

ARTÍCULOS
DE INVESTIGACIÓN



Does autonomous moral reasoning favor consequentialism?*

Caner Turan

Tulane University, New Orleans, United States
Email: cturan@tulane.edu

Recibido: 30 de marzo de 2021 | Aceptado: 15 de junio 2021
<https://doi.org/10.17533/udea.ef.345775>

Abstract: This paper addresses an important issue that has been commonly debated in moral psychology, namely the normative and metaethical implications of our differing intuitive responses to morally indistinguishable dilemmas. The prominent example of the asymmetry in our responses is that people often intuitively accept *pulling a switch* and deny *pushing* as a morally permissible way of sacrificing an innocent person to save more innocent people. Joshua Greene traces our negative responses to actions involving “up close and personal” harm back to our evolutionary past and argues that this undermines the normative power of deontological judgments. I reject Greene’s argument by arguing that our theoretical moral intuitions, as opposed to concrete and mid-level ones, are independent of direct evolutionary influence because they are the product of autonomous (gene-independent) moral reasoning. I then explain how both consequentialist and deontological theoretical intuitions, which enable us to make important moral distinctions and grasp objective moral facts, are produced by the exercise of autonomous moral reasoning and the process of cultural evolution. My conclusion will be that Greene is not justified in his claim that deontology is normatively inferior to consequentialism.

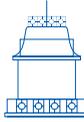
Keywords: metaethics, moral psychology, moral intuitions, trolley cases, moral reasoning, cultural evolution

* This work has been funded by Tulane University, US.

Cómo citar este artículo

Turan, C. (2022). Does autonomous moral reasoning favor consequentialism? *Estudios de Filosofía*, 65. 89-111. <https://doi.org/10.17533/udea.ef.345775>





¿Favorece el razonamiento moral autónomo el consecuencialismo?

Resumen: Este artículo aborda una cuestión importante que ha sido comúnmente debatida en la psicología moral, a saber, las implicaciones normativas y metaéticas de nuestras diferentes respuestas intuitivas a dilemas moralmente indistintos. El ejemplo más destacado de la asimetría en nuestras respuestas es que la gente suele aceptar intuitivamente *halar y niega empujar* como una forma moralmente permisible de sacrificar a un inocente para salvar a más personas inocentes. Joshua Greene atribuye nuestras respuestas negativas a las acciones que implican un daño “cercano y personal” a nuestro pasado evolutivo y argumenta que esto socava el poder normativo de los juicios deontológicos. Rechazo el argumento de Greene argumentando que nuestras intuiciones morales teóricas, a diferencia de las concretas y de nivel medio, son independientes de la influencia evolutiva directa porque son el producto de un razonamiento moral autónomo (independiente de los genes). A continuación, explico cómo las intuiciones teóricas consecuencialistas y deontológicas, que nos permiten hacer importantes distinciones morales y captar hechos morales objetivos, son producidas por el ejercicio del razonamiento moral autónomo y el proceso de evolución cultural. Mi conclusión será que Greene no está justificado en su afirmación de que la deontología es normativamente inferior al consecuencialismo.

Palabras clave: metaética, psicología moral, intuiciones morales, dilema del tranvía, razonamiento moral, evolución cultural

Caner Turan: is a doctoral candidate in the department of philosophy at Tulane University, writing his dissertation on Kantian constitutivism. He is the author of “Necessary constructivism in Kant’s moral theory” in *The Philosophy of Kant*, Ricardo Gutiérrez Aguilar Ed. (2019) and “Are ambitious evolutionary debunking arguments self-refuting?” (*Southwest Philosophical Studies*, 43, forthcoming).

ORCID: 0000-0002-4900-8578

An out-of-control trolley is speeding toward five workers who are not able to move out of the way in time. You see a switch connected to the tracks that, if pulled, will divert the trolley onto a sidetrack and kill a single worker who is standing there. Would you pull the switch to sacrifice one person to save five? Or would you push a large man over a footbridge into the path of the trolley to stop the trolley and save those workers? As widely known, the former is the trolley dilemma, and the latter is the footbridge dilemma.¹ What is philosophically interesting about these dilemmas is that people often react differently to them, even though they both involve sacrificing an innocent person to save more innocent people. Most people intuitively accept *pulling a switch* and deny *pushing* as a morally permissible way of sacrificing an innocent person to save more innocent people, when they are presented with the trolley and footbridge dilemmas (Greene, Sommerville, Nystrom, Darley & Cohen, 2001). What explains these different reactions?

Joshua Greene and his colleagues (2001; 2009) assert that most people find pushing the large man off the footbridge morally unacceptable because their evolutionary-based negative emotional reactions to actions involving “up close and personal” harm *distort* their judgments. Moreover, Greene (2008) holds that our negative reactions to actions involving personal harm result in deontological judgments. The idea is that since our deontological judgments have this epistemically suspect, evolutionary origin, they are *normatively inferior* to our consequentialist judgments, which arise from more rational thought processes such as cost-benefit analysis.² My aim in this paper is to show that Greene is not entitled to make a substantive normative distinction between consequentialism and deontology.

In the [first section](#), I lay out Greene’s main argument and the results of his initial study, and in the [second section](#), I consider an objection to Greene’s argument and present his revised position. I distinguish, in the [third](#) and [fourth sections](#), between different types of moral intuitions and claim that our theoretical moral intuitions, as opposed to concrete and mid-level ones, are independent of direct evolutionary influence because they are the product of autonomous (gene-independent) moral reasoning. Since both consequentialist *and* deontological theoretical intuitions are immune to nonmoral biases, it would be wrong to make a substantive normative distinction between the two theories. Finally, in the [fifth section](#), I describe *how* the exercise of moral reasoning and the process of cultural evolution could generate gene-independent consequentialist *and* deontological moral intuitions that allow us to grasp objective moral facts and distinctions. I conclude that (1) the theory I support offers a better explanation of our consequentialist and deontological responses to particular cases, and that (2) Greene

1 The trolley and footbridge dilemmas were introduced by Philippa Foot (1967) and Judith Jarvis Thomson (1976), respectively.

2 A deontological judgment roughly takes the form “X is wrong *per se*, independent of any consequences it may lead to,” whereas a consequentialist judgment takes the form “X is right because its outcome promotes aggregate happiness or pleasure.”

is not justified in his claim that deontology is normatively inferior to consequentialism.

1. Greene's argument

Greene claims that consequentialism is normatively superior to deontology.³ To bolster his claim, he aims to undermine the justification of our deontological beliefs by placing a focus on their evolutionary origins. That is, on his account, deontological judgments are typically fueled not by the rational pull of objective moral facts but by morally irrelevant emotional responses which evolved as a means of increasing biological fitness. These emotional responses, according to Greene, are more likely to occur when one finds oneself in a situation where there is a *personal harm rather than an impersonal one*.⁴ Greene's important claim is that whether harm is "up close and personal" or impersonal is morally irrelevant. This means, according to Greene, only the harm itself must have moral importance. Thus, strong negative responses to cases involving personal harm cannot be the result of the rational force of a deontological principle. Rather, they must be rooted in our evolutionary past: "These responses evolved as a means of regulating the behavior of creatures who are capable of intentionally harming one another, but whose survival depends on cooperation and individual restraint" (Greene, 2008, p. 43).

Greene bases his claim on the results of a study he and his colleagues conducted regarding responses to various moral scenarios, including the famous trolley dilemma. Greene et al.'s (2001) study shows that whether the harm is "up close and personal" affects people's moral judgments. In the study, subjects are presented with both personal and impersonal moral dilemmas and their brain activity is measured using fMRI when they respond. The results are interesting: (a) when subjects are presented with a personal dilemma like the footbridge case, higher activity in brain regions connected with emotional response is observed compared to their brain activity when presented with an impersonal dilemma like the trolley case; (b) people who, contrary to the majority opinion, decided to push the large man in the footbridge case took longer to finalize their decisions than people confronted with the trolley case. Greene infers two conclusions from these findings: (1) people's emotional responses to morally indistinguishable dilemmas differ according to the way harm is inflicted; (2) the longer response time shows

3 Normative superiority should be understood in terms of the *reliability* of our reasons for action. That is, consequentialism is normatively superior to deontology if our consequentialist beliefs are less epistemically suspect or distorted than deontological ones.

4 An action involves *personal* harm if it is "(i) likely to cause serious bodily harm, (ii) to a particular person, (iii) in such a way that the harm does not result from the deflection of an existing threat onto a different party" (Greene & Haidt, 2002, p. 519).

that the cognitive system in the subjects' brain *overrides* the emotional system and results in the decision to push the large man.⁵ Therefore, Greene and his colleagues think that “the crucial difference between the trolley dilemma and the footbridge dilemma lies in the latter’s tendency to engage people’s emotions in a way that the former does not” (Greene et al., 2001, p. 2106).

Greene thinks that consequentialism is normatively superior to deontology because he associates deontological judgments with the emotional system and consequentialist judgments with the cognitive system, and more importantly, he claims that the emotional system that produces deontological judgments is distorted by *morally irrelevant factors* such as the proximity of harm or physical contact. Deontologists usually concede the importance of producing best consequences, as they do in the trolley case. Nevertheless, they sometimes choose to place constraints on consequentialism on account of rights or intrinsic human dignity –constraints which apply in the footbridge case. Greene et al.’s study suggests that deontologists offer these constraints not due to the rational force of a deontological principle but due to their strong negative reaction to instances involving personal harm. Greene’s argument can be summarized as follows:

P1. The emotional system that induces deontological judgments is affected by features that render a moral dilemma personal.

P2. The features that render a moral dilemma personal are morally irrelevant.

C1. Therefore, the emotional system that induces deontological judgments is affected by morally irrelevant features (P1, P2; empirical conclusion).

P3. Moral judgments that are affected by morally irrelevant features do not reliably track moral facts.

P4. Deontological judgments do not reliably track moral facts (C1, P3).

P5. Moral judgments have a genuine rational or normative power only if they reliably track moral facts.

C2. Therefore, deontological judgments, in contrast to consequentialist judgments, lack a genuine rational or normative power (P4, P5; philosophical conclusion).

5 According to Greene, two psychological systems compete to produce a moral judgment: emotional and cognitive systems. The emotional system mostly involves quick, automatic, but not necessarily conscious responses. We mostly resort to our emotions to survive, for they are “very reliable, quick, and efficient responses to recurring situations” (Greene, 2008, p. 60). By contrast, the cognitive system involves processes such as “reasoning, planning, manipulating information in working memory, controlling impulses, and higher executive functions” (Greene, 2008, p. 40). These processes do not trigger particular behavioral responses like emotions. Rather, they are neutral since they take many factors into account and any change in these factors could lead to a different course of action.

2. Personal force

There is a quick objection to Greene's first premise. Suppose that we change the conditions in the footbridge case: as in the footbridge case, a large man is standing on a footbridge, but you are standing near a remote switch. If you pull the switch, you will knock the victim off the bridge by activating a trap door. Let's call this case *Remote Footbridge*. In Remote Footbridge, we are using someone as a mere means as we do in the footbridge case, but now the harm appears to be *impersonal*. If this variation does not change our negative reaction, then Greene is mistaken: the nature of our reaction to footbridge cases does not depend on whether the harm is personal. Rather, we have strong emotional responses to such cases because we have *good reasons* to believe that using an innocent person as a mere means is wrong.

Greene has an answer to this objection. In their 2009 study, Greene and his colleagues introduce variations of the footbridge case including Remote Footbridge. The aim of the study is to improve on the 2001 study by making pairwise comparisons between variations of the footbridge case and trying to find out which morally irrelevant factor we respond to. There are three morally irrelevant factors in the standard footbridge dilemma: (i) *spatial proximity*, (ii) *physical contact*, (iii) *personal force*. Personal force differs from physical contact in that personal force occurs "when the force that *directly* impacts the other is generated by the agent's muscles, as when one pushes another with one's hands or with a rigid object" (Greene, Cushman, Stewart, Lowenberg, Nystrom & Cohen, 2009, p. 365). Greene and his colleagues introduce three variations in addition to the standard case:

Remote Footbridge: Everything is the same except that you knock the victim off the bridge using a trap door and a remote switch. This action involves *none* of the morally irrelevant factors.

Footbridge Pole: Everything is the same except that you use a pole instead of your hands to push the victim. This action involves *spatial proximity* and *personal force*.

Footbridge Switch: Identical to the Remote Footbridge except that you and the switch are next to the victim. This action only involves *spatial proximity*.

Comparing Remote Footbridge to the Footbridge Switch isolates *spatial proximity*. Comparing Standard Footbridge to Footbridge Pole isolates *physical contact*. And comparing Footbridge Switch to Footbridge Pole isolates *personal force*.

The results reveal that there is *no* significant effect of *spatial proximity* (Remote Footbridge vs. Footbridge Switch) and *physical contact* (Standard Footbridge vs. Footbridge Pole) on our responses, but there is a significant effect of *personal force* (Footbridge Switch vs. Footbridge Pole). That is, we are inclined to find harmful actions involving personal force morally wrong. Greene thereby shows that we generally don't

react negatively to actions that don't involve personal force such as the actions in the Remote Footbridge and Footbridge Switch. The objection mentioned above, thus, collapses.

In his 2013 book *Moral Tribes*, Greene cites another evidence that could show that our gut reaction against actions involving personal force is in fact sensitive to the means/side-effect distinction. There is conflicting evidence on our sensitivity to that distinction. On the one hand, most people (81%) approve of harming someone using personal force if the harm occurs as a *side effect* rather than a means to an end (obstacle collide case). On the other hand, most people (81%) approve of harming someone as a means when there is no personal force (loop case). And most people (69%) disapprove of harming someone as a means when personal force is involved, as footbridge case shows. Greene's solution is what he calls the *modular myopia hypothesis*, according to which, our emotional system is *blind to harmful side effects* and detects only harmful actions that are intended as a means to some end. That is, most find it permissible to use personal force when the harm occurs as a side effect because they are emotionally (not cognitively) blind to the resulting harm. Greene thinks the means/side-effect distinction is morally irrelevant, and the reason we find it morally significant is that we first have these distorted intuitive judgments about the distinction, and we then imperfectly organize them into the Doctrine of Double Effect. So, Greene believes that the means/side-effect distinction is unjustified just like the judgment that pushing the large man in the footbridge case is impermissible.

3. Types of moral intuitions and autonomous reasoning

Does Greene's argument cast doubt on *all* deontological judgments? Michael Huemer's (2008) categorization of moral intuitions⁶ will help us to show that Greene is not entitled to associate selective pressures with our *theoretical* moral intuitions, be they consequentialist or deontological.⁷ On Huemer's account, there are three types of moral intuitions:

-
- 6 By intuition, I mean "immediate intellectual grasp or appearance prior to reasoning." There are roughly two ways of conceiving intuitions. We can see them as a species of belief with a distinct kind of justification. For example, once we understand the proposition "20 is less than 100", we are automatically justified in believing it. This doesn't require reasoning or inference but merely understanding the proposition. Or we can see intuitions as cognitive states different from beliefs. Considered in this way, an intuition takes the form "It seems to me that X," which is different from having a belief. Our beliefs may differ from our intuitions: a counterexample to an intuition or an evidence against it could make us disbelieve a certain intuition.
- 7 Huemer uses his categorization of moral intuitions to support his 'revisionary intuitionism,' while I use it against Greene's argument and reach different conclusions than Huemer. I also provide an account of how we acquire theoretical intuitions in the fifth section, which, I believe, is necessary to be able to classify them as 'intuitions.' I am not sure if Huemer would accept that account or my conclusions.

- (1) *Concrete intuitions*: Intuitions about particular cases such as trolley and footbridge dilemmas.
- (2) *Abstract (theoretical) intuitions*: Intuitions about general moral principles such as the consequentialist intuition that the right action is the one whose consequence promotes the overall happiness, or the deontological intuition that it is wrong to use people as mere means.
- (3) *Mid-level intuitions*: Intuitions about principles that have in-between level of generality such as the principle that one ought not to kill sentient beings for entertainment, or that one ought not to lie even if it is the only way to save one's life.

Huemer claims that concrete and mid-level intuitions are more likely to be responsive to evolutionary biases, whereas abstract or theoretical intuitions are probably not directly responsive to them. We can give two reasons for this conclusion: (1) concrete and mid-level intuitions are related to strong emotional reactions, and (2) evolutionary mechanisms select for certain types of behavior to promote biological fitness.

Consider a concrete intuition such as that “It is wrong that David had sex with his sister,” or a mid-level intuition that “It is not permissible to kill sentient beings for entertainment.” These intuitions are more likely to arouse strong emotional reactions associated with our evolutionary past than the theoretical intuition, say, “S is morally obligated to do x, only if S is capable of doing x.” Furthermore, evolution might have endowed us with intuitions that favor certain types of behavior that are more likely to increase our reproductive success. These could include, for example, intuitions about incest, infanticide, murder, and promiscuity, among many others. However, it is less likely that it has endowed us with intuitions that favor certain types of abstract principles. Admittedly, there are studies that indicate our ancestors’ involvement in moral reflection to some extent;⁸ nevertheless, it is less than obvious that there is a meaningful relation between theoretical intuitions and reproductive success. It is true that some of our shared evaluative dispositions may have arisen from properties conducive to survival and reproduction. For instance, properties such as fractal patterns and symmetry could be the evolutionary foundations for our appreciation of beauty.⁹ Theoretical intuitions, by contrast, do not seem to be very helpful in terms of survival and reproduction. Abstract moral principles (e.g., one should always maximize utility, or one should never use people as a mere means) are *too general* to qualify as useful guides to survive and reproduce. It is likely that we reach abstract moral principles through employment of autonomous reflection on our evolved reactive attitudes to particular cases.

8 Flack & de Waal (2000); de Waal (2014); Burkart, Brügger & van Schaik (2018).

9 Freeman, Graham & Emlen (1993); Jasienska, Lipson, Ellison, Thune & Ziomkiewicz (2006); Chatterjee (2014).

And over time, we convert these principles into intuitions and internalize them, as I describe it in the [fifth section](#).¹⁰

Is the distinction between different types of moral intuitions too vague? We arguably need specific criteria that an intuition must meet for us to classify it reliably as theoretical, mid-level or concrete. I propose *three* criteria which an intuition must meet to qualify as a theoretical intuition: (1) it does not arise from a specific set of environmental circumstances of time and place (e.g., trolley cases), (2) it does not require acting in a specific way (e.g., pulling a switch), and (3) it does not involve a particular person (e.g., large man). Let's look at some examples:

(D1) It is unfair when some people are worse off than others owing to differences in their unchosen circumstances.

(D2) It is our duty to fulfill a promise, regardless of the goodness of the consequences.

(D3) It is morally wrong to prosecute and punish those known to be innocent.¹¹

(C1) An action is morally good if it increases the overall level of pleasure in the world.

(C2) An action is morally bad if it increases the overall level of pain in the world.

While D1, D2, and D3 are deontological theoretical intuitions, C1 and C2 are consequentialist theoretical intuitions. These intuitions do not specify any circumstances, any particular action, or any particular individual. Thus, it seems that they are *not* generated directly by our emotional responses to particular cases but rather by autonomous moral reasoning.

What makes moral reasoning *autonomous or gene-independent*? One way to interpret the evolutionary influence on our capacity for morality and the content of our moral judgments is to claim that moral thought and behavior are the direct result of evolutionary mechanisms and have the sole function of adapting humans to their changing surroundings.¹² This strictly behavioristic and eliminativist view sees morality as a “certain type of behavioral pattern or habit, accompanied by some emotional responses” rather than as “a theoretical inquiry that can be approached by rational methods, and that has internal standards of justification and criticism” (Nagel, 1979, p. 142). The latter theoretical conception of morality does not deny the

10 One could say that our abstract deontological theories are ultimately rooted in our distorted concrete and mid-level deontological intuitions. That is, it could be that we accept the principle that one should not use someone as a mere means on the basis of intuitions such as the one we have about the original footbridge case. I will address this “garbage in, garbage out” problem in the following.

11 This is Shafer-Landau's (2003) example.

12 For example, Michael Ruse's Darwinism says that “substantive [or objective] morality is a kind of illusion, put in place by our genes, in order to make us good social cooperators” (Ruse, 2010, p. 309).

fact that we and some species of nonhuman animals share certain basic psychological dispositions and emotional responses that have been shaped by evolutionary forces. Its distinctive claim is that we also possess a superior intelligence that allows us to evaluate, systematize, and occasionally say no to these pre-reflective dispositions and emotional responses.¹³ That is, we don't automatically act on our adaptive dispositions or emotions but we also think about them, evaluate them, try to justify them by appealing to reasons, decide between them, and guide our actions in line with the judgments reached through rational reflection. Moral reasoning is autonomous, then, in the sense that moral thought is not significantly determined by specific, evolutionary-based, psychological dispositions. For example, it may turn out that dispositions that promote slavery, racism, and condemnation of homosexuality are adaptive. But after exercising moral reasoning, revising our judgments about these cases, and passing down our revised judgments to subsequent generations, many people have come to suspect that beliefs that support slavery, racism, and discrimination against LGBTQ+ people are unjustified, and accordingly they have been trying to act against these adaptive dispositions. The fact that we have the uniquely human capacity to criticize and revise our adaptive responses supports the idea that human morality is not just a totality of behavioral patterns but also a theoretical inquiry.

The fact that pre-reflective, adaptive dispositions cannot determine the content of certain mental activities points to the human capacity for autonomous (gene-independent) reasoning. For example, it would be a mistake to explain our mathematical or scientific judgments only by appeal to their psychological origins. The *reasons* behind those judgments should also be taken into account. We can justify our mathematical or scientific beliefs only when they correspond to mathematical or scientific *facts*, some of which are independent of how evolutionary mechanisms work. For instance, truths of the propositions “ $1+1=2$ ” and “Water boils at 100°C ” are independent of how organisms evolve and what kind of basic dispositions they possess. Although it may be true that our general disposition to become involved in mathematics and science is an adaptation, it is less likely that evolutionary influence is so pervasive as to determine the content of these activities. This indicates a shift in the function of our intellectual capacities: they do not only facilitate survival and reproduction, but they also help us track the truth of certain objective facts that go beyond the workings of evolutionary mechanisms.

13 As a matter of fact, some evolutionary biologists and philosophers maintain that human moral behavior is not causally related to prosocial behavior of nonhuman animals. Rather, moral behavior amounts to *rational* behavior in our case because our capacity to make moral judgments is the product of our advanced intellectual powers such as *self-awareness* and *abstract thinking* (Ayala, 2016; Prinz, 2008). In other words, since human moral behavior involves *evaluation* and *reasoning*, which requires those advanced intellectual capacities, it is *not* biologically determined like altruistic behavior of nonhuman animals.

Furthermore, even though moral codes must be consistent with our biological nature,¹⁴ “moral norms are independent of [biologically conditioned human] behaviors in the sense that some norms may not favor and may hinder [reproductive success]” (Ayala, 2016, p. 245). For example, religion and patriotism facilitate cooperation but they also give rise to beliefs about racism and genocide, which may act against biological fitness. Obedience is considered as being conducive to survival (Darwin, 1981) but it can turn into mass killings in the wrong hands. The rapid decline in birth rates in some areas of Italy in the nineteenth century is another example of human values going beyond evolutionary aims (Cavalli-Sforza & Feldman, 1981). All of these examples point to the fact that *some* of the content of our moral beliefs are *exaptations* rather than *adaptations*.¹⁵ Thus, we have good reasons to claim that human moral behavior differs from altruistic behavior of nonhuman animals due to the distinct human capacity for *reasoning* and *evaluation*, along with the effect of cultural transmission.¹⁶ Greene does not deny that moral reasoning is autonomous. Rather, on his account, autonomous moral reasoning favors consequentialism because he believes that his studies indicate that deontological judgments, unlike consequentialist ones, are shaped by our evolutionary-based emotional reactions to morally irrelevant factors. My point is that the nature of our consequentialist *and* deontological theoretical moral intuitions indicates that they *both* are the result of *autonomous* application of human intelligence in moral thinking.

Turning back to theoretical moral intuitions, one important objection to the contrast between emotional responses to particular cases and reasoning about abstract theories is the claim that *all* abstract moral theories ultimately rest upon our responses to particular cases; thus, they are subject to evolutionary influence as well. That is, we first have emotional reactions to particular cases. We then reflect on these reactions and come up with abstract moral theories. Since these abstract theories originate from our evolved reactive attitudes, they also must have been shaped by natural selection. This is the GIGO problem,¹⁷ but it applies to both consequentialist and deontological theoretical intuitions. *If* deontological theoretical intuitions are epistemically suspect because they are distorted by our evolved attitudes to particular cases, so are consequentialist theoretical intuitions because they arise from our attitudes to particular cases as well. For instance, it is highly unlikely that we arrived at C1 and C2 without reflecting on

14 Otherwise, our species would cease to exist. For example, if killing were the norm, there would ultimately be no human being to think about morality.

15 Exaptations are shifts in the function of a trait in the course of evolution, such as plumage of birds. See Gould & Vrba (1982) and Ayala (2010).

16 I discuss the effect of cultural transmission in the fifth section.

17 The idea is that if our basic psychological dispositions are heavily shaped by natural selection, and if our moral reasoning stems from those basic dispositions, then our moral reasoning must also have been shaped by natural selection. Greene (2008) calls it “the GIGO problem: garbage in, garbage out” (p. 119). Street (2006) and Kahane (2011) make the same point.

our evolved attitudes to particular cases involving specific kinds of pleasure and pain. This is one side of the coin: a possible story about how we reach moral judgments. The other side of the coin involves another equally plausible story: the fact that particular cases arouse emotional reactions does not tell us whether the actions in these cases are “*morally right, wrong, or neutral*”. To decide that, we employ autonomous moral reasoning and decide whether our reactions are *appropriate*. In cases like condemnation of homosexuality, we have come to suspect that our reactions are not appropriate, whereas in cases like looking after our children we believe that our reactive attitudes are on the right track. What I claim is *not* that the former story is absolutely wrong, and the latter is absolutely correct. Although I favor the latter story in the rest of this paper, my point in this paragraph is that the truth of *either* story would go against Greene’s claim that consequentialism is normatively superior to deontology.

The only weakness of our theoretical intuitions seems to be that they are vulnerable to exceptions. This means one could always come up with a counterexample to abstract moral principles, and once a counterexample is found the intuition at issue loses its initial credibility. This occurs quite often in ethical discussions. Deontological ethics is often criticized for being *too restrictive*. For instance, if lying amounts to using another rational being as a mere means, then lying must be forbidden in *all* circumstances, including the one in which telling the truth gives rise to death of your children. Once this counterexample is raised, we develop a more critical attitude towards the theoretical intuition that we should never use people as a mere means. Consequentialism, on the other hand, is often criticized for being *too demanding*. This is because consequentialism requires us to give away almost all our wealth to the poor, since it will increase the overall happiness.¹⁸ For many, this goes above and beyond what morality requires of us. Such counterexamples pose a challenge to generalizations that are reflected in our different consequentialist or deontological theoretical intuitions, making it quite difficult to construct a coherent moral system. They encourage revision not only in our theoretical intuitions but also in our mid-level and concrete intuitions.

4. Formal intuitions

I have argued that while concrete and mid-level intuitions are susceptible to nonmoral biases, theoretical intuitions seem to be immune to them. At the same time, theoretical intuitions of both consequentialism and deontology are vulnerable to exceptions. Should we now conclude that our moral intuitions are never trustworthy as a guide to practical reasoning? Is there no way forward? Huemer proposes that our most trustworthy intuitions are *formal intuitions*, which are a subset of theoretical intuitions. The function

¹⁸ See Singer (1972).

of formal intuitions is not to make any moral evaluation but to place formal constraints on moral theories. Consider three of the examples Huemer (2008) gives:

- (a) If x is better than y and y is better than z , then x is better than z .
- (b) If it is wrong to do x , and it is wrong to do y , then it is wrong to do both x and y .
- (c) If two states of affairs, x and y , are so related that y can be produced by adding something valuable to x , without creating anything bad, lowering the value of anything in x , or removing anything of value from x , then y is better than x (p. 386).

Formal intuitions, according to Huemer, are products of *reflection* upon what is required by the *nature* of the ‘better than’ relation, wrongness, moral evaluation, permissibility, and so on. While we arrive at non-formal theoretical intuitions first by reflecting on particular cases and then reaching a general conclusion, formal theoretical intuitions are generated by reasoning about what is *entailed* by the *nature* of moral predicates such as ‘better than’ or moral evaluation in general.¹⁹ In other words, the reasoning from particular cases to non-formal theoretical intuitions is *inductive*, whereas the reasoning from formal (theoretical) intuitions to particular moral facts is *deductive*. We can liken formal intuitions to axioms in geometry: we derive particular moral facts from ethical principles like the principle of transitivity of ‘better than.’ Since formal intuitions follow directly from the nature of evaluative/moral concepts, they do not arise from observation of particular cases. Therefore, formal intuitions are less likely to be affected by nonmoral, evolutionary biases associated with reactive attitudes to particular cases, and they also do not seem to be vulnerable to exceptions.

When a counterexample is found to a theoretical intuition, the intuition loses its initial credibility, which could even result in a rejection of it. However, when one comes up with a counterexample to a formal intuition, we generally call the case a *paradox* instead of giving up the intuition altogether. For example, take Derek Parfit’s (1984) famous ‘repugnant conclusion:’

- A: 100 people live a very high quality of life.
- B: 200 people live a slightly lower quality of life than the people in A.
- C: 400 people live a slightly lower quality of life than the people in B.
- ...
- Z: 100.2²⁵ people live in conditions barely worth living (pp. 419–430).

19 It is important to note that formal intuitions are not logical laws but principles of ethics. For example, if A punches B and B punches C , it doesn’t follow that A punches C . The schema “ $aRb, bRc, \text{ therefore } aRc$ ” is not a valid inference form in general, since it depends on the *nature* of the relation R . Transitivity principles are principles of whatever subject matter relation R belongs to, not principles of logic.

Most people have the intuition that “B is better than A,” “C is better than B,” and so on. And if the principle of the transitivity of ‘better than’ is to hold for our intuitions, then we must expect to have the intuition that “Z is better than A.” Despite that, most people have the intuition that “A is better than Z.” We seem to have three options to solve this paradox: (1) reject the transitivity of the ‘better than’ relation, (2) reject some or all of our earlier intuitions (e.g., B is better than A), or (3) accept the “repugnant” conclusion. Denying the principle of the transitivity of ‘better than’ would block the reasoning leading us to the repugnant conclusion at the cost of giving up one of the least controversial evaluative principles. Temkin (1987; 2012) and Persson (2004) espouse such a radical approach. Temkin (2012), for example, denies what he calls the ‘Internal Aspects View’ in favor of the ‘Essentially Comparative View’ to reject the principle of transitivity of ‘better than.’ Still, there are various ways of coming to terms with the repugnant conclusion.²⁰ My point here is not that we should accept the repugnant conclusion. Regardless of who is right on this matter, rejecting the transitivity of ‘better than’ is clearly a radical step that challenges one of our most pervasive evaluative principles. My point is that counterexamples to formal intuitions yield paradoxes, whereas counterexamples to non-formal theoretical intuitions do not. Moreover, one plausible way to solve paradoxes generated by formal intuitions is to reject otherwise appealing concrete intuitions in favor of preserving formal ones. Formal intuitions are, therefore, helpful in assessing and—at least sometimes—revising our moral theories.

The formal intuition (c) above seems to be a suitable candidate for a consequentialist formal intuition. Are there any deontological formal intuitions? One possible candidate is Kant’s first formulation of the Categorical Imperative (also known as the universalizability principle): “Act only in accordance with that maxim through which you can at the same time will that it become a universal law” (Kant 1998, p.31). In other words, your maxim (or the subjective principle that motivates you to perform a certain act) can qualify as a moral principle if and only if it does not lead to a contradiction when raised to the level of a universal law. For example, if everyone were to act on the subjective principle that “lying is permissible if you can get away with it” with the same regularity of laws of nature, that would lead to a contradiction. This is because the assumption of truth-telling is a necessary condition for lying. Since nobody will assume that anyone is telling the truth, it is impossible for anyone to lie under that maxim. In such cases, we think that a particular consideration gives rise to a paradox

20 Huemer (2013) claims that the ‘Money Pump Argument’ shows that intransitive betterness judgments give rise to paradoxical results. That is, in Huemer’s view, Temkin tries to solve a paradox by creating another paradox. Moreover, it may be true that the repugnant conclusion is *not* repugnant after all: why would it be repugnant to say that Z is better than A, considering that Z includes more of what is good? Since we compare the *total* quality of life, there seems to be nothing wrong with the claim that Z is better than A. There are also other ways to accept the repugnant conclusion. Tännsjö (2002) denies that the repugnant conclusion is counterintuitive by appealing to the human capacity to adapt to new conditions. Arrhenius (2000) accepts the repugnant conclusion by denying that there are any acceptable alternatives to the repugnant conclusion. For reasons why we should not try to avoid the repugnant conclusion, see also “What should we agree on about the repugnant conclusion?” *Utilitas* (2021, pp. 1–5).

about a formal intuition rather than giving up the intuition itself (the law of non-contradiction in Kant's case).²¹ Even if we accept that concrete and mid-level intuitions are unreliable due to the distortion created by nonmoral factors, we cannot reach the same conclusion for Kant's first formulation of the Categorical Imperative because it is a purely formal principle without any experiential content. Thus, it is not directly susceptible to evolutionary influences. Kant's universalizability principle itself may have nothing to do with our evolutionary history or emotions. It is devoid of any empirical content, yet it is the foundation of Kant's deontological moral theory. Emotions and evolutionary influences could come into play only when we try to derive concrete moral duties (e.g., "You ought not to commit suicide") from that formal principle.²²

It is possible that there exists both consequentialist and deontological formal intuitions that are safe from morally irrelevant factors and exceptions. Can we use these formal intuitions to make substantive normative judgments? This is the real difficulty that faces both consequentialists and deontologists when they try to prescind from the level of purely formal intuitions to particular cases. It seems that different moral theories can have different secure formal intuitions. However, once they try to fill the gap between formal intuitions and particular cases, their theories become vulnerable to exceptions and nonmoral factors. The problem that confronts all such theories is, thus, to find a way to translate formal intuitions into action-guiding general principles

-
- 21 I admit that declaring the first formulation of the Categorical Imperative to be a *deontological* formal intuition appears to be hasty. One might argue that the universalizability principle, on its face, is neutral between consequentialism and deontology. Surely, there is a 'universalizability' requirement built into the logic of moral judging, but that requirement does not seem to rule out consequentialism: the impossibility, logical or otherwise, of everyone's doing what I want to do is neither necessary nor sufficient to show that what I want to do is impermissible. So, the question is, how is it possible for a purely formal principle to produce substantive moral content? This, I think, is a legitimate concern; however, there are possible solutions. For instance, according to Sensen's (2014) interpretation of Kant's universalizability principle, the principle is recognizably moral because it represents the idea of *fairness*. That is, the universalizability principle forbids us from seeing ourselves as superior to others or making an *exception* for ourselves to the laws we regard as objectively necessary. This is Kant's *formal* principle. But we should resort to empirical sciences to discover these objective and necessary laws and to derive concrete duties from that formal principle. For example, we regard the law against killing objectively necessary because following it is necessary for us to be able to live at all. Or we regard the law of helping others as objectively necessary because it is a *necessary* means to an end *all* of us have, i.e., we will necessarily need help at some point in our lives as finite beings. It is to these laws that we should not make an exception.
- 22 One could also say that Kant's universalizability principle does not rule out consequentialism because consequentialism could be a universal law in the sense that everyone could follow the rule, say, "Always maximize utility." However, it is worth noting that what we are assessing is not *actions* themselves but *maxims*, namely the *reasons* behind actions. For example, take the consequentialist maxim "*Whenever X leads to a better consequence, do X.*" Substitute X with 'lying,' 'committing suicide,' 'committing murder,' 'not looking after one's children,' and so on, and suppose that these subjective principles become necessary and universal like laws of nature. If everyone were to act on these maxims with the same regularity of laws of nature, then either there would be *no* consequence to be improved (because everyone would die soon or older members of society would not be replaced), or it would ultimately become *impossible* for any of the mentioned actions to lead to any better consequence. Although I admit that more work has to be done to be able to claim that the universalizability principle and consequentialism are mutually exclusive, this is roughly how the universalizability principle could rule out consequentialism. In any case, even if there were no deontological formal intuitions, the existence of deontological—and non-formal— theoretical intuitions would suffice to prove the claim that some of our consequentialist *and* deontological intuitions are immune to direct evolutionary influence.

with content, without falling prey to exceptions and morally irrelevant biases. Although Greene successfully shows us the way in which *some* of our concrete and mid-level deontological intuitions *are* susceptible to distortions that arise from evolutionary forces in *some* moral dilemmas, his evidence does not seem to apply to (non-formal) theoretical and formal (theoretical) intuitions. Hence, he is not entitled to claim that consequentialism is normatively superior to deontology.

5. How to acquire theoretical intuitions

I have argued that while our concrete and mid-level intuitions are susceptible to evolutionary influence, our theoretical intuitions (formal ones included) often aren't directly susceptible to them. But are theoretical intuitions really "intuitions"? Or, are they generalizations – convictions which we develop only *after we reflect on* particular cases? Since intuitions are "*immediate intellectual grasps or appearances prior to reasoning,*" it may seem to be wrong to classify them as intuitions. In this section, I describe how we could turn abstract moral theories into intuitions.

Actions arouse emotions. Due to our shared biological nature, we are inclined to react in particular ways to particular actions. These reactive attitudes are the products of our '*concrete intuitions.*' For example, when we hear about an instance of incest, or someone beating their children, or someone cheating on their partner, our concrete intuitions make us feel that they are wrong, and we react accordingly. Then we use our capacity for abstraction (the capacity to see actions or objects as members of general categories) and find certain *kinds* of actions right or wrong (e.g., "Incest is wrong"). We call such reactions '*mid-level intuitions.*' Mid-level intuitions are similar to what Haidt (2001, p. 828) calls "*post hoc rationalizations*": They appear to originate from moral reasoning, but they are often merely expressions of our emotions (plus abstraction).²³ Then we systematically think about our concrete and mid-level intuitions and question their appropriateness, credibility, coherence with each other, and so on. We reflect on various circumstances we might find ourselves in and whether different conditions affect the rightness/wrongness of particular actions. We also try to see whether our concrete and mid-level intuitions check with our formal intuitions. After reflecting systematically on our concrete and mid-level intuitions, we reach generalizations or abstract moral theories. And when we can think of exceptions or counterexamples to our moral theories, we either reject them completely or revise them. This is roughly the way we exercise our capacity for autonomous moral reasoning. But it would be quite difficult –if not impossible– and ineffective to stop what one is doing each time

23 Our capacity for abstraction differs from our capacity to make generalizations. While the former capacity does not involve any inference (e.g., mid-level intuitions), the latter capacity is exercised through drawing inferences (e.g., theoretical intuitions).

and repeat this process for each action. We certainly need context-sensitive intuitions to effectively guide our actions. Can our generalizations or moral theories convert to intuitions? And if so, how?

To explain how we turn moral theories into “immediate intellectual grasps or appearances”, we need to focus on Greene’s distinction between the cognitive and the emotional. Assuming a sharp distinction between the emotional and the (dispassionately) cognitive is too hasty, of course. A *purely* cognitive or emotional judgment could be an impossibility due to the extremely complex nature of the brain. What we call ‘the cognitive’ could involve emotions, or what we call ‘the emotional’ could involve cognitive elements (Greene, 2008; Railton, 2014). However, it is plausible that we are *functionally* divided into two systems: emotional and cognitive systems.

The distinction between the emotional and the cognitive has been popularized by Daniel Kahneman, who calls them “System 1” and “System 2”, respectively (Kahneman, 2011). *System 1* involves automatic, effortless, and unconscious reactions to particular cases. It is like an autopilot, a stranger within, that “operates automatically and quickly, with little or no effort and no sense of voluntary control” (Kahneman, 2011, p. 20). *System 1* includes our innate, evolved, psychological dispositions that we share with other animals. It is our fast, effortless, and mostly unconscious mental system for jumping to conclusions to increase our chances of survival. Activities that are linked with *System 1* include “making a ‘disgust face’ when shown a horrible picture”, “understanding simple sentences”, “driving a car on an empty road”, and so on. By contrast, *System 2* “allocates attention to the effortful mental activities that demand it, including complex computations” (Kahneman, 2011, p. 21). It is associated with deliberate, logical, and slow reasoning that requires a lot of energy. *System 2* is an expert at solving problems, but it has limited capacity to exert effort. This is because we are not machines: it is difficult for us to maintain our focus for a long time and on different things at the same time. Activities that are linked with *System 2* include “searching memory to identify a surprising sound”, “parking in a narrow space”, “filling out a tax form”, “checking the validity of a complex logical argument”, and so on. Edward Slingerland (2014) uses the term “*hot cognition*” for *System 1* and “*cold cognition*” for *System 2*. He identifies hot cognition with “knowing how” and cold cognition with “knowing that.”

Now I can explain how we turn our moral theories into intuitions. Our Pleistocene ancestors used to live in small groups of hunter-gatherers, and they interacted mainly with relatives or people well known to them. Under these conditions, they developed psychological adaptations to facilitate and maintain cooperation such as emotions (e.g. empathy, resentment, etc.), the ability to recognize and remember faces, detect cheaters, and so on. We share these hot cognition processes with some of nonhuman animals. The important question is, “how one particular primate, us, managed the abrupt transition from our ancient hunter-gatherer lifestyle to the large-scale, urban way of life made possible by agriculture” (Slingerland, 2014, p. 175). In other words, how has it

become possible for our *hominin* ancestors, equipped only with hot cognitions, to adapt to urban life? It is less likely that hot cognitions are responsible for this adaptation because the time scale is too short for the evolution of new and complex psychological dispositions. One idea is that the abrupt transition became possible because of the introduction of new, external social institutions, which are designed to keep hot cognitions in check. However, given that cold cognition has a limited capacity to exert effort, it doesn't seem to be well-equipped for this task. Therefore, it is more likely that the transition to dense urban life "was managed not by consciously suppressing our tribal emotions but by using cold cognition to extend or redirect instincts through a process of *emotional education*" (Slingerland, 2014, p. 176; **emphasis added**). That is, instead of trying constantly to keep our pre-reflective intuitions in check, we create a set of shared values, *some* of which are the product of autonomous reasoning, and, more importantly, *we internalize* them: they become second nature to us.

The role of cultural evolution in achieving this is immense: cultural inculcation can train us to internalize the input from autonomous reasoning and react accordingly to particular cases so rapidly (even within a single generation) that biological evolution cannot be responsible for it.²⁴ Cultural evolution is "a distinctive human mode of *evolution* that has surpassed the biological mode because it is a more effective form of *adaptation*; it is faster than biological evolution, and it can be directed" (Ayala, 2016, p. 252). We call the effect of culture on transforming our ideas and beliefs 'cultural evolution' due to the striking resemblance between evolution of ideas/beliefs and biological evolution. First, there is *heredity*: ideas/beliefs are transmitted from one individual to another or from one society to another. Second, there is *variation*: beliefs, ideas, and values differ among individuals and societies. And third, there is *differential reproduction*: some ideas/beliefs are transmitted more efficiently than others among individuals or groups.

Just as some biological traits are favored by natural selection, so some ideas/beliefs are favored by *cultural selection*. Cultural ideas/beliefs compete for our attention and some of them are more successful in replicating themselves than others. What could be the reason for such a success? *Coercion* could be a reason. For example, after the Roman emperor Constantine the Great became a Christian in 312 CE, most of the Western world suddenly became Christian (Jones, 1948). *Psychological attractiveness* could be another reason. For a long time and in most cultures, people have believed that God exists. The reason for this success could be that it gives people security, love, and tranquility, i.e., the idea of God is psychologically appealing.²⁵ However, some of

24 The fact that "[w]e have no evidence for biological change in brain size or structure since *Homo sapiens* appeared in the fossil record some fifty thousand years ago" suggests that "[a]ll that we have done since then [...] is the product of cultural evolution" (Gould, 1996, p. 354).

25 This idea is supported by Dawkins' (2006) meme theory, according to which the survival of a cultural idea (or a 'meme') depends on the level of its psychological appeal.

the cultural ideas/beliefs' reproductive success is independent of biological fitness. We often opt for certain beliefs, not because they are psychologically attractive but because they *track the truth*. For example, the reason why no reasonable person believes that the geocentric model of the universe is correct is *not* because the belief in question is psychologically unappealing but because it does not track the truth. The capacity for autonomous reasoning creates a gap between biological and cultural evolution by enabling us to create cultural mutations that track the truth, regardless of their biological advantages.

Cultural evolution enables us to develop ideas and find solutions to long-term problems much faster and more efficiently than biological evolution, because we can transmit our ideas, beliefs, and values not only to our children but also to all humankind. However, our ability to use cultural information is restricted because cold cognition is slow, effortful, costly, and limited. The solution is *domestication* or *internalization*: just as we can deliberately change the behavior of animals and plants in line with our needs, so “the conscious mind can acquire new, desirable goals and then download them onto the unconscious self, where they can then be turned into habits and implemented without the need for constant monitoring” (Slingerland, 2014, p. 65). For example, when beginner drivers learn how to drive, prefrontal regions in their brains become much more active because they constantly try to maintain their focus to understand the traffic rules, how to turn the wheel, how to press on the pedals, and so on. However, after certain amount of practice, the nervous driver starts talking and making jokes while driving because their brain activity decreases. As they *internalize* driving skills, their conscious mind or cold cognition becomes less active. Likewise, beginner chess players have to think a lot about how pieces move, how to devise strategies to win, and so on. But over time, thanks to the human capacity to recognize patterns, they develop intuitions that allow them to make quick and efficient decisions without having to think or calculate for too long. (Especially in bullet chess games, where each player is given only a minute in total, players mostly play by their ‘chess intuitions’ instead of active thinking). In both examples, cold cognitions are transformed into hot cognitions over time.

In similar fashion, we turn our abstract moral theories into theoretical intuitions over time. Although we arrive at moral theories through conscious deliberation, we internalize them later. And once we turn our moral theories into hot cognitions, they start informing our concrete intuitions and affect our reactions to particular cases. That is, our theoretical intuitions help us grasp moral facts and distinctions and react to moral cases *immediately*, without thinking much about them, like a chess grandmaster playing a bullet game.²⁶ Admittedly, chess intuitions might go wrong

26 One may claim that insofar as an attitude, such as accepting a theory, was acquired through an inference, it will never become an intuition, regardless of whether it generates automatic judgments. To have an intuition is to have a mental state in which something seems true to you before you advance an argument for it. If it is possible to extend and redirect instincts and dispositions (e.g., we have redirected compassion from our relatives towards strangers; some have redirected disgust towards fatty and sugary

because they are not perfectly reliable in grasping chess facts. Chess grandmasters usually analyze games they played to improve their skills and to learn from their mistakes. And once a better move or strategy is discovered, the chess grandmaster revises or gives up their chess intuitions. Similarly, our theoretical intuitions are *not perfectly reliable* in grasping moral facts and distinctions. Therefore, when a counterexample is raised, we question the credibility of our theoretical intuitions; we either revise them, or we give them up.

6. Concluding remarks

I have argued that our theoretical moral intuitions are immune to direct evolutionary influence because they are the products of reflection on pre-reflective evolved psychological dispositions and the process of cultural evolution. Since we have both consequentialist and deontological theoretical intuitions, Greene is not justified in his claim that deontology is normatively inferior to consequentialism. Moreover, the theory I support has wider benefits. I have offered a model of how to respond to a great many evolutionary debunking arguments (EDAs). EDAs attempt to undermine the justification of our evaluative/moral beliefs by asserting that there is a *pervasive* evolutionary influence on the content of our evaluative/moral beliefs.²⁷ If my discussion of the different types of moral intuitions gets things right, then we cannot talk about a *pervasive* evolutionary influence because *some* of the content of our evaluative/moral beliefs is determined by autonomous application of human intelligence in moral thinking. I leave the task of debunking EDAs elsewhere to do justice to their complexities.

References

- Arrhenius, G. (2000). An impossibility theorem for welfarist axiologies. *Economics and Philosophy*, 16(2), 247–266. <https://doi.org/10.1017/S0266267100000249>
- Ayala, F. J. (2010). The difference of being human: morality. *Proceedings of the National Academy of Sciences*, 107(S2), 9015–9022. <https://doi.org/10.1073/pnas.0914616107>

food, even though evolution has shaped our taste so that we judge such food to be tasty), it may well be possible to redirect our intuitions from particular cases towards abstract moral theories. A moral theory may seem true to us (or we may immediately become aware of its truth) upon reflecting on it. It seems to me that we *do* have such intuitions about moral theories, and we do not necessarily think about particular cases each time to have them. Thus, I believe emotional education remains to be a plausible story.

- 27 There are different kinds of EDAs with different targets. An EDA can target all evaluative beliefs (Street, 2006), all moral beliefs (Joyce, 2006), or a certain subset of moral beliefs (Greene, 2008). However, all EDAs seem to share the assumption that evolutionary pressures have a *widespread* influence on targeted beliefs. This is the *empirical* premise that all EDAs seem to share. The inductive strength of an EDA depends on the truth of its empirical premise.

- Ayala, F. J. (2016). *Evolution, explanation, ethics and aesthetics: towards a philosophy of biology*. Academic Press. <https://doi.org/10.1016/B978-0-12-803693-8.00013-X>
- Burkart, J. M., Brügger, R. K. & van Schaik, C. P. (2018). Evolutionary origins of morality: insights from non-human primates. *Frontiers in Sociology*, 3(17), 1–12. <https://doi.org/10.3389/fsoc.2018.00017>
- Cavalli-Sforza, L. L. & Feldman, M. W. (1981). *Cultural transmission and evolution*. Princeton University Press. <https://doi.org/10.1515/9780691209357>
- Chatterjee, A. (2014). *The aesthetic brain: how we evolved to desire beauty and enjoy art*. Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199811809.001.0001>
- Darwin, C. (1981). *The descent of man and selection in relation to sex*. Princeton University Press. <https://doi.org/10.5962/bhl.title.70891>
- Dawkins, R. (2006). *The selfish gene*. Oxford University Press.
- Flack, J. C. & de Waal F. B. M. (2000). Any animal whatever: darwinian building blocks of morality in monkeys and apes. *Journal of Consciousness Studies*, 7(1–2), 1–29.
- Foot, P. (1967). The problem of abortion and the doctrine of double effect. *Oxford Review*, 1(5), 5–15.
- Freeman, D., Graham, J. & Emlen, J. (1993). Developmental stability in plants: symmetries, stress and epigenesis. *Genetica*, 89(1), 97–119. <https://doi.org/10.1007/BF02424508>
- Gould, S. J. (1996). *The mismeasure of man*. W. W. Norton & Company.
- Gould, S. J. & Vrba, E. S. (1982). Exaptation—a missing term in the science of form. *Paleobiology*, 8(1), 4–15. <https://doi.org/10.1017/S0094837300004310>
- Greene, J. (2008). The secret joke of Kant's soul. In W. Sinnott-Armstrong (Ed.), *Moral psychology*, Vol. 3, (pp. 35–79). MIT Press.
- Greene, J., Cushman, F. A., Stewart, L. E., Lowenberg, K., Nystrom, L. E. & Cohen J. D. (2009). Pushing moral buttons: the interaction between personal force and intention in moral judgment. *Cognition*, 111(3), 364–371. <https://doi.org/10.1016/j.cognition.2009.02.001>
- Greene, J. (2013). *Moral tribes*. The Penguin Press.
- Greene, J. & Haidt, J. (2002). How does moral judgment work? *Trends in Cognitive Sciences*, 6(12), 517–523. [https://doi.org/10.1016/S1364-6613\(02\)02011-9](https://doi.org/10.1016/S1364-6613(02)02011-9)
- Greene, J., Sommerville, B., Nystrom, L. E., Darley, J. M. & Cohen, J. D. (2001). An fMRI investigation of emotional engagement in moral judgment. *Science*, 293(5537), 2105–2108. <https://doi.org/10.1126/science.1062872>
- Haidt, J. (2001). The emotional dog and its rational tail: a social intuitionist approach to moral judgment. *Psychological Review*, 108(4), 814–834. <https://doi.org/10.1037/0033-295X.108.4.814>
- Huemer, M. (2008). Revisionary intuitionism. *Social Philosophy and Policy*, 25(1), 368–392. <https://doi.org/10.1017/S026505250808014X>

- Huemer, M. (2013). Transitivity, comparative value, and the methods of ethics. *Ethics*, 123(2), 318–345. <https://doi.org/10.1086/668905>
- Jasienska, G., Lipson, S. F., Ellison, P. T., Thune, I. & Ziolkiewicz, A. (2006). Symmetrical women have higher potential fertility. *Evolution and Human Behavior*, 27(5), 390–400. <https://doi.org/10.1016/j.evolhumbehav.2006.01.001>
- Jones, A. H. M. (1948). *Constantine and the conversion of Europe*. Hodder & Stoughton.
- Joyce, R. (2006). *The evolution of morality*. MIT Press. <https://doi.org/10.7551/mitpress/2880.001.0001>
- Kahane, G. (2011). Evolutionary debunking arguments. *Noûs*, 45(1), 103–125. <https://doi.org/10.1111/j.1468-0068.2010.00770.x>
- Kahneman, D. (2011). *Thinking, fast and slow*. Farrar, Straus and Giroux.
- Kant, I. (1998). *Groundwork of the metaphysics of morals*. (M. Gregor, Ed. & Trans.). Cambridge University Press. <https://doi.org/10.1017/CBO9780511809590>
- Nagel, T. (1979). Ethics without biology. *Mortal Questions* (pp. 142–146). Cambridge University Press. <https://doi.org/10.1017/CBO9781107341050.012>
- Parfit, D. (1984). *Reasons and persons*. Clarendon Press.
- Parker, S. T. & Gibson, K. R. (1990). “Language” and intelligence in monkeys and apes: comparative developmental perspectives. Cambridge University Press. <https://doi.org/10.1017/CBO9780511665486>
- Persson, I. (2004). The root of the repugnant conclusion and its rebuttal. In J. Ryberg & T. Tännsjö (Eds.), *The repugnant conclusion: essays on population ethics* (pp. 187–199). Kluwer Academic Publishers. https://doi.org/10.1007/978-1-4020-2473-3_10
- Prinz, J. J. (2008). Is morality innate? In W. Sinnott-Armstrong (Ed.), *Moral psychology*. Volume 1: *The evolution of morality: adaptations and innateness* (pp. 367–406). MIT Press.
- Railton, P. (2014). The affective dog and its rational tale: intuition and attunement. *Ethics*, 124(4), 813–859. <https://doi.org/10.1086/675876>
- Ruse, M. (2010). The biological sciences can act as a ground for ethics. In F. J. Ayala & R. Arp (Eds.), *Contemporary debates in philosophy of biology* (pp. 97–315). Wiley-Blackwell.
- Sensen, O. (2014). Universalizing as a moral demand. *Estudos Kantianos*, 2(1), 169–184. <https://doi.org/10.36311/2318-0501/2014.v2n1.3815>
- Shafer-Landau, R. (2003). *Moral realism: a defence*. Oxford University Press. <https://oxford.universitypressscholarship.com/view/10.1093/0199259755.001.0001/acprof-9780199259755>
- Singer, P. (1972). Famine, affluence, and morality. *Philosophy and Public Affairs*, 1(3), 229–243.
- Slingerland, E. (2014). *Trying not to try: the art and science of spontaneity*. Crown Publishers.

- Street, S. (2006). A darwinian dilemma for realist theories of value. *Philosophical Studies*, 127(1), 109–166. <https://doi.org/10.1007/s11098-005-1726-6>
- Tännsjö, T. (2002). Why we ought to accept the repugnant conclusion. *Utilitas*, 14(3), 339–359. <https://doi.org/10.1017/S0953820800003642>
- Temkin, L. S. (1987). Intransitivity and the mere addition paradox. *Philosophy and Public Affairs*, 16(2), 138–187.
- Temkin, L. S. (2012). *Rethinking the good: moral ideals and the nature of practical reasoning*. Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199759446.001.0001>
- Thomson, J. J. (1976). Killing, letting die, and the trolley problem. *The Monist*, 59(2), 204–217. <https://doi.org/10.5840/monist197659224>
- de Waal, F. B. M. (2014). Natural normativity: the ‘is’ and ‘ought’ of animal behavior. *Behaviour*, 151(2–3), 185–204. <https://doi.org/10.1163/1568539X-00003146>