

DRAFT: Please do not quote without permission of the author; comments welcome at [drkelly@purdue.edu](mailto:drkelly@purdue.edu)

To appear in the *Routledge Handbook of Evolution and Philosophy*, Ed. R. Joyce, New York: Routledge Press.

Last changes: 9-18-16

Word Count: 9811 total; 6717, main text

# Moral Cheesecake, Evolved Psychology, and the Debunking Impulse

By Daniel R. Kelly

## I. Introduction

In his famous discussion of inductive inferences, David Hume gives one of the most notorious debunking arguments in Western philosophy. His target is causation; despite what we naturally believe or how it intuitively seems to us, Hume is skeptical of the common idea that there are necessary connections of cause and effect that can hold between objects and events in the world. Hume also provides an account of the source of the mistake, of where the idea of causation comes from, and why belief in causal connections is so natural and persistent. That account appeals to custom, or mental habit, and the mind's general "propensity to spread itself on external objects" (Hume 1978, 1.3.14). In the case of cause and effect, he says:

"We remember to have had frequent instances of the existence of one species of objects; and also remember, that the individuals of another species of objects have always attended them, and have existed in a regular order of contiguity and succession with regard to them. Thus we remember to have seen that species of object we call *flame*, and to have felt that species of sensation we call *heat*. We likewise call to mind their constant conjunction in all past instances. Without any farther ceremony, we call the one *cause* and the other *effect*, and infer the existence of the one from that of the other. In all those instances, from which we learn the conjunction of particular causes and effects, both the causes and effects have been perceived by the senses, and are remembered: But in all cases, wherein we reason concerning them, there is only one perceived or remembered, and the other is supplied in conformity to our past experience." (Hume, 1978 1.3.6, italics in original)

Hume's strategy here is instructive even if one rejects it and its conclusion that causation is not a feature of the external world. Perhaps most relevant is his appeal to characteristics of our psychological capacities, and his reasons for doing so. Hume gives an alternative explanation for why we seem to 'see' causal connections in the world, why we so naturally interpret the unfolding of events through the lens of cause and effect, even though (he claims) there are no such connections, no causal relations out there in the world to be seen or otherwise perceived. Instead, he directs our attention to what is going on in our heads, suggesting that such relations are a feature of ourselves, associations between our ideas that we mistakenly project onto the objects and events we perceive, misattributing them to the

world we are observing. Hume locates the source of belief in external causal connections in “a productive faculty” of our minds, a psychological process that, in “gilding and staining all natural objects with the colours, borrowed from internal sentiment, raises in a manner a new creation” (Hume [1751] 1983: 88). Thus, Hume’s debunking strategy aspires to give an explanation of its target phenomenon (causation, and a cluster of intuitive beliefs thereof) that 1) illuminates that phenomenon in a new way, and in so doing 2) shows not just *that* previous ways of construing or accounting for the phenomenon went wrong, but *how* and *why* they did as well.

But did Hume really debunk causation, or did he just put it on fresh ontological footing? Did he show that causation does not exist and that belief in it mistaken, or did he just give a better account of what it really is? Perhaps his appeal to our habits of expectation and the operation of other productive psychological faculties serves to clarify and correct some of our intuitive beliefs about the relations between unfolding events, rather than show the whole package to be false. If Hume’s account is understood this way, as explaining why causation seems to be out there in the world, but is in fact a “new creation” produced in our own heads and added to experience by our perceptual apparatus, did he debunk causation, or did he supplant an old and mistaken explanation for a newer and better one?

I’ll return to questions like these, concerning how to interpret applications of this kind of broad Humean strategy and what counts as “debunking,” at the end of the paper. By way of getting there, in section 2 I canvass some of the recent literature on evolution, psychology, and morality, separating out some of the different ways the debunking impulse as manifested there, and endorsing the recent trend towards a more selective, divide and conquer approach. Next in section 3, I turn to recent work on the evolution of the human mind, the kinds of mental structure and innate psychological endowment it suggests, and what this implies about our moral capacities and the mechanisms that underpin them. In section 4 I focus on the relationships that can hold between different components of our moral capacities, the environments and contexts in which they operate, and the functions they perform in those environments and contexts. I argue accurate assessments of those components and their contributions to moral judgment and behavior needs to be selective and fine-grained. I also emphasize that a full assessment needs to take into account the fact that human minds are cultural minds (Henrich 2015), and today operate in cognitive niches that are intensely and unprecedentedly constructed, filled with supernormal stimuli intentionally designed to shape moral attitudes and elicit particular moral responses by their designers. I conclude with some remarks on the Humean lineage of this line of thought, and on how it fits with other recent attempts to square scientific images of the world with their putatively corresponding manifest images.

## II. Morality and The Debunking Impulse

While Hume’s treatment of causation is most concise and well known, he was famously skeptical about many other things besides. Indeed, much of his epistemology was focused on how and why we get things wrong, and many of his explanations of our epistemological mishaps appeal in one way or another to the features and functioning of our minds. Here I’ll be focusing on discussions specifically about morality, and the different ways that the debunking impulse has been expressed and debated in recent work by moral philosophers. Despite their differences, contemporary discussions typically share three

related features that set them apart from Hume's. First, the state of science and the scientific picture of the world have developed enormously since Hume's day in the mid eighteenth century, and so attempts to locate morality with respect to that picture, and contemporary attempts to bring science to bear on moral issues, will be accordingly more sophisticated. Second, a centerpiece of that scientific development and its accompanying naturalistic worldview is evolutionary theory, and discussions of morality have been looking closely at the relationships between the two, trying to discern what evolutionary theory might tell us about human moral nature, and the epistemology and ontology of morality itself. Obviously Hume was unable to do this; indeed, with Darwin's publication of the theory of evolution by natural selection still a century off, Hume was unable to come up with a rebuttal to the Argument from Design that was convincing or intellectually satisfying even to himself. Finally, and related to the two previous points, the models of the mind that feature in many of these contemporary discussions in moral theory depart in fairly radical ways from the broadly empiricist, associationist model Hume worked with (though there are some deep affinities between the two as well; see Fodor 2003). Of particular interest are those models that combine contemporary psychology and contemporary evolutionary theory, on which I will briefly focus in section 3.

A full tour through the recent outgrowth of literature at the intersection of evolutionarily informed behavioral science, empirical moral psychology, and moral philosophy would try most readers' patience, but a few distinctions and landmarks will help orient the discussion to come. In an early *locus classicus*, Mackie (1977) presents an error theory about morality, defending the view that all moral judgments are false (in error), and gives a rudimentary explanation of our persistence in making the error of believing some of them (usually our own) are true. Other important agenda setting contributions to these debates, especially those that dwell on the implications of evolutionary theory for questions in metaethics, include Gibbard (1990), Street (2006), and Joyce (2007). The latter two present distinct but complementary versions of arguments that move from premises about the evolution of our moral capacities to conclusions that serve as skeptical challenges to venerable metaethical doctrines; Street's challenge is to moral realist views while Joyce's challenges the idea that any of our moral judgments are properly justified. The discussion, much of it in response to these key texts, fans out in many interesting directions from there. Clarke-Doane (2012) explores an analogy between moral and mathematical beliefs, arguing that beliefs of both types stand or fall together in the face of evolutionary arguments. Cline (2015) defends a form of moral nativism against a provocative argument made by Machery and Mallon (2010) that morality did not evolve, by which they mean that morality, understood as a distinct category of normativity, is not an adaptation resulting from natural selection; Cline argues against them that debunking arguments that include the premise that human psychology contains some innate, evolved moral cognitive machinery remain viable. Mason (2010) focuses less on details of psychology and more on the structure of genealogical debunking arguments of both moral and religious judgments, and lays out an account of whether and when the genealogy of a judgment is "truth-mooting". He explores the possibilities of applying such truth-mooting debunking strategies to both cognitivist as well as non-cognitivist accounts of judgments in those domains (also see Joyce 2013 for

discussion of the viability of debunking arguments in terms of genealogy and non-cognitivism).<sup>1</sup>

Along side this, another interesting and important line of literature has sprung up. It is largely animated by the spirit of experimental philosophy, and typically proceeds by investigating, first and foremost, folk psychological beliefs about metaethical issues like the objectivity or relativity of moral judgments. Given what is discerned by empirical investigation of folk metaethical beliefs, inferences are typically made about the nature of the cues and psychological mechanisms responsible for such beliefs, or about the implications of those folk beliefs and their sources for positions in metaethical theory itself (Nichols, S. 2004, Goodwin, G., and Darley, J. 2008, Wright, J., Grandjean, P., and McWhite, C. 2013, Uttich, K., Tsai, G. and Lombrozo, T. 2014, and many of the contributions in Knobe and Nichols 2014). While many of these authors see their empirical results as fit to inform questions about metaethics, sometimes ruling out certain possibilities or serving as evidence against some family of metaethical views, they do not often present their conclusions explicitly in terms of “debunking”. For instance, Doris and Plakias’ approach (2007) brings a range of empirical psychological and behavioral facts about evaluative diversity to bear on philosophical debates about metaethical realism via the argument from disagreement. They tentatively endorse a “patchy realism”, arguing that with respect to some but not all moral issues, there is good reason to think that moral disagreement would persist even if all factual disagreement could be resolved, and so claim that with respect to those, but only those, moral issues wherein moral disagreement would persist, the right conclusion to draw is an anti-realist one: is no fact of the matter (see Sommers 2012 for a similar approach to moral issues specifically concerning free will and moral responsibility). A notable exception is Nichols (2015), who explores what he explicitly calls “process debunking” approaches, which draw on empirical findings about the psychological processes that form different moral beliefs. Nichols distinguishes between using such an approach to draw implications for metaethical questions, on the one hand, and for normative ethical questions, on the other. While he is optimistic about the prospects for the former, he is less sanguine about the later.

Two more distinctions useful in taxonomizing different expressions of the debunking impulse can be drawn from Nichols’ discussion. The first is between those that seek to establish metaethical conclusions and those that seek to establish first order normative ethical conclusions. This is more of a continuum than a categorical distinction, but clear, and representative instances of each kind can be easily identified. For instance, most of those mentioned above – Mackie’s error theory, Street’s anti-realism, and Joyce’s moral skepticism – fall in the first category, and these tend to be most commonly associated with the idea of moral debunking. However, many authors have recently advanced positions on questions in the second category, about the normative significance of findings in the evolutionary and empirical sciences, for instance Berker (2009), Kumar and Campbell (2012), Railton (2014), Rini (2013, 2016), Kumar (forthcoming).

---

<sup>1</sup> The literature in this vein is getting more articulated by the day it seems; for some other highlights see Lillehammer, H. (2003), Enoch (2010), Wielenberg, E. (2010). Shafer, K. (2010). Kahane, G. (2011), Shafer-Landau, R. (2012), Fraser, B. (2014), Buchanan, A., and Powell, R. (2015), FitzPatrick, W. (2015), and Vavova, K. (2015); for a book length discussion framed in terms of moral naturalism rather than debunking, see Kitcher (2011).

This leads to a second distinction, this one between global and selective approaches to debunking, or, if not debunking, to simply to bringing science into dialogue with moral theory. Global strategies typically seek to treat all of morality at once (or all moral beliefs, or all moral judgments, or all moral norms, or all of our moral capacities, etc.) taking it (or them) to be a kind unified enough to stand or fall together in the face of attempted debunks. Most attempts to establish metaethical conclusions are global in this sense (with Nichols 2015 and Doris and Plakias 2007 being notable exceptions), and those noted above who attempt to establish normative ethical conclusions proceed with this kind of globalist assumption as well. Selective expressions of the debunking impulse, on the other hand, take more of a divide and conquer approach. They tend to separate out different components of morality and our moral psychology for more specific focus and detailed examination. Selective approaches, implicitly or explicitly, find in morality a number of variegated phenomena that differ in ways that are directly relevant to the sorts of empirically and evolutionarily informed considerations that might be brought to bear in attempts to debunk or simply understand them. They see variety in our moral beliefs, expect different components of our moral capacities to be better or worse at producing moral judgments and behaviors, and as such they typically look more closely at the properties of mechanisms and processes that produce different subsets of moral beliefs and behavioral tendencies, or the evolutionary history and selective regimes that shaped different particular components of our moral psychology.

I favor the selective approach, and in previous work (Kelly 2013) have articulated and defended an argument form that underlies many instances of it<sup>2</sup> Indeed, there has been a recent flowering of philosophical work drawing normative implications for some selective part of morality, or about the influence of some specific component or components of our psychological machinery on behavior and moral judgment. An early instance of this is Singer's argument against a subset of moral intuitions, namely those that drive many objections to utilitarianism (2005). Greene (2013) advances a similar argument while also expanding the base of cognitive neuro-scientific evidence on which the argument draws (See also Joyce 2016 for discussion). Other work has singled out for special focus the character and influence of particular emotions. After languishing in relative obscurity for decades, disgust has come in for much recent attention, with both its variously qualified advocates (Kass 1997, Kahane 1998, 1999, Plakias 2013, Fischer 2015, Kumar ms) and skeptics (Nussbaum, Kelly 2011, Kelly and Morar 2014). Others have directed their attention to empathy (Prinz 2013, also see Bloom 2013), contempt (Bell 2013), hope (Martin 2014), romantic love (Brogaard 2015), anger (Wiegman 2015), and jealousy (Kristjánsson 2016).<sup>3</sup> Despite differing levels of engagement with the cutting edge evolutionary and empirical literature, all of these use facts about the respective target of their selective focus to draw out

---

<sup>2</sup> The key premise of the argument sets up a modus ponens: "If some particular psychological mechanism can be shown to be problematic in a relevant way, and the intuitions or judgments influenced by that psychological mechanism can be identified, then we should disregard, discount or discredit those intuitions and be suspicious of the judgments that they influence, to the extent that we can" (137). This remains neutral on how to individuate mechanisms, and, more importantly, on what 'problematic' amounts to; see section 4 of the main text of this paper for more on the later.

<sup>3</sup> Also see Barrett et al (2016) and the line of research that it is responding to for an attempt to wrestle with the ways that mechanisms that discern and ascribe intentions contribute to moral judgment, especially in light of the evidence they present showing that those contributions can vary across cultures.

normative implications about if, how, and when that component of our psychology can best inform various aspects of morality and moral theory.

As is evident, when it comes to morality, the debunking impulse can take many forms, and philosophers have recently explored many of them. What they all have in common is that they are based on the idea that facts about the character and evolutionary history of our psychological moral capacities can help us understand those capacities, and perhaps improve their functioning and the judgments and practices they support. Many also suggest a suspicion that in one way or another, morality is sometimes not as it seems, and that some portion of our intuitive beliefs about our own moral judgments and their subject matter is mistaken.

### **III. Evolution, Cognitive Architecture and the Adapted Mind**

Setting aside for a moment the varieties of philosophical debunking, let us consider in more detail the picture of the human mind that emerges out of evolutionarily informed cognitive science, and how it construes the mind's component parts, mechanisms and processes, including those that underpin our moral capacities. Though there are of course differences in detail, emphasis and terminology, most contemporary approaches to studying the evolution of the human mind share some common assumptions.

Skipping ahead a couple hundred years from Hume to the turn of the millennium brings with it a radical shift not just in our understanding of biology, i.e. the rise of evolutionary theory, but also in psychological theorizing. Chomsky and Fodor made general appeal to innate ideas and mental structure intellectually reputable again (see especially Fodor 1983), and while neither of them is particularly enthusiastic about evolution, other theorists took their ideas and developed them in an evolutionary context. The most prominent examples of this are Evolutionary Psychologists like Barkow, Tooby & Cosmides (1992), Pinker (1997) and Buss (2005), who hold that the study of human behavior, and more importantly the psychology that produces it, should be informed by adaptationist thinking. Since human minds evolved, they can and should be investigated like the products of natural selection they are. However, the human mind should not be thought of as a single adaptation or one (very) important trait. Rather, it is best conceived of and studied as a collection of loosely affiliated but distinct adaptations.

This is perhaps the most central and common commitment of contemporary evolutionarily informed cognitive science. Rather than seeing the mind as one trait, a single all purpose, domain general information processor, they share a vision of the mind as being comprised in large part (how large varies from one theorist to another) of a number of distinguishable components or subsystems, distinct traits, many of which have their own particular adaptive history, specific domain, proprietary database and algorithms, and functional profile. So the human mind has, for instance: a suite of mechanisms dedicated to interpreting the mental states of others based on their behavior, mechanisms that evolved to detect predators, to avoid pathogens and parasites, to categorize and make inferences about biological entities, to predict the movement of objects based on an intuitive understanding of the physics of medium sized dry goods near the surface of the earth. Each of these evolved to solve a particular problem or related cluster of adaptive problems that were

recurrent in our species history, and each brings to bear a specialized body of innate structure and information to help deal with its associated domain.

Adherents of the Evolutionary Psychology research programme have defended a particularly stark version of this kind of picture. Their massive modularity hypothesis holds that the core machinery of all human minds is composed of a large number of different and distinguishable modules, psychological subsystems whose functioning is marked by a cluster of properties; they are typically specialized, fast, automatic, domain specific, semi-autonomous, informationally encapsulated, and cognitively impenetrable. (See Machery, this volume.) They also appeal to some suggestive metaphors, likening the mind to a Swiss Army Knife, comprised of a variety of ‘elegant machines’, each designed by a specific set of selection pressures to efficiently and effectively solve a particular adaptive problem that was recurrent in our species’ evolutionary past. Like the distinct tools of a Swiss Army Knife, however, each of the ‘elegant machines’ of the human mind operates fairly autonomously from the rest. Moreover, just as, say, the corkscrew is wonderful for opening a bottle of wine but clumsy if not outright useless for most other tasks, so too is each component of the human mind well-fitted to its associated task. That specialization comes at the same kind of cost in the psychological cases as well, though, and the performance and effectiveness of a mechanism can fall off abruptly when it operates outside of its evolved domain.

Another area to which researchers have brought this kind of broad picture to bear, and for which they have posited a number of such innate psychological mechanisms is, roughly, that of morality. Though it remains unclear exactly how to delimit the scope of this domain, human moral capacities are often taken to involve mechanisms dedicated to producing altruistic behavior, navigating the social world and facilitating cooperation, distinguishing intentional from unintentional behavior, acquiring and complying with social norms, detecting and punishing norm violators, and recognizing and responding to group membership and status.<sup>4</sup> In the next section, I will focus on two features of these capacities and the mechanisms that underlie them: that they are likely to be especially sensitive to cultural influences and shifting context, and that they can be assessed along a number of orthogonal dimensions that can be easily conflated, but are importantly distinct.

#### **IV. Assessing Our Moral Capacities in a Hyper-Constructed Cultural Niche: Fitness, Functions, and Supernormal Stimuli**

The picture of the mind coming out of cognitive science and evolutionary thought about human cognition is that it is composed of a collection of different systems, mechanisms, processes, each of which evolved in the face of different selective pressures to solve different adaptive problems, and which are to some extent isolated from each other, either operationally, informationally, or both.<sup>5</sup> This picture dovetails and supports the

---

<sup>4</sup> In addition to those works cited in the main text, see Doris et al (2010) and Henrich (2015) for useful overviews of much recent research in this vein.

<sup>5</sup> While a picture that includes at least some mechanisms and mental structure of this type is accepted by most theorists, debate about that structure continues to rage along a number of dimensions, including how much and which parts of it are universal, how much can be accounted for by social learning or how much credit for certain behavioral capabilities and practices can be attributed to structured environments and constructed niches, how much and which parts of internal mental structure are innate, how much innate content is likely to

selective, divide and conquer strategy separated out in section 2, because it raises the very real possibility that the credentials of the different components of our moral cognition need to be assessed separately, and on their own terms. For instance, some components might be equipped with “shoddy software” or bear architectural features that make them ill-suited to contribute to morality in the way they typically do. Others might be well-suited, providing exactly what we would most want in guiding our moral judgments, deliberations, and actions. Indeed, the assessment of these, in this piecemeal way, can turn up many different options or concepts on how to conceive the relationships between the mechanisms, the contexts in which they operate, and the contributions they are likely to make to moral deliberation, judgment, and action.

For instance, we might conclude, after considering the operational and evolutionary specs of some particular component of the moral mind, that it is an elegant machine. This would suggest that the component evolved specifically to perform a particular function, and it carries out the job in a concise, efficient, elegant, and largely accurate way. A common alternative to this is that we might see other moral mechanism as a kludge. Saying a piece of the mind is a kludge is to claim that while it is able to perform whatever function it does, it does so inefficiently, often in a roundabout way; it works, but with all of the streamlined optimality of a Rube Goldberg machine.<sup>6</sup> The connotations of the term “kludge” also suggest that the mechanism itself was not originally designed (or in the case of evolutionary explanations, did not initially evolve) to perform the particular function we find it performing, or if it is multi-functional, the particular function we are interested in assessing.

Though these two common sets of concepts are sometimes applied to the mind as a whole, they can easily be transferred to evaluate individual components of it, and are arguably more usefully applied in this more selective way. Both also presuppose a distinction between the function, on the one hand, and mechanism that performs that function, on the other.<sup>7</sup> This distinction is particularly useful, and can be developed in several ways. It is even more useful when paired with the distinction between the *proper* function of a trait or mechanism, i.e. the function it was selected to perform, and whose performance explains its existence, why it initially spread through a population and was passed along to subsequent generations, and, on the other hand, the *actual* function of the trait, the function that it performs in its current environment.<sup>8</sup>

Appreciating the difference between proper and actual functions points to a final important distinction, this one between a mechanism itself and the function or functions it

---

be included with those structures and mechanisms, and about what topics and in what representational formats. These remain open empirical questions.

<sup>6</sup> See Marcus (2007) for elaboration of the idea into a general account of the mind, and Stich (2007) for discussion of the distinction between elegant machine and kludge accounts of human moral psychology.

<sup>7</sup> Philosophical debates about functions are vexed, but see Cummins (1975) for an important causal role account of functional analysis, Millikan (1989) and Neander (1991) for accounts of proper functions and etiological or selected effects accounts of functional analysis, and Walsh (1996) for an account that attempts to capture the virtues of both etiological and causal role accounts.

<sup>8</sup> Sperber (1996) introduces and develops this distinction. Also note that it is complicated by the idea of an exaptation. An exaptation is a trait, mechanism, or other feature of an organism or its mind that, because it was not selected for or to do anything, does not have a proper function, but rather initially came into existence as a byproduct of selection for some other trait. Despite this history, an exaptation may have nevertheless acquired one or more actual functions that it performs in its current environment.



performs. Assessing a mechanism as an elegant machine or a kludge means assessing it in terms of how well it performs a particular function. But functions can change, and with such changes the mechanism itself can be assessed differently. It could very well be the case that a particular mechanism is elegantly designed to perform its proper function, and does so (or did so) efficiently and effectively. However, that same mechanism could acquire other functions, functions that it is able to perform with some degree of effectiveness, but that it does not or cannot perform in anything like an elegant, optimal manner. (I have defended a view similar to this about disgust, and the differences in its fairly efficient performance of its primary functions dealing with poisons and parasites and the rather kludgy performance of its acquired functions in the social domain.) Thus a mechanism might appear to be an efficient, elegant machine in one context, and from the perspective of one standard of evaluation, but a kludgy, clunky Rube Goldberg-esque machine in another. Big Mistake or Mismatch hypotheses are extreme versions of this kind of account. Such accounts are often invoked to explain the existence of a trait or tendency that appears to be not just inefficient, a kludgy, wonky solution to some adaptive problem, but as outright maladaptive, actively decreasing the organism's fitness. Such putatively maladaptive traits present a puzzle for evolutionary theorists, and Big Mistake and Mismatch hypotheses offer a solution by construing the trait or tendency as only producing maladaptive behavior because it is operating in an environment or context much different than the one in which it initially evolved. While it produced adaptive benefit in that initial environment, the contextual shifts are such that it is mismatched to its current one, so only the behaviors it currently produces look like a "mistake" from the point of view of evolution.

These distinctions, between psychological mechanisms and the functions they perform, between the mechanisms and the context in which they perform those functions, between elegant machines and kludges, do not yet say anything about morality, or assessments of the *ethical status* or *moral significance* of any particular mechanism or the intuitions, deliberations, judgments and behaviors it contributes to. Rather, they are assessments made largely from an engineering perspective. Adding an additional, separate moral standard of evaluation to the mix presents a further complication, but the preceding discussion suggests a network of distinctions into which it can be systematically fit. For example, it clarifies the possibility that a single mechanism can be an elegant machine with respect to its proper function, and that its proper function can also be its actual function, i.e. it is well-designed and efficiently performs the function that it evolved to perform, which it continues to perform in its current environment. Even in light of this, the possibility remains that in performing that function the mechanism is also doing something morally objectionable, making a contribution to moral cognition and behavior that runs afoul of our moral standards of assessment. In such cases, the function the mechanism performs is itself morally undesirable.

Put another way, elements of our moral psychology might, in a particular context, perhaps even in current actual environments, be doing exactly what they were designed to do, and doing it well. However, they could have been selected to generate intuitions, judgments and actions that our current ethical perspective condemns, that we see in our wisdom are immoral. Our "moral" psychology may have a dark side, and evolution may have equipped us with a set of capacities for navigating our social worlds some of which are well engineered, evolutionarily effective – fitness enhancing – but morally nasty. Such mechanisms are not "kludges" in the traditional sense at all; they are elegant machines, they

are just elegantly designed to, and are actually efficiently doing, something unethical. This is one way to develop and apply the more general point that the categories of engineering efficiency, evolutionary fitness (of whatever sort), and moral worth all cross cut each other. Mechanisms that drive male proprietary attitudes towards women, or those that support in-group and out-group biases, ethnocentrism, cronyism, nepotism, and other types of what we now consider untenable forms of discrimination and prejudice may fit this description; well designed, efficient, fitness enhancing, but immoral.

A slightly different way our moral psychology can go awry is by producing immoral byproducts. Intuitively, a mechanism can attempt to do something that is morally acceptable, to perform an ethically unobjectionable function, but just do it poorly. However, we are now in a position to see that “poorly” has several distinct senses, and the differences can be specified thus: our moral assessment of the task or function is positive, i.e. the aimed at contribution to behavior and judgment is morally acceptable; but the engineering assessment of the mechanism’s performance of that function is negative, i.e. the piece of our moral psychology actually performing that function is kludgy and suboptimal from an engineering perspective. Finally, however, the specific inefficiencies and byproducts generated by the kludgy performance of that function by that mechanism are themselves morally undesirable. The ways in which the mechanism fails to optimally perform its task generate psychological and behavioral effects that, e.g. impel us to do unethical things or induce us to make morally unacceptable judgments.<sup>9</sup>

This brings me to a final way in which appreciating the evolved features of our minds and the way their innate elements interact with the environments in which they operate can shed light on how components of moral psychology can go awry, or in this case, be actively led astray. This intriguing possibility has received little attention, at least under this description, but seems increasingly relevant to moral assessment of the deliverances and influences of different components of human moral psychology, especially in today’s world. It bears a similarity to mismatch hypotheses and the Evolutionary Psychologists’ idea of a Stone Age Mind in a Modern World in that it again focuses on the relationship between the functioning of a psychological mechanism and the environment in which it operates, but its emphasis is slightly different. Consider the general idea of a supernormal stimulus (or superstimuli): an exaggerated version of a stimulus for which there is an existing response tendency, which is embellished in ways that elicit stronger and more exaggerated instances of that response tendency than normal.<sup>10</sup> A familiar example is cheesecake:

“We enjoy strawberry cheesecake, but not because we evolved a taste for it. We evolved circuits that gave us trickles of enjoyment from the sweet taste of ripe fruit, the creamy mouth feel of fats and oils from nuts and meat, and the coolness of fresh water. Cheesecake packs a sensual wallop unlike anything in the natural world

---

<sup>9</sup> I have argued that the production of concerns about stigma, ethical taint, spiritual pollution and moral contamination that accompany moral judgments driven by disgust fit this description (Kelly 2011); also see Plakias (2013) for an opposing view of the value of disgust’s contamination sensitivity when operating in the social domain.

<sup>10</sup> The concept has roots in classic ethology (Tinbergen 1951, Lorenz 1981), but has also appeared in more contemporary discussions concerning humans and human cognition; see De Block and Du Laing (2010) for examples and assessment of different applications of the idea in the social sciences.

because it is a brew of megadoses of agreeable stimuli which we concocted for the express purpose of pressing our pleasure buttons” (Pinker, 1997, p. 524, cited in De Block and Du Laing 2010).

Moving from broadly physiological and sensory “pleasure buttons” to more cognitive and psychological ones, Hurley, Dennett, and Adams (2011) incorporate the idea of superstimuli into their theory of humor, portraying jokes as analogous to a range of other familiar elements of contemporary life:

“Some human artifacts – paintings and sculptures, and pornography, but also music and even aspects of religion – have been devised as supernormal stimuli that (over) stimulate our instinctual systems, producing more intense reactions than they were designed (by natural selection) to deliver. ... jokes are prime examples of supernormal stimuli that take advantage of our natural propensity for humor-detection in much the same way that perfumes, makeup, artificial sweeteners, music, and art give us exaggerated experiences with respect to the natural world. Thanks to their refined designs, they tend to have the power to induce in us a far stronger and richer sense of the ludicrous than everyday ‘found’ stimuli, however humorous” (159).

The picture of the mind here should look familiar; it assumes a broadly modularized structure described above, composed of domain specific mechanisms serving different purposes, producing distinctive responses to a set of proprietary stimuli relevant to the adaptive problems they evolved to deal with in the past. Several more points can help connect this concept back to the main line of discussion about selective debunking.

First, the buttons that supernormal stimuli push need not activate pleasurable responses; they can trigger negative or neutrally valenced responses as well.<sup>11</sup> Second, nothing in the notion of a supernormal stimulus would render it unsuitable for use in thinking about our moral capacities, and the types of cues different mechanisms that underpin those capacities respond to. There is nothing confused or contradictory about the idea of moral cheesecake; though if your daily experience (especially media consumption) is

---

<sup>11</sup> The internet spawned literary genre of “creepypasta” provides good examples of intricate, joke-like supernormal stimuli intentionally designed to push a negatively valenced psychological button. Here is an example called “Bad Dream”:

‘Daddy, I had a bad dream.’  
You blink your eyes and pull up on your elbows. Your clock glows red in the darkness — it’s 3:23.  
‘Do you want to climb into bed and tell me about it?’  
‘No, Daddy.’  
The oddness of the situation wakes you up more fully. You can barely make out your daughter’s pale form in the darkness of your room. ‘Why not, sweetie?’  
‘Because in my dream, when I told you about the dream, the thing wearing Mommy’s skin sat up.’  
For a moment, you feel paralysed; you can’t take your eyes off of your daughter. The covers behind you begin to shift.

As Lake describes it, “‘Bad Dream’ resembles a joke in set-up and punchline but, rather than a laugh, it delivers Lovecraftian dread. As in an effective joke, it is pared down to the absolute minimum, and the scene is set with literary economy. Very little could be cut without severing the thing’s vital tendons” (<https://aeon.co/essays/creepypasta-is-how-the-internet-learns-our-fears>).

anything like mine you might think that supernormal stimuli that push our moral buttons are much less likely to be designed to trigger positive responses like admiration, elevation, or awe and much more likely to be designed to trigger negative responses like disgust, spite, or outrage.<sup>12</sup> Third, while both cheesecake and jokes are intentionally designed cultural artifacts, supernormal stimuli needn't be. Indeed, the concept does not purport to pick out a phenomenon unique to humans, and exaggerated versions of proprietary stimuli might occur naturally, or in the course of normal social interactions (benefitting from a particularly brave or selfless act, or witnessing an especially cruel one). That said, cheesecake and jokes are good illustrations because most instances of supernormal stimuli pertinent to humans and modern life are indeed designed by people, crafted specifically to activate targeted psychological components and elicit selective responses and behaviors. This leads to a fourth point, which is that the extent to which many of us now live in environments that are massively culturally constructed, littered with intentionally designed supernormal stimuli of this kind, is unprecedented in human history.<sup>13</sup> The cognitive niche in which our psychological machinery, including our moral capacities and their underlying mechanisms, are being asked to operate is media drenched, saturated with advertising and political messaging backed by sophisticated marketing research and funded by entities with virtually bottomless pockets. We – we WEIRDos, anyway<sup>14</sup> – are constantly inundated with moral cheesecake and other supernormal stimuli that, along with other subtle and not so subtle technologies of perception shaping and narrative crafting, are in constant competition to capture and hold our attention, and then to push our psychological buttons in ways that nudge us to form particular beliefs and desires, engage in certain kinds of behaviors, and, often, make specific moral judgments, all of which will serve the agendas and ends of their creators; buy this, click on that, vote for this candidate, support that referendum, be very suspicious of people who look like this.<sup>15</sup>

From the point of view of the discussion above, producing and propagating supernormal stimuli for a particular psychological mechanism is one way of actively changing the environment in which that mechanism operates, and in extreme cases can change or expand the actual function of the mechanism in question. Failing to take this into account when performing a moral assessment of the deliverances of those mechanisms, or ignoring this possibility when considering the normative significance or ethical implications of what we know about the evolution and functioning of different components of human moral psychology, seems foolish, especially given the informationally engineered contexts in which they are typically operating in the modern world. Philosophers have an understandable

---

<sup>12</sup> For recent work on the evolution, psychology and signaling function of moral outrage see Jordan et al. (2016), and for spite see Forber and Smead (2014).

<sup>13</sup> One insightful commentator on this situation dubbed it the Total Noise of contemporary culture (American contemporary culture, specifically; Wallace 2007, also see Kelly 2015 for discussion).

<sup>14</sup> That is: Western, Educated, Industrialized, Rich, and Democratic. See Henrich et al. 2010.

<sup>15</sup> Thaler and Sunstein (2009) defend the use of such nudging techniques in public policy for creating default strategies and choice architectures that will engage selective features of our psychology in ways designed to promote public good and individual health. Resistance in the name of personal autonomy to the use of such techniques in the public domain, and for such putatively benign purposes as, e.g. boosting the number of organ donations, seems to me almost quaint in light of the rampant use of similar (and probably more sophisticated version of) techniques by entities, firms, corporations, etc. in the service of ends that overlap with the good of the public or most individuals contingently if at all; indeed, such resistance seems to me not only quaintly old-fashioned but misguided (see Kelly and Morar in press for some discussion.)

tendency to want to formulate and address questions in their most general, universal, and timeless form, but here I think it does them a disservice.<sup>16</sup> Moral cognition ‘in the wild’ today isn’t just extremely social (cf. Doris and Nichols 2012), but takes place in an increasingly meticulously constructed cognitive niche loaded with meticulously constructed cultural artifacts, pieces of moral cheesecake designed to push hard on our moral buttons, inducing specific actions and attitudes. A selective approach informed by the tools of evolutionary theory and cognitive science will be able to help us understand the different ways in which this happens, and what conclusions we might draw from it, and what its moral significance might be.

## **V. Conclusion: The Debunking Impulse Revisited**

I began this essay with Hume’s argument about causation, and by posing some questions about whether or not that argument actually debunked anything. I then separated out different ways the debunking impulse has manifested in recent debates about morality, particularly those debates that draw on current empirical psychology and evolutionary theory. I advocated a divide and conquer approach to drawing implications from science for moral theory, and described a common picture of the structure of the human mind and its moral machinery, innate and otherwise, that I hold fits with and supports that selective approach. Particular applications of that selective strategy may result in negative conclusions, in what would comfortably be called the debunking of a circumscribed set of claims or views. But other applications might suggest implications that are neutral, positive, or perspective altering in ways not naturally described as debunking.

Even if Hume’s argument is accepted in full, I am not sure if it would be accurate to say that it debunked causation, or showed causation to be an illusion. I am likewise unsure about what would count as the fully fledged debunking of different metaethical or normative ethical views, or under what conditions it would be more accurate to say we are just coming to a deeper and better appreciation of the subject matter in question. In these domains, as in others, as our scientific understanding of the phenomena develops, the scientific and manifest images of them will become increasingly dissimilar, growing further apart. The concepts and vocabulary developed to produce a satisfactory explanation for those phenomena will, if they are to be genuinely explanatory, differ from the concepts and vocabulary used to describe the phenomena in the first place. Over time the needed explanatory concepts may depart radically from the folk theoretic concepts of the phenomena, and may also appeal to entities foreign to or at odds with common sense. In many cases it will be useful to allow gains in our scientific understanding of a phenomenon to feed back into and inform what we mean by the concepts and words we use to conceive of and talk about it. None of this need threaten belief in the existence of the phenomena themselves, though. Explanans are always pitched at a different level and stated in different terms than their explananda, and no easy ontological or sweeping ethical conclusions can be drawn from the fact that the kind of language useful for describing a set of patterns or

---

<sup>16</sup> For an interesting parallel discussion of philosophy being ill-served by its urge toward the timeless, see Millgram (2015, especially chapters 1 and 2).

participating in a set of practices departs, even in radical ways, from the language needed to explain different aspects of those patterns and practices (Fodor 1974, 1998, cf. Stich 1996).<sup>17</sup>

I see no reason to think the phenomena of morality are different from any others in this respect, even if the boundaries demarcating the phenomena that are distinctively moral from those that are not remain difficult to bring into focus, e.g. Sinnott-Armstrong and Wheatley (2012). As Hume's argument about causation showed, the invocation of the workings of the mind to explain, illuminate or otherwise inform an area of inquiry is not unique to morality, and invocations of our current, empirically and evolutionarily informed understanding of those workings continue in a range of areas of philosophy to this day (e.g. Maudlin 2010, Goldman 2015, Paul 2016). But whether or not those who draw them wish to frame their conclusions and implications in terms of debunking, moral theory will be best served if philosophers continue to engage deeply with the details of current cognitive science and evolutionary theory, drawing out careful, selectively focused implications for the issues that most concern them.

---

<sup>17</sup> Chalmer's (2001) suggestion about how to describe the beliefs of someone who is a brain in a vat is intriguing, and points to a response one might make to global debunking arguments of moral realism that is similar in spirit to my comments here in the main text. According to Chalmers, a person who is a brain in a vat does not, as is typically assumed, suffer from massive delusion, possessing mostly false beliefs about the external world ("I am in Seattle", "I am typing on my laptop with my two hands", "there is a stack of books next to a half full bottle of Liefmans Cuvee-Brut on the table in front of me", etc.) Rather, most of the envatted person's mundane, particular beliefs about the to-ing and fro-ing of things in 'the external world', including those just listed, are true. She only has a few false beliefs, namely those more abstract ones about their underlying explanation and the fundamental ontology of the world in which she lives, i.e. she is wrong in her belief that the world and its concrete objects are at root fundamentally physical and governed by physical laws. Her 'external world' does in fact contain things like cities and laptops, indescribably good Belgian beers, hands that type and books that stack, but those things are not completely as she thinks they are, because they are at root computational entities rather than, as she believes, physical ones. One might construct a similar line of argument preserving the truth of particular moral beliefs ("don't murder innocent babies for fun" "give to charity when possible" "treat others as you wish to be treated") even if the underlying explanation and fundamental ontology of their subject matter are not as common sense (or even philosophic elaborations thereof) might suppose.

## References

- Barkow, J., Cosmides, L., & J. Tooby. (1992). *The Adapted Mind: Evolutionary Psychology and the Generation of Culture*. New York: Oxford University Press.
- Barrett, C. et al. (2016). "Small-scale societies exhibit fundamental variation in the role of intentions in moral judgment," *Proceedings of the National Academy of Sciences of the United States of America*.113(17): 4688-4693.
- Bell, Macalester. (2013). *Hard Feelings: The Moral Psychology of Contempt*. New York: Oxford University Press.
- Berker, S. (2009). "The normative insignificance of neuroscience," *Philosophy and Public Affairs*, 37: 293–329.
- Blackburn, S. (1998). *Ruling Passion*. London, Oxford University Press.
- Bloom, P. 2013. "The baby in the well: the case against empathy," *The New Yorker*, May 20.
- Brogaard, B. (2015). *On Romantic Love: Simple Truths about a Complex Emotion*. New York: Oxford University Press.
- Buchanan, A., and Powell, R. (2015). "The Limits of Evolutionary Explanations of Morality and Their Implications for Moral Progress," *Ethics*, 126(1): 37-67
- Buss, D. (2005). *The Handbook of Evolutionary Psychology*. New York: Wiley.
- Carruthers, P. (2006). *The Architecture of the Mind*. New York, Oxford University Press.
- Chalmers, D. (2001). "The Matrix as Metaphysics," available at <http://consc.net/papers/matrix.html>
- Clarke-Doane, J. (2012). "Morality and Mathematics: The Evolutionary Challenge," *Ethics*, 122(2): 313-340.
- Cline, B. (2015). "Nativism and the Evolutionary Debunking of Morality," *Review of Philosophy and Psychology*, 6: 231–253
- Cummins, R. 1975. "Functional Analysis", Reprinted in E. Sober (ed.), 1994. *Conceptual Issues in Evolutionary Biology*, 49-69. Cambridge, MA: The MIT Press.
- De Block, A, and Du Laing, B. (2010). "Amusing ourselves to death? Superstimuli and the evolutionary social sciences," *Philosophical Psychology*, 23(6): 821 – 843.
- Doris, J. and The Moral Psychology Research Group (eds.), (2010). *The Moral Psychology Handbook*. New York: Oxford University Press.
- Doris, J. and Nichols, S. 2012. "Broadminded: Sociality and the Cognitive Science of Morality." In E. Margolis, R. Samuels, and S. Stich (eds.), *The Oxford Handbook of Philosophy and Cognitive Science*, pp. 425-53. Oxford: Oxford University Press
- Doris, J. & Plakias, A. (2007). "How to Argue About Disagreement: Evaluative Diversity and Moral Realism." In W. Sinnott-Armstrong (Ed.), *Moral Psychology, vol. 2: The Biology and Psychology of Morality* (pp. 303-332). Oxford: Oxford University Press.
- Enoch, D. 2010. "The Epistemological Challenge to Metanormative Realism: How Best to Understand It, and How to Cope with It," *Philosophical Studies* 148(3): 413–438.
- Fischer, R. 2015. "Disgust as heuristic," *Ethical Theory and Moral Practice*, 1-15.
- FitzPatrick, W. (2015). Debunking evolutionary debunking of ethical realism," *Philosophical Studies*, 172: 883–904
- Fodor, J. (1974) "Special Sciences," *Synthese*, 28, 97-1 15.
- Fodor, J. (1983). *The Modularity of Mind*. Cambridge, MA: The MIT Press.
- Fodor, J. (1997). "Special Sciences: Still Autonomous After All These Years" *Noûs*, 31(11): 149-163.
- Fodor, J. (2003). *Hume Variations*. Oxford: Oxford University Press.

- Forber, P. and Smead, R. (2014). "The evolution of fairness through spite," *Proceedings of the Royal Society B*, 281: 20132439.
- Fraser, B. (2014). "Evolutionary debunking arguments and the reliability of moral cognition," *Philosophical Studies*, 168: 457 – 473.
- Gibbard, A. (1990). *Wise Choices, Apt Feelings*. Cambridge: Harvard University Press.
- Goldman, A. (2015). "Naturalizing Metaphysics with the Help of Cognitive Science." In *Oxford Studies in Metaphysics* (vol. 8), edited by Karen Bennett and Dean Zimmerman, 171- 214. Oxford: Oxford University Press.
- Goodwin, G., and Darley, J. (2008). "The psychology of meta-ethics: Exploring objectivism" *Cognition*, 106: 1339–1366.
- Greene, J. (2013). *Moral Tribes: Emotion, Reason, and the Gap Between Us and Them*. New York: Penguin Press.
- Henrich, J., Heine, S. and A. Norenzayan. (2010). "The weirdest people in the world." *Behavioral and Brain Sciences*. 33 (June): 61-135.
- Henrich, J. (2015). *The Secret of Our Success: How Culture Is Driving Human Evolution, Domesticating Our Species, and Making Us Smarter*. Princeton, NJ: Princeton University Press.
- Hume, D. (1751/1983). *An Enquiry Concerning the Principles of Morals*, Indianapolis: Hackett.
- Hume, D. (1978). *A Treatise of Human Nature*.
- Hurley, M., Dennett, D., and Adams, R. (2011). *Inside Jokes: Using Humor to Reverse Engineer the Mind*. Cambridge, MA: The MIT Press.
- Jordan, J., Hoffman, M., Bloom, P. and Rand, D. (2016). "Third-party punishment as a costly signal of trustworthiness," *Nature*, 530, 473–476.
- Joyce, R. 2007. *The Evolution of Morality*. Cambridge: MIT Press.
- Joyce, R. 2013. "Irrealism and the Genealogy of Morals," *Ratio*, 26 (4): 351 - 372.
- Joyce, R. (Forthcoming). "Reply: Confessions of a modest debunker," to appear in *Explanation in Mathematics and Ethics*, U. Leibowitz & N. Sinclair (eds.).
- Kahane, G. (2011). Evolutionary debunking arguments. *Nous*, 45, 103–125.
- Kahan, D. (1999). The Progressive Appropriation of Disgust. In (ed.) Susan Bandes, *The Passions of the Law*. New York. New York U Press. pp.63-80.
- Kahan, D. 1998. The Anatomy of Disgust in Criminal Law. *Michigan Law Review*, 96(May): 1621-1657
- Kass, L. (1997). "The Wisdom of Repugnance," *The New Republic*, 216(22), available online at [http://www.catholiceducation.org/articles/medical\\_ethics/me0006.html](http://www.catholiceducation.org/articles/medical_ethics/me0006.html) and in *Life, Liberty, and the Defense of Dignity: The Challenge to Bioethics*, (New York: Encounter Books, 2002).
- Kelly, D. (2011). *Yuck! The Nature and Moral Significance of Disgust*. Cambridge, MA: The MIT Press.
- Kelly, D. (2013). "Selective Debunking Arguments, Folk Psychology and Empirical Moral Psychology," *Advances in Experimental Moral Psychology: Affect, Character, and Commitments*, Eds. Wright, J.C. & Sarkissian, H. New York: Continuum Press.
- Kelly, D. (2015). 'David Foster Wallace as American Hedgehog', *Freedom and the Self: Essays in the Philosophy of David Foster Wallace*, Eds. S. Cahn and M. Eckert. New York: Columbia University Press, pages 109 - 132.
- Kelly, D. and Morar, N. (2014). "Against the Yuck Factor: On the Ideal Role for Disgust in Society," *Utilitas*, 26(2): 153-177.
- Kelly, D. and Morar, N. (in press). "Nudging and the Ecological and Social Roots of Human Agency", commentary on Douglas MacKay and Alexandra Robinson "Registering



- Organ Donors: Nudges and Respect for Autonomy,” to appear in *The American Journal of Bioethics*.
- Kitcher, P. (2011). *The ethical project*. Cambridge, MA: Harvard University Press.
- Knobe, J. and Nichols, S. (2014). *Experimental Philosophy, Vol. 2*. Oxford: Oxford University Press.
- Kumar, V. (forthcoming). “The Ethical Significance of Cognitive Science,” to appear in *Current Controversies in Cognitive Science*, Eds. Leslie, S. and Cullen, S.
- Kumar, V. (manuscript). “Foul Behavior”
- Kumar, V. & Campbell, R. (2012). "On the Normative Significance of Experimental Moral Psychology," *Philosophical Psychology*, 25: 311-30.
- Lillehammer, H. (2003). “Debunking morality: evolutionary naturalism and moral error theory,” *Biology and Philosophy* 18: 567–581.
- Lorenz, K. (1981). *The foundations of ethology*. New York: Springer.
- Mackie, J. L. (1977). *Ethics: Inventing right and wrong*. New York: Penguin Books.
- Machery, E. and Mallon, R. (2010). ‘Evolution of Morality’, *The Oxford Handbook of Moral Psychology*, Eds. J. Doris et al. New York: Oxford University Press.
- Marcus, G. (2007). *Kluge: The haphazard construction of the human mind*. New York: Mariner Books.
- Martin, A. (2014). *How We Hope: A Moral Psychology*. Princeton, NJ: Princeton University Press.
- Mason, K. (2010). “Debunking Arguments and the Genealogy of Religion and Morality,” *Philosophy Compass* 5/9: 770–778.
- McGeer, T. (2008). “Trust, Hope and Empowerment,” *Australasian Journal of Philosophy*. 86(2): 237 – 254.
- Maudlin, T. (2010). *The Metaphysics Within Physics*. New York: Oxford University Press.
- Millgram, E. (2015). *The Great Endarkenment: Philosophy for an Age of Hyperspecialization*. New York: Oxford University Press.
- Millikan, R.W. (1989). “In Defense of Proper Functions”, *Philosophy of Science* 56: 288-302.
- Neander, K. (1991). “Functions as Selected Effects: The Conceptual Analyst’s Defense”, *Philosophy of Science* 58: 168-184.
- Nichols, S. (2004). *Sentimental Rules: On the Natural Foundations of Moral Judgment*. New York: Oxford University Press.
- Nichols, S. (2015). “Processing Debunking and Ethics,” *Ethics*, 124(4): 727-749.
- Nussbaum, M.2004. *Hiding from humanity: Disgust, shame, and the law*. Princeton, NJ: Princeton University Press.
- Paul, L. (2016). “Experience, Metaphysics, and Cognitive Science,” *A Companion to Experimental Philosophy*, (Eds). Sytma, J and Buckwalter, W. New York: Blackwell. p. 417-433
- Pinker, S. (1997). *How the mind works*. New York: W.W. Norton & Co.
- Plakias, A (2013). “The Good and the Gross,” *Ethical Theory and Moral Practice*, 16(2): 261-278.
- Prinz, J. (2011). “Against Empathy,” *The Southern Journal of Philosophy*, 9: 214–233.
- Railton, P. (2014). “The affective dog and its rational tale,” *Ethics*, 124(4): 813-859.
- Rini, R. (2013). “Making Psychology Normatively Significant,” *Journal of Ethics*, 17(3):257-274.
- Rini, R. (2016). “Debunking Debunking: A Regress Challenge for Psychological Threats to Moral Judgment,” *Philosophical Studies*, 173(3): 675-697.
- Schukraft, J. (2016). “Carving Intuition at its Joints,” *Metaphilosophy*, 47(3): 326 – 352.

- Shafer, K. (2010). "Evolution and Normative Skepticism," *The Australasian Journal of Philosophy*, 88 (3): 471-488.
- Shafer-Landau, R. (2012). Evolutionary debunking, moral realism and moral knowledge. *Journal of Ethics and Social Philosophy*, 7, 1–37.
- Singer, P. (2005). "Ethics and intuitions," *The Journal of Ethics*, 9: 331–352.
- Sinnott-Armstrong, W. & Wheatley, T. (2012). "The Disunity of Morality and Why it Matters to Philosophy," *The Monist* 95(3): 355-377.
- Sommers, T. (2012). *Relative Justice: Cultural Diversity, Free Will, and Moral Responsibility*. Princeton, NJ: Princeton University Press.
- Sperber, D. (1996). *Explaining culture: A naturalistic approach*. New York: Blackwell Publishers.
- Sterelny, K. (2003). *Thought in a Hostile World: The Evolution of Human Cognition*. New York, Blackwell.
- Sterelny, K. (2012). *The Evolved Apprentice: How Evolution Made Humans Unique*. Cambridge, MA. The MIT Press.
- Stich S. (1996). *Deconstructing the Mind*. New York: Oxford University Press.
- Stich S. (2006). "Is morality an elegant machine or a kludge?" *Journal of Cognition and Culture*, 6(1), 181-189.
- Street, S. (2006). "A Darwinian Dilemma for Realist Theories of Value," *Philosophical Studies*, 127(1): 109-166.
- Thaler, R. and Sunstein, C. 2009. *Nudge: Improving decisions about health, wealth, and happiness*. NY: Penguin Books.
- Tinbergen, N. (1989). *The study of instinct*. Oxford: Clarendon (Original work published 1951).
- Vavova , K. (2015). "Evolutionary Debunking of Moral Realism," *Philosophy Compass* 10/2: 104–116.
- Uttich, K., Tsai, G. and Lombrozo, T. (2014). "Exploring Metaethical Commitments: Moral Objectivity and Moral Progress," *Advances in Experimental Moral Psychology: Affect, Character, and Commitments*, Eds. Wright, J.C. & Sarkissian, H. New York: Continuum Press.
- Wallace, D (2007). 2007. "Deciderization 2007—A Special Report." In *The Best American Essays 2007*, ed. D. F. Wallace and R. Atwan. Boston: Mariner.
- Walsh, D. (1996). "Function and Fitness," *British Journal of Philosophy of Science*, 47: 553 – 574.
- Wiegman, I. (2015). "The Evolution of Retribution: Intuitions Undermined," *Pacific Philosophical Quarterly*, doi: 10.1111/papq.12083.
- Wielenberg, E. (2010). On the evolutionary debunking of morality. *Ethics*, 120, 441–464.
- Wright, J., Grandjean, P., and McWhite, C. (2013). "The meta-ethical grounding of our moral beliefs: Evidence for meta-ethical pluralism," *Philosophical Psychology*, 26(3): 336 - 361.