children. These findings are not surprising for students of animal behaviour, but the general public, of course, still sees animals as wild and uncontrolled.

On the positive side there is all the work on empathy and genuine altruism in animals (de Waal 2008), including nowadays neuroscience studies on empathy in rodents (Burkett et al. 2016), which indicate that caring for others, even if there is nothing necessarily to be gained by the altruist, can be found in other species. By taking all of these tendencies into account, moral psychology can ground itself in evolutionary biology, which—I would say—is the only sensible grounding for any field that concerns itself with human behaviour.

Hanno Sauer

With recent developments in moral psychology and experimental philosophy, there's no going back for the way philosophy is done. Would you agree?

Yes. Empirical evidence shows that our powers of introspection are frail and prone to self-deception. We simply don't know where our conceptual intuitions come from and what influences them. Naïve conceptual analysis is dead.

Can moral psychology help answer moral or ethical questions?

Yes. It cannot answer moral questions on its own; but neither can empirically empty allegedly "pure" moral theorizing. More specifically, empirical information can be brought to bear on issues of normative import by (i) debunking the empirical presuppositions regarding moral agency that various normative theories incur, by (ii) debunking people's moral intuitions as epistemically defective, and by (iii) reflexively enabling people to improve their moral judgements and actions in light of (i) and (ii).

What role should moral psychology and neuroscience play in shaping law and public policy?

Given the actual extent to which law- and policy-makers seem to pay attention to evidence and reason, they should at the very least play a much *larger* role. It could also be tremendously useful in identifying and counteracting the various epistemic limitations of jury members, in reassessing the conditions for criminal responsibility, and in gauging the long-term effect of criminal "justice." Properly taking into account empirical evidence in general, not just psychological and neuroscientific evidence, but also social scientific and economic insights, would likely lead to drastic reforms of the current penal system.

How important is it to be multidisciplinary when doing research in moral psychology? What are the main difficulties in achieving this?

All-important—it simply cannot be done unilaterally. The main difficulty, it seems to me, is to reap the benefits of the epistemic division of labour while avoiding the costs that come with it. People come from different backgrounds and have different abilities. It is extremely tricky to coordinate people's work in the absence of central oversight (which would likely be undesirable anyway).

Paolo Heywood

With recent developments in moral psychology and experimental philosophy, there's no going back for the way philosophy is done. Would you agree?

Whilst I think there's a lot in the way of insight to be gained from experimental philosophy, particularly when it comes to cultural diversity, I also think—and I am of course bound to say this as an anthropologist—that quantitative methods in the social sciences have their limits. Responses to survey questions about abstract cases can tell you plenty of things, but they cannot tell you the same things that observing the ways in which people deal with moral and ethical concerns in their everyday lives can. Which of those one is more interested in obviously depends on one's aims. And, for what it's worth coming from a layman, I see no particular reason why philosophers should abandon conceptual work in favour of methods already employed by sociologists and psychologists, unless we have come to think the kinds of results produced by the latter are in some way or another superior, more cost-effective or more "impactful" than the former. And if that's the case, then it's worth asking why. Philosophically, might I add.

Can moral psychology help answer moral or ethical questions?

Of course. Again, though, I would highlight the word "help" in that question. Moral psychology, neuroscience, philosophy and anthropology can all "help" answer moral or ethical questions because they provide answers of different forms to such questions, not because any one of them has hit upon the correct form answers should take.

What role should moral psychology and neuroscience play in shaping law and public policy?

It's a bit difficult to have much faith in the value people will continue to place on "experts" in the wake of recent political events. And since I am neither a moral psychologist nor a neuroscientist, it's not really for me to attempt to specify their place in public life. That said, as I have already indicated, I am rather wary of the ways in which academic disciplines are increasingly required and effectively extorted into having "impact." The more that academic disciplines are put in hock to whatever people happen to think is "useful" at any particular moment, the more vulnerable they are to rapid changes in assessments of utility—as we have recently seen—and the less they are capable of doing what they are best at: questioning our assumptions (regarding, e.g., what it means to be "useful").

How important is it to be multidisciplinary when doing research in moral psychology? What are the main difficulties in achieving this?

As I've suggested, I think interdisciplinarity is at its best when it is complementary, rather than integrative, and actually that a significant difficulty lies in ensuring that "being multidisciplinary" doesn't end up meaning taking one totalizing approach that also happens to draw from a range of disciplines. I personally think it would be more productive if we all kept on arguing with one another over approaches, rather than stifling such debate in an attempt to find an ideal approach that doesn't in fact exist.

Verena Wieser

What are the philosophical developments that shape our understanding of morality in marketing research and practice?

Morality always has been—and still is—a contested concept in marketing and consumer research and practice. The discipline features lively debates concerning what 'doing good' or 'doing bad' means in marketing contexts and how those meanings develop in contemporary consumer societies (e.g. Caruana 2007a, b; Stoeckl and Luedicke 2015). I would like to share one or two observations on these debates here.

The vast amount of morality research in marketing follows a *techno-rationalist marketing discourse* (Caruana 2007b), which views morality as one discriminating factor in consumption choices. From a micro-marketing perspective, morality competes with pragmatic factors such as price and quality when consumers decide, for instance, between conventional and fair-trade products in their daily routines. A rich pool of research traces how, when and why consumers couple their purchase decision with—or decouple their purchase decision from—societal moral norms and personal concerns (see Grayson 2014 for a summary of articles on this issue in the Journal of Consumer Research).

The discipline's focus on consumer choices, however, leaves blind spots in the moral marketing landscape. Whilst consumer researchers consistently spot a gap between consumers' moral attitudes and actual purchase behaviours, business scandals and brand crises unveil the substantial limitations of the logics of efficiency and corporate self-control. The overestimation of the "empowered" and "responsible" consumer (Caruana and Crane 2008; Giesler and Veresiu 2014; Izberk-Bilgin 2010) has called reformist perspectives on the marketing agenda which endorse the conversion of protected moral values, such as the respect for human life or for ecological balance, into golden rules of marketing conduct. However, the modernist endeavour of reducing moral ambiguity in consumers' lives increasingly fails in its attempts to translate the abstractness of unifying ideals into concrete marketing measures. Besides other barriers, a lack of global governance systems makes it both difficult to agree on universal moral duties and to monitor compliance, respectively.

An emerging *moral pluralist discourse* (Eabrasu 2012) promotes a view that corporations accept and promote more than one morally acceptable set of commitments in the postmodern world. Supported by the responsiveness of digital media, marketers build the moral identity of their brands in a sociocultural flux. On one hand, brands compete on claims to be more sustainable, more ethical or, at least, less evil relative to other market participants. On the other hand, normative branding projects attract cynical comments that label moral marketing campaigns as "pseudo-moral," "greenwashing," or "blue-washing." Research will show whether an inflation of moral messages in marketing activities leads to a loss of moral sensitivity in the marketplace (Bauman and Donskis 2013) or to a more nuanced and attentive public opinion on moral concerns.

What are the current hot topics and directions in consumer and marketing research concerning morality?

Morality research in marketing monitors closely how brands dynamically navigate the blurry frontiers between good and bad. Extreme cases—when brands break taboos (e.g. in shock advertisement campaigns) or exceed the limits of legal tolerance (e.g. in corruption scandals)—show how marketers, consumers, regulators, the media and other brand stakeholders deal with morally ambivalent marketing activities.

Marketing research on morality will further set the focus on consumers' moral reflexivity and self-awareness and other/market awareness. Study programs increasingly trace the moral footprints of consumers online (e.g. through capturing the moral tone of consumer feedback in social media environments), compare how consumers define morality in various consumption contexts (e.g. in mundane spheres like food consumption versus in extraordinary experiences like holiday consumption) and investigate how consumers develop their moral competences over time and vis-à-vis contextual premises (e.g. socio-economic developments, cultural trends, social group/family traditions).

Going beyond consumers' purchase decisions, cultural and historical marketing research reveals how moral values form and evolve in consumer subgroups (e.g. in neighbourhoods or online brand communities) and how consumers use morality in combination with consumption goods and experiences to enact identity work (e.g. in moral conflicts between fans and enemies of luxury brands). Finally, interdisciplinary research pushes methodological boundaries and investigates how consumers experience morality with their bodies and through moral sentiments (e.g. through anger, anxiety, disgust or guilt).

How should we understand the nexus between morality and regulation?

The question of how morality translates into regulation is also a question of authority. One facet determining authority is moral language; legal authority dominantly rests on negative judgements of "what is wrong," "what is unjust," "criticizable" or "impermissible," on ensuing obligations, interdictions and penalties. However, at the other side of the morality coin, positive moral judgements simultaneously build moral authority, like notions of "praise" of "what is good," "obligatory" or "heroic" (Bartels et al. 2015). To mention just one of many areas of interest, morality research will need to pay more focused attention to the cultural and regulatory qualities of both positive and negative moral language in consumption and marketing contexts and beyond.

A Discussion Between Edouard Machery and John M. Doris

EM: John, good to talk to you. So today we're going to be talking about empirically oriented moral psychology and its growth in philosophy and psychology over the last 10 or 15 years. I think it would be useful to start with the obvious question—

what were things like in the beginning, 10 or 15 years ago? What do you think moral philosophy and moral psychology were like about that time?

JMD: You and I were in interestingly different circumstances, because I was an ethics graduate student who got interested in cognitive science, and many of my colleagues, like you and Shaun Nichols, were people working in cognitive science who got interested in ethics.

EM: That's right.

JMD: There was a lot of resistance, but our experiences of that resistance might have been pretty different. At my end, resistance was often just benign neglect; people didn't think to do empirical work or empirically informed theorizing. When some of us proposed doing it in ethics, the response was usually based on concerns about normativity—that, you know, you couldn't import empirical facts into moral philosophy without distorting ethics' distinctively normative character.

From your end things might look a little bit different. There was the thought in psychology—I think there still is in the mainstream psychology journals—that science doesn't deal in evaluative discourse. So, to caricature just a bit, philosophers thought values were good, facts bad; psychologists that facts were good, values bad. From both directions we were doing something that went counter to the dominant ideology.

EM: I agree entirely. There was also this sense when I was finishing my PhD in the early 2000s, that the real psychology was not social psychology and, more generally, not the psychology of "real-life behaviors": what we eat, how we love, what we do in everyday life, etc. The real psychology—the one we philosophers of psychology should be excited about—was cognitive psychology. So for a philosopher of cognitive science in the late 1990s and early 2000s—for graduate students like me at the time—it was really not obvious why philosophers of cognitive science should worry about morality.

It was not even clear there was a good psychology of moral judgement. I think things have changed tremendously in 15 years. Now more psychologists are interested in morality, but at the time there was very little interest in it from cognitive scientists.

Did you have that impression too? That social psychology and psychology related to "real-life" behaviors were not well respected in the philosophy of psychology and perhaps even in psychology until maybe 10 years ago?

JMD: Maybe something like that. Anyway, "serious" philosophy of cognitive science was focusing on issues I call architectural.

EM: That's right.

JMD: Architectural questions, and the empirical work that was relevant to this was on very low-level cognition. So philosophers of cognitive science weren't interested in psychology treating what philosophers like me would think of as questions of broad human interest.

EM: That's the way I felt. So how did you get interested in empirical moral psychology—what we think of as real moral psychology? Why did you as a graduate student at the time get into it?

JMD: Actually for philosophical reasons! Two of my heroes—then and now—were Bernard Williams and Alasdair MacIntyre. (Bernard Williams is deceased of course, and much missed in philosophy.) And I took them to be saying that if philosophical ethics is going to get better, it's going to need a more lifelike moral psychology. This is a point, of course, made before by Anscombe—though on my view she did little to contribute to the cause. For Williams and MacIntyre, "lifelike" meant thinking more about character. My thought, a thought they probably thought flat-footed, was, "well that means we should go talk to psychologists!" But both of them were very supportive when I talked to them about it.

Originally—and this is kind of funny—I was interested in the thematic apperception test and motivational psychology, which of course is a species of personality psychology. Then I happened to date a personality psychologist who was Mischelian and she said, "you really need to think about his critique of traits if you're thinking about moral psychology."

And so one day I was in the library, back when people went to the library (and this is actually why maybe it would be good if people still went to the library) and I saw this book called *The Person and the Situation*. This kind of seemed relevant to what I was thinking about. So I opened it up and realized one author was at Michigan. So, I got my adviser Allan Gibbard to arrange an introduction, and I went to meet Dick Nisbett.

I must have been a sight: I used to have long hair, and so this shaggy giant came into Dick's office and said, "have you ever wondered about how all this stuff that you do relates to morality?" And he said, "I've been waiting for years for someone to knock on my door and ask me that."

And then after that I was off to the races. You know: you have an idea that you can do something—empirical moral psychology—but you need to have a good example for traction. And I think the traction was that character theorists and virtue ethicists very much took themselves to be in pursuit of a lifelike moral psychology. So as it were, they invited me in, which gave me and others like Gil Harman license to dirty the carpet with those messy facts.

EM: Let me just follow up on that. How did people around you react? You meet Nisbett, and his research is obviously relevant for your interest in moral psychology. Clearly you're right on target, but how do the Michigan folks—you were a graduate at Michigan at the time—how did they react?

JMD: Well, it helped that I got Nisbett on my committee. He was—is—a very big deal there; he already had a University Professorship.

So having Dick's stamp of approval helped a lot. But I do remember one of my teachers saying about my character skepticism, "I don't know what you could say to convince me of this."

Michigan of course had excellent moral philosophers of all stripes, you know, conspicuously Darwall, Gibbard, and Railton, and all of them were sympathetic to naturalism.

And of course you might see the kind of work we started to do as enabled by the kind of theoretical groundclearing that people like Peter Railton and my under-

graduate teacher, Nicolas Sturgeon, did when they showed that there's a kind of ethical naturalism where ethics doesn't need to fear science. So I think Michigan people were pretty supportive.

Of course it's always a little hard to sort out the philosophy from the sociology since for much of graduate school I spent a lot more time doing martial arts than philosophy. So I certainly had more than a few moments of impatience from my professors, but Michigan was probably one of the best places to do moral psychology. None of the faculty then did quite what we do now, but they were pretty sympathetic, and of course my adviser Allan Gibbard is just an incredibly intellectually curious guy—he wanted to see arguments but he was very supportive. I don't think there are many other places where I could've made that fly because obviously I was a very beginning philosopher and at the time I was not going to have the best possible arguments. One doesn't imagine that I would've been able to do what I did at many other major graduate programs.

EM: The other places that became important for moral psychology were Rutgers around Stephen Stich and Princeton around Gil Harman. I don't know exactly when Stich and Harman got interested in moral psychology—they taught a graduate seminar together I believe.

JMD: With John Darley in 2000. That's where many of us met.

EM: That's right, yes so it was 2000.

JMD: Gil had been thinking about that for a few years because he'd been working on the fundamental attribution error, and Steve had a paper in 1993 about mental representation in ethics. But I don't think it was clear to either of them that it was going to be, as we say nowadays, a thing.

EM: It's noteworthy that Stich didn't develop his interest in moral psychology immediately after that 1993 paper. It's a very good paper and an influential piece of work, but it did not lead to an explosion of work in moral psychology either by him or by his students and colleagues.

Personally I got into moral psychology through Stich because I was at Rutgers in the early 2000s, when Steve actually was starting to take moral psychology extremely seriously and to do important research in this area.

I was influenced by the work that had already been published at the time, including yours. Psychologists were getting involved. I read John Mikhail's dissertation when I was still a graduate student.

Of course evolutionary psychology also got me interested in the psychology of "real-life" human behavior, including the psychology of morality. Evolutionary psychologists were doing work that was at the intersection of cognitive psychology and social psychology. And that led me to pay more attention to social psychology and, as a result, moral psychology.

Do you have a sense of when psychologists themselves got involved? Was Marc Hauser an early adopter?

JMD: It's kind of interesting how to think about this. From the 1960s to early 1980s, we have what we can call that the golden age of social psychology.

There were all of those studies on helping and prosocial behavior and the figure who looms so large then is the great John Darley. Then John turned to other stuff.

Of course Stich is a lifelong friend of Nisbett, so he was kind of in the picture for many of us.

If you think of the first meeting of the Moral Psychology Research Group that Stich organized in 2003, there were very few scientists. I think both Joshua Knobe and Josh Greene were there, but they were originally philosophically trained. Fiery Cushman and Liane Young, both Harvard graduate students in psychology, joined the group later.

Think of the scientists that visited the Moral Psychology Research Group in our early days. Marc Hauser came, but many of the visiting scientists weren't working directly on morality per se: Paul Rozin, George Lowenstein, Marty Seligman. A lot of us were really influenced by psychologists, but it didn't seem like these psychologists or their students were really quite our fellow travellers or colleagues. Maybe that came a little bit later.

Maybe an exception here would have been Jonathan Baron; he cared a lot about morality and moral philosophy. And of course then we have Jon Haidt, who might have been one of the first moral psychologists.

EM: Indeed. I remember when I was at Rutgers as a visiting graduate student in the early 2000s, we read a lot of Jon Haidt's work and he was being discussed by graduate students around Steve Stich.

He was clearly very influential at the time in leading us to think that the psychology of moral judgement was relevant to philosophical questions and vice versa.

JMD: Of course Jon had a very talented graduate student, Jesse Graham, who's now one of our colleagues in the Moral Psychology Research Group. But interestingly, it might be that Jon was more influential amongst philosophers than psychologists.

EM: I wouldn't be surprised. It may be worth saying a few things about the Moral Psychology Research Group (MPRG), which we've mentioned a few times. There may still be a few people out there who don't know enough about that group, so it's time to enlighten them. When was the group created exactly? 2003 was the first meeting, is that right?

JMD: Yeah, 2003 as far as I can remember. I must've been working in Santa Cruz. Steve organized the meeting and I remember it was a sticky New Jersey grey day and I had trouble finding the venue and then there couldn't have been more than 8 or 10 people. I'm quite sure Walter Sinnott-Armstrong, Gil Harman, Jesse Prinz, and Shaun Nichols were there.

EM: I was not there. I may have been back in France at the time or I may have been in Germany. Josh Knobe must've been there, and perhaps Chandra Sripada and Dan Kelly.

JMD: And as I say, if scientists were represented, it was Princeton trained philosophers. That's interesting though; Princeton is not a very empirically oriented program, and two of their very best known recent products are very empirical.

Another crucial moment for MPRG was the really big conference on "The Psychology and Biology of Morality" Walter Sinnott-Armstrong put together at Dartmouth in 2004. A lot of the early MPRG types were there, together with other

good philosophers and many scientists studying morality but not necessarily yet collaborating with philosophers, like Kent Kiehl.

So it was kind of a coming out party where MPRG started to connect with a wider community. A big moment at that conference was Josh Greene presenting his early work—I don't think we've mentioned experimental philosophy so far, but this then new movement attracted huge attention, and a lot of the "X-phi" work was on morality.

There was an MPRG held right after the big conference. Joshua Knobe, another founding X-phi-er, was there, and he and I presented something on responsibility. Walter sent me the program from that MPRG not long ago, and the business meeting was titled something like, "Drinks & Planning Session: Where Do We Go From Here?" [laughs]

We've come a long way—there's now something like five of the Sinnott-Armstrong Moral Psychology volumes. Walter's been a force all along, first because he was respected as a philosopher's philosopher who knew his way around the arguments, which brought credibility to empirical approaches, but also because of his institution building skills.

X-phi also cross-polinated back to psychology, as it was influential for younger psychologists studying morality, like Fiery Cushman and Liane Young, who have gone on to do important work. X-phi seems pretty well-established now, too, with the *Oxford Studies in Experimental Philosophy*, edited by Joshua Knobe, Shaun Nichols, and a psychologist, Tania Lombrozo, slated to appear regularly. Fingers crossed!

EM: A watershed moment for X-phi was the preconference before the 2008 Society for Philosophy and Psychology annual meeting in Philadelphia. It brought together all the philosophers and psychologists pushing forward what was, and still is, one of the most exciting developments in philosophy: Eddy Nahmias, Bertram Malle, John Mikhail, Jonathan Baron, Liane Young, Eric Schwitzgebel, Brian Scholl, Ron Mallon, Tania Lombrozo, Shaun Nichols, Josh Knobe, Ernest Sosa, Jonathan Weinberg, and myself. It's remarkable that half of them are MPRG members!

It's also really worth highlighting how important the MPRG was in creating a community of likeminded philosophers and then psychologists, people who had similar views about how to develop a moral psychology that was relevant for philosophy. Instead of each of us working in our little niche alienated from both philosophy and psychology, somehow it felt that we could be a force. And we were a force! MPRG was actually extremely important in changing the sociology of philosophy, and I hope to an extent anyway, of psychology.

JMD: Certainly to some extent. You know, now there are all these scientists who characterize themselves as moral psychologists, and who I have never even heard of. That's how much the field has grown. And that's what the MPRG did: We edited the *Moral Psychology Handbook*, which was a nice touchstone, but more importantly the group generated hundreds of collaborative publications. And many group members pollinated across disciplines and continue to do so.

Of course the fact that Stich, with all his influence and energy, was some sort of protector for the group pushed us forward in the early days. In the early days, it was absolutely critical having people like Stich and Harman to give the group credibility. Then we got lucky with some publications that people wanted to talk about and as it spread, we've been able to attract young people.

I take it a big reason for the success of moral psychology is that it's just kind of fun. I mean everybody is different. Some people are worried about external world skepticism. Some people are worried about whether dishrags persist through time. And that's fine! It's a great thing about philosophy that there are a lot of different questions, but a lot of people thought the questions in moral psychology were really cool. Pick your favorite example and it's just fun to read that stuff and try and figure it out.

EM: I agree. If someone asked me why moral psychology was so successful in philosophy and in psychology, I would mention some of those things you've mentioned. The fact that Steve Stich and Gil Harman were already extremely influential in philosophy gave us some credibility, as you said.

Also moral psychology is fun, no question about that.

And we were lucky in attracting some of the best and brightest in both philosophy and psychology at the time, and the type of research we were doing was just extremely good. People could see it was good and interesting.

Something you haven't mentioned is the spirit of what was going on. I mean the atmosphere of what was going on between us was quite different from the usual atmosphere in philosophy. It was very friendly, we were collaborating with one another. It was always constructive. We were trying to help each other. This spirit has now become slightly more common; it's more common now to hear that philosophers should be less critical of one another, less combative. But it was not like that 10 years ago.

In any case, very early on we had this idea that we wanted to help each other even when we were criticising each other. And that was actually a very useful way of creating a research community that ended up being quite successful.

JMD: There was a real feeling of, you know, group connectedness; people were friends; people generally delighted in one another's success.

EM: Yep.

JMD: Now there is more of a breadth in both the group and the field. Valerie Tiberius is the person who first comes to mind, but involving people with more mainstream interest in normative ethics or ethical theory made possible a supportive environment for people from a broad spectrum of methodological orientations to have, you know, to have some fun. And be supported. So yes, I think the MPRG and empirical moral psychology have been a big success.

But you know, as we think about what we want to do as a group and individually going forward, we've been sort of having this suspicion that maybe we haven't figured out what our next big thing is and what would excite us. So it's not unreasonable to ask: how successful has it been really?

I guess this is kind of a mid-life crisis.

EM: I know. [laughs] Well it's…it's not entirely clear which metrics we should use to decide how successful we've been. Clearly many of us have been successful from an academic and professional point of view and moral psychology was part of our success. It did contribute to our academic success, to getting read, to putting some of our ideas out, and getting discussion going around our work. So in that respect we've been successful.

JMD: Citation, dissertations about the work, right?

EM: That's exactly right, by all these measures we've built a successful community and led a successful project. In other respects it's less clear how successful we've been. Have we really changed ethics and philosophy? If you open some of the main journals in philosophy you may feel that you're stuck in the 1960s. I'm of course exaggerating a bit, but you know there isn't that much work of the kind we've been pushing that gets published in the top two journals in ethics and the best generalist journals in philosophy like *Noûs* and *Philosophical Review*. There is the occasional paper, but I think many philosophers still do non-empirical moral psychology. So that's a benchmark which is a little bit more depressing than the first benchmark.

JMD: It's correct to say there are empirically oriented moral psychologists who have had enviable careers. But, I take it the two highest visibility journals in moral philosophy are *Ethics* and *Philosophy and Public Affairs*.

These are journals that I wouldn't really think of submitting an empirically oriented piece to. There have been a few exceptions but they are few and far between. We do sometimes get things in *Noûs* and *Philosophy and Phenomenological Research*; they've been generally sympathetic, because Ernest Sosa edits them and is genuinely philosophically open-minded and has a good eye.

EM: They have been. That's right.

JMD: And they are amongst the best mainstream journals. Obviously *Philosophical Psychology* and *Mind and Language* are sympathetic journals, but they are less mainstream. On the other hand, Peter Momtchiloff is at Oxford University Press, and he has been supportive of good quality interdisciplinary work, so we do get monographs at the best house for philosophy.

On the psychology side, I get the same sense on the journals, right?

In social and personality psychology the *Journal of Personality and Social Psychology* is the flagship, and they don't do a ton of moral psychology either. So the journal benchmark may be not so good. What's your take on the sort of departmental composition benchmark? What are the big graduate programmes doing in both disciplines?

EM: Let me add something about journals: the only exception in psychology would be *Cognition*, which has become extremely friendly to moral psychology. But of course it's not a big journal in social psychology. It's a very good journal, well respected, but it's not the central journal in social psychology.

JMD: It's not *Psychological Science* and they've always had kind of a theoretical orientation.

EM: This may say something interesting about the MPRG as a research group. We are a community of researchers, by some measures a very successful one, but

we've not tried to control institutions from the inside. We did not try to control leading journals, such as *Philosophical Studies* or *Noûs* or whatever, where we could publish our things. Nor did we try to control some academic institutions in philosophy: We never had a plan to control the APA or to be very much involved in the planning of the APA conferences, such that our work could be well represented. Still, we were successful. It's worth noting because not every interest group in philosophy has behaved like that, you know.

Now about departments, it's a good question. It's a bit of a mixed bag as well, you know. Many of us are in good departments. I work at Pitt.

JMD: Not accidently in the History and Philosophy of Science department.

EM: True enough. You're in a top department for the philosophy of cognitive science, with the Philosophy-Neuroscience-Psychology program at Washington University in St. Louis, Shaun Nichols at Arizona, Jesse Prinz at CUNY, Steve Stich at Rutgers and Gil Harman at Princeton

JMD: Although Steve and Gil did not ride into town on moral psychology.

EM: That's exactly the point.

JMD: Moral psychology rode into town on them.

EM: That's exactly right. So we are blessed, but again we are in a sense the exception, right, that confirms the rule. We're sort of outliers. We did well but most of the top departments don't really do empirically informed moral psychology, I would say.

JMD: Here's one way to think about it: who besides Gil Harman is at an Ivy League grad program? (Adina Roskies is a leading moral psychologist at Dartmouth.)

EM: Yeah.

…and who at the University of California? I guess San Diego would be the exception there. David Brink and Dana Nelkin think about science seriously and Manuel Vargas, one of our friends at the MPRG, has just moved there, so maybe San Diego is an exception, but certainly not UCLA or Berkeley.

JMD: Not so good on that kind of measure. Happy enough to note that we're not missing meals, but it does not yet seem that graduate programs feel like they have to have one or two moral psychology types.

In contrast, at many places it's acknowledged that there would be something wrong if they didn't have one or two specialists in ancient philosophy.

EM: I agree.

JMD: This gets us to the question of what's going to happen in 10 years: What's the future looking like for our ilk?

EM: Yup, it's a good question. Moral psychology is booming in psychology. The number of papers that get to be published has increased dramatically over the last 10 years.

Moral psychology may even have reached a ceiling in psychology and in neuroscience. It's not clear to me how much bigger the field of moral psychology in psychology can become. Now in philosophy I'm not utterly optimistic.

Maybe I'm reaching a point in my life where I see things in darker shades than I used to. I do feel philosophers are really hard to move and I also feel that it goes through cycles of interest, and that after 10 or 15 years interests fade and philoso-

phers move to other things. And I do already feel that there's a bit of that going on in philosophy at this point: There was a lot of interest in empirically informed moral psychology—including experimental philosophy.

"This window is closing" is a bit too strong, but perhaps it is starting to be less open. Do you have a similar pessimistic look or am I just…is it my bad night that's speaking?

JMD: Well I don't know, that's a good question, whether it's just being up with your child…

I think that there was a kind of optimism in the old days that was sort of—"we the happy few who are about to die." A real sense of mission, and we're all doing it as close friends and any victory was a big deal. But now the Moral Psychology Research Group is much more diverse and we have people doing very different kinds of work. So I think things feel more diffuse. I'm not sure that's worse.

EM: I agree.

JMD: One way people make things less exciting is by succeeding…

EM: That's true.

JDM: In any case, it's certainly not guaranteed that the gains that we've been celebrating in this conversation are here to stay. What's a thing that people don't talk about anymore that was a big deal, that everybody had to have a view on? In philosophy or psychology, a thing that fizzled?

EM: Modularity would be one of them; people are much less interested in modularity than they were 10 years ago. Ten years ago everybody had to have a view about whether the mind was modular or not, and dozens, hundreds of papers were written by psychologists and philosophers on that topic. I haven't seen very much on that topic lately, and it's not a topic I would really recommend for a graduate student.

JMD: A good case. So should you think that the moment has passed or should you think that the general idea that the mind has a lot of bits and pieces that are often doing their own thing, the most generic way of describing modularity, is now part of the water?

So one way we could think about the future of moral psychology is, jeez, it doesn't quite seem like that there's a bunch of angry young men and women gravitating towards moral psychology the way we were, and it's hard to think of people who are going at it in quite the same way.

But another way to think about it is, everybody talks about interdisciplinarity. So there are all these virtue ethicists writing books that claim to be developing empirically adequate theories, and among philosophers working on emotion, like my graduate student colleagues Justin D'Arms and Dan Jacobson, it's utterly expected that you're going to have some facility with the psychology of emotion.

EM: I'm not sure which of these two descriptions is the right one, and I don't exactly know whether topics like modularity have disappeared or whether they've become part of the air we breathe.

JMD: I vote for the air we breathe then! But it's funny, it makes it kind of harder. In my case, although I've lately been working on character again, that was never really what I was about. I was about figuring out how you could do moral psychology and take empirical work seriously, but still take ethics seriously. And now it's

clear to me that's going on all over, methodologically. You know, the dog has talked, now what should he say?

EM: To switch topic slightly, do you expect some kind of backlash from more traditional philosophers, from moral philosophers?

JMD: I think there has been backlash the whole time…

In the bad old days when we did convention interviews for the job market, I had several interviews with people lecturing me about how wrong-headed everything I was doing was.

There is one thing that I do think is hopeful, especially in light of the current troubles in psychology. Some people think it's a crisis, some people don't think it's a crisis. I don't think it's a crisis; we know what to do to do psychology better. But the perception of a crisis might make philosophers more suspicious of consuming psychology.

The flip side of that is that we're so much more sophisticated about consuming psychology than we used to be. There are some people like you who have research interests in statistics and can do their own experiments. But even somebody like me, who's still very much a philosophical theorist, I routinely collaborate with psychologists and so I pick some of the relevant knowledge up, and I just think we're so much better at it.

My students at WashU take statistics, and there are all these avenues of research that are open to them, that are not open to me.

EM: I agree: philosophers have improved dramatically in their use of science more broadly, and psychology in particular. People have become less naïve, more sophisticated, better at distinguishing bad from good science. There has been progress in this respect, and graduate students are, I have to say, much more sophisticated in this respect than I was when I was a graduate student.

In this respect I'm optimistic about philosophy because the graduate students we train are very good. They are usually very good philosophers and they are savvy from a scientific point of view, from a psychological point of view, they understand very well how psychology gets done, much more than I used to when I was a graduate student.

JMD: Indeed. Is that a good note to stop on?

EM: I think it is. Good talking to you, John.

References

Bartels, D. M., Bauman, C. W., Cushman, F. A., Pizarro, D. A., & Peter McGraw, A. (2015). Moral judgment and decision making. In G. Keren & G. Wu (Eds.), *The Wiley Blackwell handbook of judgment and decision making*. Chichester, UK: Wiley.

Bauman, Z., & Donskis, L. (2013). *Moral blindness: The loss of sensitivity in liquid modernity*. Cambridge, UK: Polity Press.

Brosnan, S. F., & de Waal, F. B. M. (2014). The evolution of responses to (un)fairness. *Science, 346*, 1251776.

Burkett, J. P., Andari, E., Curry, D. C., de Waal, F. B. M., & Young, L. J. (2016). Oxytocin-dependent consolation behavior in rodents. *Science, 351*, 375–378.

Caruana, R. (2007a). Morality and consumption: Towards a multidisciplinary perspective. *Journal of Marketing Management, 23*(3–4), 207–225.

Caruana, R. (2007b). A sociological perspective of consumption morality. *Journal of Consumer Behaviour, 6*, 287–304.

Caruana, R., & Crane, A. (2008). Constructing consumer responsibility: exploring the role of corporate communications. *Organization Studies, 29*(12), 1495–1519.

de Waal, F. B. M. (2008). Putting the altruism back into altruism: The evolution of empathy. *Annual Review of Psychology, 59*, 279–300.

de Waal, F. B. M. (2013). *The bonobo and the atheist*. New York: Norton.

Eabrasu, M. (2012). A moral pluralist perspective on corporate social responsibility: From good to controversial practices. *Journal of Business Ethics, 110*(4), 429–439.

Giesler, M., & Veresiu, E. (2014). Creating the responsible consumer: moralistic governance regimes and consumer subjectivity. *Journal of Consumer Research, 41*(3), 840–857.

Grayson, K. (2014). Morality and the market place. *Journal of Consumer Research, 41*(2), 6–9.

Hume, D. (1985). *A treatise of human nature*. Harmondsworth, UK: Penguin. (Original work published 1739).

Izberk-Bilgin, E. (2010). An interdisciplinary review of resistance to consumption, some marketing interpretations, and future research suggestions. *Consumption, Markets and Culture, 13*(3), 299–323.

Stoeckl, V. E., & Luedicke, M. K. (2015). Doing well while doing good? An integrative review of marketing criticism and response. *Journal of Business Research, 68*(12), 2452–2463.

Westermarck, E. (1912/1917). *The origin and development of the moral ideas* (2nd ed.). London: Macmillan . (Original work published 1908).

Index

A
Academic philosophy, 1
Adaptation, 33, 36, 38
Adaptive problems, 36
Adherence, 60, 68–71
Algorithmic level, 59, 68
Altruistic punishment, 67
American cultural anthropology, 46
American National Election Study (ANES), 106
Amygdala, 73
Ancestors, 34
Ancestral hominins, ecology of, 31
Ancestral problems, 31
Anthropologists, 45, 53
Anthropology, 43
  ethics in ethnography, 51–55
  ethics as social norms, 43–48
  Foucault and freedom, 48–50
  problems with moral psychology, 44
  relativism, 46, 47
  virtue ethics, 50
Anti-social behaviors, 63
Arbitrary superstition, 5
Aristotle's doctrine, 137

B
BBC Prison Study, 63, 64
Behavior, 31–34, 44, 62
Behavioral economics, 60
Brain, human, 30

C
Charity, 36
Christianity, 22
Cognitive miser model, 97
Cognitive neuroscience, 60, 73
Cognitive psychologists, 32
Cognitive science, 9, 59, 68, 70, 74
  of morality, 60
Cognitive scientist, 1, 2, 59
Cognitive system, analysis, 59
Computational level, 59
Computational modeling, 1, 74
Condemnation, 33
Conditional cooperation, 64
Conscience, 33
Consequentialism, 8–10
Contemporary moral psychology, 6
Contemporary normative ethics, 9
Cooperative behavior, 65
Cooperative interactions, 64
Cooperative relationships, 32
Cross-cultural validity, 46
Culturalist, 44
Culture, 47
Culture of honor, 12

D
Debunking arguments, 22
Democratic dilemma, 109
Deontology, 9, 10
Determinism, 14

B.G. Voyer, T. Tarantola (eds.), *Moral Psychology*,
DOI 10.1007/978-3-319-61849-4

Dictator game, 67
Disagreement, 11, 12
Distal causes, 20
Doubtless, 121
Drift diffusion model, 70, 71
Dual process model of moral cognition, 9, 10
Durkheim, E., 45
Durkheimian sociology, 45

E
Ecology of ancestral hominins, 31
Economists, 32
Emotion in moral decision-making, 68
Emotion manipulation, 18, 20
Emotionally charged intuitions, 18
Empirical data, 9, 10, 21
Empirical moral psychology, 6
Empirical research, 11, 23
Empirically informed normative inquiry, 21
Empirically supported challenges, 13
Enforcement, 60
Ethics in ethnography, 51–55
Evaluations of Government and Society (EGSS), 106
Evolutionary biologists, 65
Evolutionary debunking, 19, 20
Evolutionary dynamics, 74
Evolutionary psychology, 8
Evolutionary theorists, 64
Experimental philosophy, 146, 148, 149
Experimental psychology, 64

F
Fallacy, 7, 46
Filters, 30, 31
Fitness maximization, 35
Flat Earth theory, 29
Foucault, 48–54
Footbridge dilemma, 69
Friend, 35
Fundamental moral disagreement, 11

G
Game theoretic modeling, 31
General relativity theory, 29
The gap, 6–10, 12, 13, 17
  bridging, 21–23
Government, 1
Greene/Singer strategy, 19
Guesswork, 5

H
Habitus, 49
Herding economy, 11
Homosexuality, 8, 15
Human, 1, 64
  and ancestors, 34
  and animals, 37
  brain, 30
  cooperation, 64
  interaction, 1, 64
  mind, 29, 59
  moral judgment and agency, 6
  morality, 2
  nature, 6
  psyche, 1
  psychology, 64
Human agency, 12
Human beings, 5
Hypersensitivity debunking, 22
Hypersocial species, 1
Hyposensitivity debunking, 22

I
Ignoble origins debunking, 22
Illusions of control, 13
Inconsistency debunking, 22
Individual process, 68–73
  norm adherence, 68–71
  norm enforcement, 72, 73
Information, 68
Interactions, 64–67
Interdisciplinary moral psychology research
  behavioral sciences, 125
  cognitive neuroscientists, 129
  cognitive science, 138
  collaborating, 138
  consensus, 122
  critical scrutiny, 125
  curious circumstance, 121
  deductively tight, 122
  descriptive statements, 121
  embarrassment, 126
  ethics and moral, 122
  fame and material goods, 134
  foreign disciplines, 120
  hating philosophy, 121
  horizontal metaphors, 128
  human personality, 122
  inferential barrier, 122
  literature surveys, 119, 137
  meta-analysts, 126
  mistake, 133
  moral psychology, 120

myths, 124
neuroimaging study, 136
null hypothesis, 129, 133
obviousness, 119
optimistic enough, 120
optional stopping, 131
permissive inclusion criterion, 127
personality, 121
posterior probability, 132
p-value, 133
publication bias, 130
real scientific progress, 136
reverse inference, 136
rhetorical space, 123
significance level, 130
scientific literature, 123
scientifically tractable, 120
size matters, 131
smoking and cancer, 126
standard deviation, 132
statistical significance, 131
testimony, 123, 124
theoretical agenda, 128
Intrapersonal trade-offs, 15
Intuitive morality, 22
Inward-facing mechanism, 2, 32–37

K
Kallars community, 52

L
Laboratory experiments, 66
Large-scale survey study, 73
Liberation, 52
Long-term rewards, 34

M
Male mating psychology, 30
Mating mechanisms, 34
Meta-analyses, 126, 134
Methodological orthodoxy in anthropology, 46
Milgram's obedience studies, 12
Moore, G.E., 7
Moral agency, 6, 12–14
character, situation and virtue, 12, 13
freedom of will, 13, 14
Moral assessments, 15
Moral cognition, 10
Moral conviction, 100
Moral decision-making, emotion, 68
Moral dumbfounding, 18
Moral economy, 46
Moral intuition, 6
evolutionary debunking, 19, 20
rationalism and sentimentalism, 18, 19
reliability of intuition, 20
Moral and nonmoral judgment, 6
moral luck, 17
personal identity, 15, 16
Moral norms, 8
Moral philosophy, 10
Moral psychology, 1, 2, 6, 43, 59, 68, 71, 74, 75
consequences of moral conviction, 104–108
conservative advantage, 101–104
de Waal, F., 146–148
Doris, J.M., 152–161
emotions and morality, 88–95
evolutionarily informed study of, 38, 39
Heywood, P., 149
immorality in political discourse, 108–112
intuitionist model, 98–100
moral conviction, 100–108
morality policy and politics, 84–88
political behavior, 95–100
political compromise, 112, 113
political science, 81–84
rationalist model of morality, 95–97
Sauer, H., 148
social context models, 97, 98
Wieser, V.E., 150, 151
Moral relativism, 11, 12
Moral system, 2, 60, 61, 75
Morality, 1, 2, 6, 7, 29, 45, 46, 53, 54
Morality policy, 87, 88
Morally irrelevant factors, 10

N
National Rifle Association (NRA), 97
Naturalistic fallacy, 7
Neural network, 75
Neuroscientists, 32
Nonmorality policy, 88
Non-trustworthy people, 67
Norms, 61–64
adherence, 68–71
enforcement, 72, 73
Normative implications, 6
Normative inquiry, 21
Normative questions, 7

Normative theory, 6, 9–12
consequentialism and deontology, 9, 10
moral relativism, 11, 12

O
Objectivist fallacy, 46
Obsoleteness debunking, 22
Off-track debunking, 22
Oft-repeated Foucauldian, *see* Foucault
Optional stopping, 131
Outward expressions of offense, 37
Outward-facing mechanism, 2, 32, 33, 35–38

P
Parkin, D., 45
Perpetrators, 33, 37
Personal identity, 15
Persons, 15
Philosophers, 3, 5, 6, 12, 18, 29, 32
Philosophical moral psychology, 5
Philosophical theories of human nature, 5
Philosophy, 7
Phineas Gage effect, 15
Physical level, 59
Political discourse
democracies, 84
immorality, 108–112
Political psychology, 95
behavior, 82
science, 81
Political science
economics, 96
moral psychology, 81–84, 89, 94
Political scientists, 3
Politics
attitudinal and behavioral consequences, 88–95
morality policy, 84–88
Practice theory, 49
Primitive cognitive processes, 10
Principia Ethica, 7
Proceedings of the National Academy of Sciences, 135
Psychological debunking arguments, 22
Psychopathic individuals, 18
Public goods game, 65
Punishment
costs of, 67
forms of, 66
functions of, 14
practices of, 14
second-party punishment, 66
third-party punishment, 38, 66, 67, 73, 75

R
Rationalism, 18, 19
Real-life atrocities, 13
Relativism, 46, 47
Reliability of intuition, 20
RepliGate controversy, 136
Reproductive fitness, 35
Reputation, 65
Research scientists, 29
Retributive punishment, 14
Robust intuition, 35

S
Satisficing, 97
Second-hand smoke, 62
Second-party punishment, 66
Self-construal, 64
Self-identification, 127
Sentimentalism, 18, 19
Sexual morality, 8
Single neuron, 60
Situationism, 13
Snake-avoidance mechanism, 31
Social anthropology, 45, 46
Social conformity, 63
Social conservatives, 101
Social norms, 64, 71
Social psychologists, 63
Sociologists, 61, 62
Stanford Prison Experiment, 63. *See also* Zimbardo's prison experiment
Structure-agency problem, 48
Superstition, 5
Switch dilemma, 68

T
Temporoparietal junction (TPJ), 71, 73
Terrestrial primate, 31
Theaetetus, 46
Third-party moral condemnation, 38
Third-party punishment, 38, 66, 73
Transgressors, 67, 68
Treatise of Human Nature, 7
Trivers' theory of reciprocal altruism, 33
Trolley problem, 68
Trustworthy people, 67

U
Ultimatum game, 66, 67
Unconditional authority, 18
Utilitarianism, 15

V
Ventromedial prefrontal cortex (vmPFC), 71, 72
Virtue ethicists, 8, 12
Virtue ethics, 12, 13, 50

W
World religions, 51

Z
Zimbardo's prison experiment, 12. *See also* Stanford Prison Experiment