

# Journal of Cognition and Neuroethics

## Should We Distrust Our Moral Intuitions? A Critical Comparison Of Two Accounts

**Felix Langley**

Kings College London

### **Biography**

Felix Langley is a recent graduate in philosophy from Kings College London. He specialises in the philosophy of cognitive science and neuroethics with particular focus on the role of intuitions in moral judgment.

### **Publication Details**

*Journal of Cognition and Neuroethics* (ISSN: 2166-5087). January, 2018. Volume 5, Issue 2.

### **Citation**

Langley, Felix. 2018. "Should We Distrust Our Moral Intuitions? A Critical Comparison Of Two Accounts." *Journal of Cognition and Neuroethics* 5 (2): 51–74.

# Should We Distrust Our Moral Intuitions? A Critical Comparison Of Two Accounts

Felix Langley

## **Abstract**

In this paper, I explore the question of whether we ought to defer to our moral intuitions across a range of situations by critically comparing two of the major views on this debate. The views I compare are those of Gerd Gigerenzer and Joshua Greene. Despite both having influential and opposing views, they have never engaged with each other in print and are not often directly compared. Gigerenzer is of the view that our moral intuitions are, broadly speaking, adaptive, whilst Greene takes the opposite view. The main contention that I focus on is Greene's supposition that our moral intuitions are maladaptive in what he calls 'unfamiliar' moral situations (i.e., problems that have arisen in our recent history; e.g., global poverty, terrorism, trolley problems, etc.). My conclusion is that Gigerenzer's thesis is either trivial or false because the areas that Greene identifies as being unsuited to our intuitions are precisely the areas that we should care about, and the conceptual tool that Gigerenzer employs to avoid this (ecological rationality) cannot plausibly solve these problems. The normative framework I employ to judge an intuition as 'better' or 'worse' is one to which both parties can agree.

## **Keywords**

Joshua Greene, Gerd Gigerenzer, Moral Psychology, Intuition, Cognitive Science, Neuroethics, Heuristics, Behavioural Economics

## **1. Introduction**

When one reads any great work of ethics in the western cannon – Kant, Hume, Hobbes, and Aristotle – an account of how we ought to be is intimately connected with an account of how we are. The central thesis of this paper is that across a significant range of moral cases we ought to distrust our moral intuitions. I will argue for this by critically comparing two major contemporary accounts of moral psychology, one which seeks to defend our intuitions and one which challenges them. The accounts in question are that of Gerd Gigerenzer and Joshua Greene; specifically, I will compare the opposing prescriptive claims these accounts make regarding how our intuitions should be employed in moral problems. My conclusion will be that Joshua Greene is correct in his claim that, across a significant class of cases in the moral domain, we should distrust the conclusions of our intuitive moral judgment, and, instead, default to our more reflective moral judgments.

The layout of this paper will be as follows: in the first half, I will explicitly state my normative framework and expound the two major theories, their normative prescriptions, and what the central conflict between them rests upon. In the second half, I will show that Gigerenzer's account fails in its prescriptive project because: (1) his claims can ultimately be shown to be philosophically trivial; and, (2) the concept that this theory could use to resist this charge, Ecological Rationality, cannot be applied in the relevant cases. By (1), I mean that across what we might think of as 'significant cases' (i.e., morally important, high stakes cases) Gigerenzer's account seems to fail, even if it succeeds across more mundane or non-moral cases. By (2), I mean that the concept of ecological rationality, upon which Gigerenzer leans heavily, is an implausible strategy in the cases with which we are concerned.

### 1.1. Normative Framework

It is important to establish, before moving forward, what exactly will be meant in the following by 'good' and 'bad' and 'adaptive' and 'maladaptive'. In order to compare the two theories, we must have some consistent common conception of 'good'. In order to avoid certain moral disputes, both metaethical and normative, the framework I will employ here will be as minimal as it can be whilst still being useful. I will suggest that a feature of our moral psychology is maladaptive if it leads us to make choices based on factors which we would not consciously agree are relevant. What I mean by this is that, for any factor X, if X is shown to effect our choices but we would not when asked agree that X is morally relevant, X can be called maladaptive. Things that would fall into this category would be factors like the order in which information is presented, spatial proximity, and whether a harm was caused physically or remotely – factors, which, when asked, people would say shouldn't be relevant to our decision making. To use a further hypothetical, though as we will see not entirely absurd, example, if it transpired that a person's height factored into whether we felt it was morally wrong to harm them, then it would seem that this tendency could be called bad or maladaptive since height just isn't the sort of thing which we, both lay people and ethicists, take to be morally relevant.

In addition to this idea of normative irrelevance, features of common sense ethics like 'all other things being equal inconsistency is bad' and 'all other things being equal it is better that less people be harmed than more' will factor into this framework. These are the kind of normative features which Greene describes as being "uninteresting" (Greene 2014, 771), insofar as they are claims we all seem to accept and do not tend to be the major salient features of moral argument. I accept that many elements of this normative

framework can be subject to metaethical scrutiny. For example, it has been argued that the legitimacy of prevailing our stated preferences over our relieved one's can be called into question. However, I take this framework to be plausible enough that I will assume it to be true for the purposes of this essay, since: (a) even if elements of it can be questioned it is the morality most of us seem to accept and live by, thus it remains an interesting question to see how these competing models fair under it; and, (b) a deeper metaethical argument would take us beyond the scope of this paper.

## **2. Kahneman, Tversky And Two System Models**

The model of decision making being proposed by Greene has its origins in the work of Daniel Kahneman and Amos Tversky (1974, 1124–1131). Furthermore, much of Gigerenzer's work is an explicit reaction their project. Thus, in order to correctly understand both Greene and Gigerenzer's accounts, it becomes important to understand the broader approach that informs them. Kahneman and Tversky's project, often referred to as the 'heuristics and biases account', posits a two system model of decision making wherein our minds can 'switch' between a slow, deliberative system of thinking (system two in Kahneman's terminology), and a faster, seemingly effortless, and less deliberative system that makes greater use of heuristics and mental shortcuts (Kahneman 2003, 697–720).

System Two is what seems to allow human beings to solve complex, novel problems and adapt quickly to change (Kahneman 2003, 669). Using it, we can consciously and deliberately focus our attention to a given problem (e.g., solving an equation, choosing a gift for our significant other, coming up with a philosophical argument, and learning how to play an unfamiliar instrument etc.). In short, system two seems to largely capture what we mean when we talk about reasoning or deliberate problem solving. In terms of describing its phenomenal character, we might think of the difference between driving on familiar roads, absentmindedly performing turns in a way which seems automatic but then being confronted with an unfamiliar diversion. The diversion creates a 'cognitive load' which causes us enter into a state of conscious problem solving, calling to mind alternative routes and performing rough calculations of time, perhaps assessing what could have caused it and assigning these causes varying degrees of probability. Similarly, when I arrive at class, the process of walking to the room, sitting down, unpacking my bag, happens almost entirely without any 'input' from me in a way in which I'm aware. However, as soon as the lecturer asks if anyone can see the flaw in an argument that's been presented or knows what school of philosophy Simone de Beauvoir is most

associated with, then cognitive load is applied and my thought requires input in a way that I am aware of and which appear to me more effortful than walking to class etc.

In short, system two is the 'precision tool' of our cognitive architecture. To be clear, I don't mean here that system two is more accurate *per se* or that it isn't capable or even prone to error, rather than is it 'precise,' in that it exists to solve novel, specific problems that require cognitive flexibility and deliberate reasoning. However, obviously, we cannot reason in this way all the time; the trade off we make with system two is that it is difficult and time consuming, hence our reserving it for specific situations where cognitive load is applied. For all other situations, we have System One.

System One is the inverse of the above, trading off precision and adaptiveness for speed and ease. System one thoughts are what we tend to think of as intuitive, reactions and process which come to mind fully formed, and indeed often don't 'come to mind at all'. This is the system which makes greater use of Heuristics. Heuristics, under this account, are (usually unconscious) mental 'rules of thumb' that our brains use in lieu of more complex decision procedures, as they tend to isolate one feature of the situation and make the choice based upon that. Prominent examples studied by Kahneman and Tversky include the availability heuristic, which bases the probability of an event occurring exclusively on how easily one can call to mind an example of it, the representative heuristic, which assesses whether A is a member of class B based on A's approximate resemblance to a mental model or stereotype of B. For example, people, when asked, overwhelming thought that a character with a meek, orderly description is assessed to be more likely to be a librarian than a list of other professions without consideration of other factors, such the statistical distribution of these professions.

There is an important point to make here that is easy to overlook when considering the interplay between these two systems. System Two makes all judgments, but it does not necessarily modify all judgments. This is the difference between judgments made under cognitive load, wherein there is deliberation, and those made intuitively. Thus, if someone were to ask me, on the spot, which has a greater population, Detroit or Grand Rapids, I might intuitively choose Detroit because of the recognition heuristic. If I am asked to think harder about it, system two might engage in deliberation, or if I am not, then I might simply settle on Detroit. This is the difference between an intuitive and non-intuitive judgment, both are made ultimately by system two, but the intuitive judgment is made entirely on the basis of the information that system one provides.

Having laid out the board strokes of the heuristics and biases account, I will now go on to expand why it is termed the 'biases' account'. As is perhaps already becoming clear, for all the dual system's elegance, our propensity to be hugely more sensitive to certain

sorts of information when making choices than to others is likely to have epistemically unhygienic consequences. According to Kahneman's account, as their work in this field progressed they began to increasingly find that these heuristics and mental shortcuts were, across a number of domains, leading us to error on a systemic level. These errors are termed 'biases'. To clarify this picture, I will expound some of the specific features of our intuitions which can cause them to lead us astray:

### 2.1. Accessibility

The 'pull' that our intuitions seem to have on us, in terms of their appeal as options, appears to stem from the ease with which we can access them. As has been discussed, conscious deliberation to answer a question is slow but our intuitions seems to pop into mind fully formed without (from our perspective) us needing to exert cognitive effort. This appeal of accessibility means that we are likely to, for example, take the frequency of events (e.g., terror attacks or air travel disasters) to correlate strongly with how easily we can bring to mind an example of one occurring. Now, naturally, in a great many everyday cases this heuristic will, in fact, lead us to truth, but it is very easy to see that, across a range of examples, this won't hold true, and, indeed, we might further worry that many of the examples in which it won't hold true are situations that invite dangerous outcomes (e.g., a voting population that demands harsh checks on personal freedom because of a perception that terror attacks are occurring several orders of magnitude more commonly than they are).

### 2.2. Sensitivity To Framing Effects

A further significant concern is the degree to which our intuitive judgments are sensitive to framing effects (i.e. presenting a problem in a particular way or using some words rather than others leads to substantial changes in the outcomes of choices in a way which we might think is deeply problematic) (Tversky and Kahneman 1981, 453–458). One major example of this effect is that people are highly sensitive to how losses and gains are presented such that the majority of people, when presented with two strategies for preventing a disease, will overwhelmingly choose the less risky option when it is presented in terms of the lives it saves (e.g., will save 200 lives) but overwhelmingly not when it is presented in terms of lives lost (e.g., 400 lives will be lost). This is despite the fact that, in both cases, participants knew the total number of lives at stake. Thus, the only thing that altered most people's choice was a slight change in how the information was framed.

The problem runs even deeper than this though. It is tempting to think that more 'rational,' intelligent people would rely less on heuristics or intuitions and thus be less susceptible to biases; however, this is not the case. Not only does greater intelligence fail to track freedom from bias, but also highly intelligent people are not free from biases in domain specific situations in their own fields. One of Kahneman's studies found that graduate statisticians at one of the top colleges in the US were led to a conclusion by the representative heuristics that basic statistics tells us is impossible (Kahneman 2003, 712).

### **3. Greene: Automatic And Manual Morality**

Now that we have the broader theoretical framework in place, we can discuss Greene's approach from a place of understanding. To a large degree, Greene's account consists in importing the heuristics and biases account unchanged into the moral domain. He explicitly accepts Kahneman and Tversky's model (Greene 2014, 695–726) and shows how it can be cashed out in moral situations. The analogy Greene uses to build on their system one and two notion is the 'Camera analogy'. The camera analogy compares system one to the automatic settings on a digital camera, preconfigured settings which allow the photographer not to worry about adjusting the variables themselves, thus making it highly efficient but inflexible. By contrast, the manual setting allows each individual variable to be adjusted manually, meaning that it is highly flexible but deeply inefficient. Having both these 'settings' make both a camera and a human mind highly efficient.

The central test cases that Greene employs to demonstrate how this model works in practice are 'trolley problems'. Trolley problems (Foot 1967, 5–15) are a broad class of philosophical problems and have been studied since the 1960s as 'test cases' for moral theories and various adaptations and iterations of these moral theories. Like fruit flies in biology, trolley problems are thought to tease out some of the most basic elements of our moral views. The basic problem is as follows: a runaway rail cart (or trolley) is heading down a track towards a group of five workmen and you can save them by pulling a switch, which will change the course of the trolley onto a track that will only kill one workman.

The switch case seems like a relatively easy choice, a trade of one life for five, but now let's say we're faced with a different version of the problem, namely the 'footbridge' case. In the footbridge case, we see a trolley heading towards five workmen, but, this time, we cannot pull a switch to stop it. What we can do, however, is push a man off an overhead footbridge into its path, which will save the five workmen but kill the one man

on the footbridge. As with the above, this is a one for five trade; however, this strikes us as a much more morally difficult choice, as our intuitions seem to resist pushing the man in a way that they don't resist pulling the switch.

The explanation for this disparity can be easily given in terms of the two system model. The footbridge case, unlike the switch case, involves factors to which our intuitions (and the heuristics that inform them) are sensitive, the most notable in the footbridge case being personal force. As Greene et al describe it (2009, 364–371), it seems that our moral intuitions (i.e., our system one moral judgments) are highly sensitive to the application of physical personal force in a way which 'counts against' the decision being taken.

Multiple studies by Greene et al involving brain imaging; lesion studies, and a variety of self-report studies, seem to bear out this above hypothesis. Cases which are constructed in ways that trigger the sensitivities of our system one tend to produce moral judgments which are: (a) more emotional; and, (b) more associated with what we would typically call 'deontological' moral judgments (Greene et al 2004, 389–400). To unpack this a little, cases in which we are required to push someone off a bridge, or more viscerally smother a child to prevent soldiers being alerted, involve features like direct personal force which trigger an emotional system one response. This then comes into conflict with our more deliberative system two response, often in a way which overrides the cost benefit analysis that characterises this sort of response.

The major contention that Greene draws from this descriptive account, in line with Kahneman and Tversky, is that, in spite of its many adaptive features, system one is often sensitive to features of situations that couldn't possibly be relevant to the decision being taken. We can understand this as a form of 'moral biasing.' I will go on to expound the details of this moral biasing view, which we can draw from Greene's work, and what prescriptive claims he makes in light of this.

### 3.1.Moral Biasing

Moral biasing, analogously to the wider systemic biasing discussed previously, is a worry that our system one is sensitive to features of situations that are not simply different to system two but different in a way which is maladaptive. To use Greene's main example, it cannot possibly be morally relevant that a person is dropped from a bridge by remote control rather than pushed, yet when the question is framed in these terms the number of people willing to kill in the footbridge case more than doubles (2013, 215).



This, it seems, is deeply troubling given the ability of our system one to override what we might think of as being our more considered judgments.

The evidence for this moral biasing seems to be fairly substantive, Greene's own extensive work in the field has, as previously discussed, used a wide variety of experimental methods to bear this point out (Greene 2014, 701–705). Additionally, these findings have been borne out elsewhere, with one of the more notable examples being Singer's drowning child case (Singer 1972, 229–243). The case is one in which a young child is drowning in a shallow pond and it would cost us only the price of a ruined suit or pair of shoes to save them. Clearly, not saving the child would be morally unthinkable to both our reflective and intuitive moral judgments. Yet, as Singer and other researchers note, for the same small sum of money the life of a small child on the other side of the world could be saved; however, people are drastically less likely to donate even small sums of money to charities that could save these children's lives. It seems, then, that mere spatial distance (i.e., spatial distance in the absence of some other morally important factor) bears on our moral judgment.

A similar study compared two cases where you personally witness a humanitarian tragedy in a country and are asked to donate verses a case where your friend is in the country and shows you a video before asking you to donate. The difference here is clearly not a morally important one, yet it appears to affect people's judgment, in that, people who are not imagining being physically present "drastically" (Greene 2014, 769) less likely to donate. Additionally, it appears race and in-group identification has a fairly substantial bearing on these choices in some situations (Swann et al 2010, 1176–1183) and this is before we get into the substantial literature on racial bias in jury decisions (Sommers 2007, 171–187).

### 3.2. Greene's Prescriptive Claims: Changing Norms

Off the back of the above research, Greene then goes onto make his major prescriptive claims. In being mindful of the Is-Ought distinction (Hume 1738, 3.1.1) (the description that states you cannot derive a normative claim from a merely descriptive one) Greene posits, as I did earlier in this paper, that we can motivate these prescriptive claims on the basis of uncontroversial normative beliefs that we already have. To wit, the descriptive claims gain normative force by being parasitic on our common normative belief set. Thus, no is-ought transgression occurs.

Greene's central prescriptive claim is that we ought to default away from, and indeed distrust, our moral intuitions across a range of situations. This notion is most clearly

expressed in what Greene calls the 'no cognitive miracles principle' which he describes thus:

The No Cognitive Miracles Principle: When we are dealing with unfamiliar\* moral problems, we ought to rely less on automatic settings automatic emotional responses and more on manual mode conscious, controlled reasoning, lest we bank on cognitive miracles. (Greene 2014, 715)

To unpack this principle, Greene takes a 'miracle' here to be a situation wherein, given the way in which our moral intuitions evolved and what they are for and sensitive to, it would be miraculous if they lead us to good moral conclusions. To illustrate this first with a non-moral example, imagine if a quantum physicist presented me with evidence for the correctness of a particular claim and I replied by saying 'that seems wrong to me, intuitively'. This seems like a poor response since quantum physics just isn't the sort of thing about which we should expect human beings to have accurate intuitions. The specific argument laid out by Greene as to whether we should expect our intuitions to be accurate is based upon the notion that our intuitions are primarily based upon experience, be it evolutionary (i.e., useful capacities developed in response experiences by our ancestors being passed down to us genetically) cultural (passed down by cultural experience) or of course personal experience.

In light of this, we can start to determine which sorts of moral problem our intuitions will be able to lead us to good decisions and which situations we ought to adopt a more reflective, cognitively engaged approach. The key, according to Greene, is moral problems that arise from recent (in relative terms) and thus unfamiliar developments. Examples of this would be things like climate change, bioethics, public health, global poverty, terrorism, existential risk (i.e., risk to the continued existence of humanity from nuclear annihilation, pandemic, or AI risk), race and gender relations, etc. This claim, that 'unfamiliar' problems are likely to be where our intuitions lead us astray is supported by work from Cass Sunstein (2005, 531–573) who posits, with Greene, Kahneman, and Tversky, that though our moral intuitions might be useful in everyday situations once they are confronted with what he terms 'exotic' problems, they seem to lead us astray.

One troubling observation a reader might make here, and this will prove critical later, is that the areas where it seems our intuitions might lead us astray are the most high cost, both morally and practically. Although, as Greene himself notes, these problems might well have components which are 'familiar' to us, the broad strokes of them will not be, as these are all problems that are recent in the grand scheme of human culture and

evolution, and since these things inform our intuitions, their absence suggests that our intuitions are not likely to be adaptive in these areas.

Our strategy then, it seems, should be that we ought to systemically distrust our intuitions across the above class of situations and default away from them when making decisions. When we have a conflict between our intuitions about a case and our more reflective system two thoughts, then we should go with our more reflective thoughts.

To preempt a tempting, though wrong-footed line of objection to the above, why don't our intuitions about 'unfamiliar' cases simply grow more accurate over time? If they are indeed grounded in experience, then surely we should expect our intuitions about these cases to grow more accurate, as we have more experience of them and thus not need to default away from them. This line of criticism fails because, as we found in Kahneman's research, certain sorts of problems are unfamiliar in some deep sense that doesn't seem to change with experience. Recall that people who had spent years of their life studying statistics had no more accurate intuitions about them than anyone else. Now, clearly, their ability to make reasoned, deliberative choices about statistics would be vastly better than average; however, this doesn't seem to effect their intuitions. Similarly, as will be discussed later, people who consider moral problems for a living appear to be subject to the same biases in their moral intuitions as everyone else. We can explain this disparity in terms of the inflexibility of system one judgments, as their speed and efficiency derives from their insensitivity to a broad range of information, which makes them very hard to alter. Thus, though prima facie appealing, this line of argument does not offer a solution to the problem.

### 3.3. Greene's Prescriptive Claims: The Specifics

Aside from its broad normative implications, what specific impact would Greene's prescriptions have? By which I mean, how specifically should we change our behaviour in response to this account? It seems that, when faced with a choice in areas where we shouldn't expect our intuitions to be useful, we should opt for our more deliberative reasoning instead of our intuitions. We should, as Socrates advises, 'Follow the argument where it leads' (Plato 1966) and go with the conclusion that our deliberative reasoning leads us to, despite intuitive resistance. Additionally, the program that Greene is suggesting implies that we should be more cautious if we only have intuitions about a topic. Say that you're confronted with a question about some complex issue of public health policy, you might know nothing about the topic but have a fairly strong intuition

that a certain policy would be best. Under this account, the prescription appears to be that we ought to remain agnostic.

Additionally, Greene's account, despite appearing, at first glance, somewhat pessimistic about the ability of human beings to make moral choices, has some extremely useful recommendations for how to improve our discourse on matters of ethics, politics, and public policy (Greene 2013, 295–298). He preempts an obvious criticism that simply saying 'think harder about tricky problems' is noble but perhaps useless advice; after all, we seem to already think that we have educated, well-founded opinions on complex problems when, in reality, these are principally intuition-driven. He retorts to this by pointing to a body of research by Fernbach, Rogers, Craig, Fox and Solomon (2013), which appears to show that people can be lead to change strongly held stances on issues of politics and policy by being asked to lay out, in detail, the problem or their solution to it. On discovering that they don't understand the problem, or their solution isn't as well thought out as they thought (or indeed at all), they either abandon their position or lower the credence they have in it.

#### **4. Gigerenzer: Fast And Frugal Heuristics**

Having laid out one half of the debate, I will now go to expound the major opposing perspective, namely Gerd Gigerenzer's 'Adaptive toolbox' account. Gigerenzer's 'Adaptive toolbox' account of cognition is at once similar and extremely different to the Heuristics And Biases account. As under Greene, Kahneman, and Tversky's model, Gigerenzer gives an account of cognition wherein heuristics have a central role in our decision making, and indeed gives a very similar descriptive account of heuristics themselves (i.e., that they ignore much of the available information, they lead us to fast decisions, they are in some sense automatic etc). However crucially, unlike the Heuristics And Biases model, Gigerenzer takes our intuitive cognition to be largely adaptive (Gigerenzer and Brighton 2009, 107–143). By this, I mean that he doesn't, broadly, conceive of our intuitions as being systemically biased and indeed holds that they can in fact make better decisions than conscious deliberation in many cases. In the following section, I will go on to expound Gigerenzer's thesis, the key concepts on which he relies, and how his normative claims interacts with those of the Heuristics And Biases Account.

Similarly to the previous account, Gigerenzer defines heuristics as being:

A strategy that ignores part of the information, with the goal of making decisions more quickly, frugally, and/or accurately than more complex methods. (Gigerenzer and Gaissmaier 2011, 454)

Gigerenzer himself frequently returns to the example of the Gaze Heuristic (Gigerenzer, and Brighton 2009, 108) to illustrate the above, this being a capacity to track moving objects (i.e., baseballs) through the air and to adjust our speed and direction to end up at the spot where it will land. He notes that the baseball playing academic, who attempts to work out where the ball will end up 'manually' will invariably fail since the capacity of any person to do the required calculations using system two is vastly slower, and less accurate, than the heuristic will be. With a single piece of information, namely 'keep your eye on the ball', the heuristic allows us to constantly complete the complex task of catching fast moving objects in a way that feels to us almost effortless.

Interestingly, this example crystallises precisely why and how Gigerenzer takes heuristics to be an adaptive decision procedure, principally through the notion of Ecological rationality.

#### 4.1. Ecological Rationality

I will now go on to carefully expound what the notion of Ecological rationality means in this context, since it is critical to Gigerenzer's account and thus to this debate more broadly. In its most simple form, the idea of Ecological Rationality is that decisions have to be thought of as an interwoven function of mind and environment and cannot be meaningfully talked about as being rational or irrational outside of the environment in which they are taken. Thus, rules which suggest things like, 'for any decision to be rational it must conform to rules X,Y,Z,' and are invariant across situations, are meaningless under this account. The notion is neatly captured by the idea of Simon's Scissors, which Simon coined by positing that:

Human rational behaviour is shaped by a scissors whose blades are the structure of task environments and the computational capabilities of the actor. (Simon 1990, 7)

The ideas here seem simple enough: human behaviour is grounded inexorably in the situation in which they must behave and we have a set of evolved cognitive capacities that produce behavioural tendencies when combined with the incentive structures created by environments. However, the implications of this account appear to be surprisingly radical.

If we take the account of ecological rationality seriously, we find that certain counter intuitive notions, like Gigerenzer's claim that heuristics that ignore information, even if there would be no cost to acquiring it, can make better choices than a system with greater information, become perfectly rational. Once a definition of rationality becomes

tied to a synthesis of mind and environment, it makes sense that actions that appear to violate rules of choice making (i.e., by not taking into account all available information) can be rational since the incentive structure created by the environment makes them rational. Additionally, this can explain how the moral adaptiveness or otherwise of an intuition can be explained in terms of environment. One of Gigerenzer's main examples is the 'status quo heuristic,' under which people will tend to do what everyone else is doing and ignore other (seemingly very important) information that also exists. The twin examples of how its goodness or badness depends on the environment are relayed presently. His first example (1) is a true example, wherein a group of ordinary Polish policemen during the second world war partook in the brutal massacre of Jewish civilians from their own nation (Gigerenzer 2008). The men were given the opportunity to step forward if they did not want to participate, and only a dozen of the 500 men present chose to not to partake. Gigerenzer goes to great lengths to point out that these men were mostly older, ordinary police men, not hardened members of the SS, and that there is good historical evidence that they were not particularly anti-Semitic. His point here is that these men were motivated by the status quo heuristic to not break ranks, and that this heuristic was powerful enough to override their intuition that murdering civilians is wrong.

His second example (2) notes that, in Britain and America, where organ donor laws are opt in (meaning you have to register to be one yourself), rates of individuals registered for organ donation are 17% and 28%, respectively. By contrast, in France and Hungary, where the laws are opt out (meaning you are automatically registered and must deliberately chose not to be) donor rates are 99% and 99%, respectively. His point is that each of these cases involve the status quo heuristic, yet, whilst in (1) it led to egregious acts of horror in (2) it leads to tens of thousands more lives being saved every year. The difference is the incentives unique to each environment, as the heuristics themselves are morally neutral.

#### 4.2. Gigerenzer's Prescriptive Claims: Environmental Design

The major Prescriptive claim that Gigerenzer's account produces, and with Greene he does so by combining interesting empirical facts with uninteresting normative notions, is that moral behaviour is best improved by focusing on how environments and the incentives they give rise to can be better designed to produce the kinds of behaviours we want. This can include simple acts such as minor alterations to the framing of situations (e.g., the officer in charge of the policemen asking anyone who *did* feel able to carry

out the task to step forward, rather than singling out those who did not) (Gigerenzer 2008, 6). It can also include higher level policy decisions such as making laws regarding organ donation opt in or, as proponents of nudge theory suggest, things like reducing the size of glasses to reduce excessive drinking, making salads rather than fries a default to encourage healthy eating, etc. Though these last examples are more geared towards public health, it seems that something like this mechanism is what Gigerenzer is advocating. However, it is critical to note that, whilst Gigerenzer is an advocate of something like nudging, broadly defined, he is opposed to its justification on the basis of human irrationality and 'libertarian paternalism' (Gigerenzer 2015, 361–383). To wit, he views the standard justification for nudging as buying into the systemic biases account of human cognition that he explicitly rejects.

It is important to clarify here that, whilst Gigerenzer rejects major elements of the nudging program, he is also explicitly advocating for environmental engineering as a way to promote adaptive behaviours (Gigerenzer 2008, 5–6; Gigerenzer 2010, 542). Just as the talented runner must have paths and tracks to run on, in order to make use of her natural gift, so too does Gigerenzer seem to suggest that our intuitive moral thinking will lead us to good choices, if only the environment allows it. To wit, for Gigerenzer, engineering environments is more a matter of allowing people to flourish than leading them because their judgment cannot be trusted. I want to make this point very clear, both because I wish to accurately convey Gigerenzer's stance and because, in light of this, there are certain criticisms of the nudging program that will apply to him and others that won't.

Before I move on to my major argument against Gigerenzer, it is interesting to note how both theories converge on the inflexibility of intuitions. Even Gigerenzer, whose account is deeply sympathetic towards our intuitions, gives prescriptions for how they could be improved by changes to incentive structures and decision environments. Both accounts recognise that we cannot simply will our intuitions to function differently to how they do. Under the descriptions of both accounts, this makes sense for the same reasons. Gigerenzer makes much of how useful it can be that our heuristics are informationally frugal; however, it seems, as has been stated, that it is this resistance to the majority of information that leads them to be so inflexible. In short, both accounts find it necessary to try and account for the inflexible nature of our intuitions.

## **5. The Central Conflict: Changing Norms Vs Changing Environments**

Cashing out the major prescriptive claims of each theory allows us to see what the central conflict between them is. On one hand, we have Greene, who suggests that we

need to change our norms and decision procedures, and, on the other, Gigerenzer, who suggests that we need to change the structures of our decision environments to let our intuitions function better. In this section, I will lay out the central arguments we can bring to bear between the two positions over this key issue and assess them.

I will begin by noting that there is some very trivial sense in which Greene can concede that, yes, our intuitions are only maladaptive given certain decision environments and problems. If we lived in a possible world where we never faced any 'unfamiliar' problems, then, naturally, our intuitions would be unproblematic. So, in one sense, the ecological rationality point is true, but in a very broad, uninteresting sense. The interesting question, upon which the conclusion of this essay will turn, is going to be which of these prescriptions is more plausible in our world, as is.

There is an interesting quote from Sunstein who, when describing the confidence we have in our intuitions regarding 'exotic' cases, says the following:

They might not deserve to be so firm, simply because they have been wrenched out of the real-world context, which is where they need to be to make sense. (2005, 541)

This quote seems to agree with Gigerenzer's point regarding ecological rationality, yet uses it to support an argument of the kind that Greene is making. Although, as I've suggested, we don't need to look to non-'real-world' examples to find situations where our intuitions won't help us. Now, while this quote seems to make a point that Gigerenzer would agree with (i.e., that our intuitions can become maladaptive if they're removed from an environment with the correct features), I hold that the broader point of the quote is highly damaging to his case. Consider that, if the only situations in which our moral intuitions are useful are ones which are not morally important, then it appears that Gigerenzer's case is reduced to triviality. To clarify more formally why this argument is so damaging:

1. Gigerenzer's view of his own program is that people like Greene et al. are inaccurately characterising our moral choice making as systemically flawed and that, in reality, our moral intuitions can be adaptive in helping us achieve our own moral ends. This is because, under the adaptive toolbox account, our heuristics are capable of quickly and accurately responding to the incentives of the environment. Indeed, he makes much of the fact that the adaptive



toolbox account can be prescriptive. Clearly, the ability for a theory to be meaningfully prescriptive is an important theoretical virtue.

2. However, as Sunstine and Greene's research seems to suggest, the situations in which this occurs are in everyday, morally trivial cases and not in important, morally high-stakes cases.
3. Thus, we can agree with Gigerenzer up to a certain point, but, at the same time, reduce his claims to relative triviality, since, when we talk about 'moral decisions,' we are normally talking about precisely the kinds of situations in which our intuitions fail us. Ergo, our intuitions are not 'morally' useful, in any interesting sense of the word.

To clarify 2 and 3 further, as has been previously stated, our moral intuitions can be useful in familiar decision environments. Aversion to personal harm, defaulting to the status quo etc. seem to be clearly useful tools of maintaining peace and stability, this much the skeptic of intuitions can concede. However, the range of dilemmas in which our intuitions fail us are precisely that, dilemmas. Our intuitions may hold us back from punching someone who we dislike, or move us to comfort a distressed child, but whether or not we ought to do these things isn't really up for question, as we don't agonise over them and they don't appear to be captured by what we mean when we talk about 'moral' choices. By contrast, the exotic dilemmas which our intuitions fail to help us with generally are the sorts of things that the phrase 'moral choice' gets at. Should we give up civil liberties to protect against terror? Could we ever justifiably carry out a permeative strike nuclear strike? If I would save a child dying right in front of me for a small sum of money what else does that commit me to? Whilst the domain of familiar problems, in which Gigerenzer's prescriptions are useful, doubtless contains some moral problems it seems that the majority of them (and certainly the vast majority of high-stakes problems) exist in the domain of the unfamiliar, where Gigerenzer's prescriptions are not useful. Thus, if this reasoning holds, Gigerenzer's prescriptive case is trivial when it comes to solving moral problems.

### 5.1. Gigerenzer's First Retort

A reply to the above, on behalf of Gigerenzer, would be that, yes, it might be that morally high-stakes decision environments cause problems for our intuitions as it currently stands, but it is a leap in logic to then claim that Greene's solution is the correct one. Given the effectiveness of our intuitions in situations where they do work, it seems that it would be better to try and change these decision environments rather than change our norms. This is true for both a positive and negative reason. The positive reason is that, as Gigerenzer's research seems to bear out, once the correct environment is discovered or created, our intuitions function with startling ease and accuracy – sometimes more so than conscious deliberation. Consider, for example, trying to engage people's system two to convince them to become organ donors (which appears to fail) to simply changing the laws to opt out (which seems to succeed). If such a workaround can be found for other high-stakes moral situations, and, contrary to the above argument, organ donation is an extremely high stakes moral situation, then the model becomes decidedly non-trivial. The negative reason to support the adaptive toolbox account over its competitor is that, as Gigerenzer notes, we can point to fairly damning flaws in cost-benefit thinking when it comes to moral problems (Gigerenzer 2008, 20–23), in that, in an uncertain and complex world, working out a solution on the basis of computational reasoning will prove far too complicated to be useful.

### 5.2. A Reply to Gigerenzer's Retort: Deep Features of Environments

To reply to the negative point first, whilst the above would be a damaging point if the only options were a heuristic approach or a purely computational one, this is not the state of affairs we find ourselves in. Something like rule consequentialism occupies a middle ground between these two positions. By this, I don't mean a commitment to rule Consequentialism per se but the kind of decision procedure that it implies (i.e., a deliberative process of establishing a rule for a range of situations and defaulting to it). A critic might, at this point, quibble with the degree to which this is really a departure from Gigerenzer's position. I posit that the departure is significant, though subtle. Whilst this sort of rule based decision procedure is perhaps superficially similar to a heuristic based decision procedure (since heuristics are rules), it is a rule established and shaped through deliberative reasoning, which we have access to, rather than by the various unconscious aspects of our cognition, which we don't.

The second element of his case requires more detail to address. This, fundamentally, is the major tension at play between the two accounts (i.e., the issue of ecological

rationality). At base, if Gigerenzer's strategy of environmental engineering and institutional design are not plausible for the problems with which we're concerned, then this entire prescriptive case is implausible. Something that I would suggest makes Gigerenzer's strategy implausible is the notion that decision environments have 'deep features,' for which there is no plausible method of engineering that is workable or non-coercive.

It strikes me that all the concrete examples that Gigerenzer presents of environmental design being successful, or cases where they could be successful, revolve around seemingly surface level features of environments (Bennis et al 2012). Changing the organ donor law to opt-out, though undeniably very effective, involves a relatively simple change in the law. By contrast, people failing to give to charity because of proximity related biases seems trickier. After all, the feature of the environment which causes the problem (i.e., that the people in question are vast distances away) is not something which can be engineered away. Despite the regular presence of advertisements from various charities showing us distressing images of plight, the fraction of people who give to charity, and the amount they give, remains tiny (McKenzie and Pharoah 2011). Thus, it seems that simply trying to use advertising to tug at peoples intuitions isn't working as a strategy. This is more than just a one-off case, for the vast majority of people don't take steps (such as voting for particular policies, etc.) to reduce climate change, despite a wealth of evidence of its effects being regularly displayed on the news that most people consume. We might attribute this too to a kind of proximity bias, albeit a temporal one, insofar as being in full possession of the facts does nothing to effect our intuitions. What ties these, and the other high-stakes cases together, is that their unfamiliarity (and thus propensity to fall victim to our systemic biases) is tied to features of the situation that it would be either impossible or at least far from optimal to remove.

Now, naturally, there are other measures that could be employed to get people to comply with their stated preferences. In the case of global poverty, we could simply make it a default that companies must donate a portion of their employees salary to charity, though, of course, they could opt-out if they wished. This method, however, begins to seem more like an example of coercion than simply allowing people to flourish. This more direct path becomes even more murky when it comes to matters of public policy (note that many of the high-stakes examples concern things like terror policy, immigration, nuclear weapons, etc.) and, as such, there doesn't seem to be a way to influence choice architecture that doesn't also favour one political party undemocratically. As we noted previously, Gigerenzer has already explicitly rejected the more 'heavy handed' approaches to nudging favoured by people like Thaler and Sunstine.

A significant additional problem, which this account falls victim to, is precisely the criticism that Gigerenzer, I believe falsely, levels against accounts like Greene's, namely over complexity. Consider that, for this approach to work, it would need to be the case that for all high-stakes moral problems, a certain account of environmental engineering would need to take place. This, in turn, would require a detailed understanding of the ins and outs of what informs our heuristics on an extending broad and diverse range of issues. This kind of problem seems to be intrinsic to any account that would seek to change the world around us rather than attempt to change us.

By contrast, Greene's decision procedure seems far more flexible, insofar as when we are required to act or form an option on some moral issue, and an intuition presents itself, we need only ask questions which appear to have obvious answers. Does it concern an issue which is, in the grand scheme of history, new or unfamiliar to us? If so, we ought to assume that our intuitive judgment isn't likely to be accurate and default away from it.

### 5.3. Gigerenzer's Second Retort: Greene's Account Is Useless for Most People

A different line of argument that a defender of Gigerenzer might take would be to posit that Greene's account applies only to cases which the majority of people don't need to deal with. This functions as the inverse of my argument that Gigerenzer's account fails if it only applies in mundane cases, since they might suggest that, by the opposite token, a theory that only applies in extreme cases has equal claim to being trivial. Specifically, they might suggest that the moral sphere of the average agent (i.e., the range of moral problems that they need to engage with) includes things like 'should I drunk drive' or 'should I cheat on my spouse,' etc. Cases of the sort that Greene's theory better applies to are issues that fall within the moral sphere of policy makers and a narrow range of experts. Thus, the critic might suggest that Greene's account represents a decision procedure for a very specific kind of agent making a specific kind of decision. Indeed, Greene himself admits that our moral intuitions work for most people most of the time. The theoretical motivation here might be that if both theories are liable to the accusation of being trivial, then it can hardly be argued that Greene's account is superior to Gigerenzer's in virtue of that criticism.

The above argument is certainly concerning; however, I would suggest it fails, insofar as it incorrectly characterises the moral sphere of the average ethical agent. By this I mean that, even if we think that it matters more that certain people have more carefully considered moral positions on high-stakes issues than others (i.e., policy makers), we can still put forward a convincing case that all moral agents ought to have more careful

attitudes in high-stakes cases. This is because there are at least two significant ways in which the average moral agent can impact these high-stakes issues. Firstly, there are the obvious, 'hard' impacts an agent can have (i.e., voting for a given policy or giving their money to a given charity), and here we might point to the fact that there is significant public division on issues where there seems to be an extremely high expert consensus (Johnston and Ballard 2016, 443–456) such as immigration, which appears to be an unfamiliar moral issue. The second way in which the average moral agent can effect these issues is the 'soft' way (i.e., by contributing to a general culture of what sort of solutions to problems are acceptable and what behaviours we allow). In general, there are a variety of things we can all do to move the Overton window (the name given to the hypothetical range of acceptable policy suggestions and opinions. This is where the work of Kahneman becomes relevant again. Kahneman notes that the act of gossip can be successful when it comes to shaping peoples attitudes, since being able to predict how your action would be gossiped about, and gossiping about the acts of others, can lead to more adaptive behaviours (Kahneman 2013, 406–409), because we are highly attuned to the faults of others and to how we might imagine others see our faults. My point here is that, whilst the average moral agent is unlikely to ever be in the position of deciding the exact nature of, for example, bioethics policy, personally, they each contribute in both soft and hard ways to the climate which dictates what the policy will be.

#### 5.4. A Third Criticism of Greene: Impossible Prescription

The final criticism that might be levelled against Greene's account is the one that I take to be the most problematic for it. The critic might concede the previous points, that Greene's account is preferable to Gigerenzer's *insofar as it's possible* but might counter that Greene's prescriptions simply aren't possible. For Greene's account to work, agents would need to be capable of defaulting away from their intuitions in unfamiliar decision making. The critic might suggest that we simply don't have that kind of cognitive control, that our intuitions will always be more appealing to us than our more reflective judgments. The weight of this criticism is added to by the fact that advocates of the heuristics and biases view, such as Kahneman and Sunstine, are themselves skeptical of how effectively people can be 'debiased'. Additionally, there is research from, among others, Kushman et al, which suggests that professional philosophers do not seem to be any less susceptible to biases than anyone else, even in examples with which they are familiar (Schwitzgebel and Cushman 2012). Thus, people whose job it is to think reflectively about moral issues still seem to be unable to resist their intuitions. In light

of this, it might be suggested that, even if it is suboptimal when compared to Greene, something like the ecological rationality view is the only game in town.

I would suggest that the above line of criticism, though forceful and intuitive, is mistaken. The first, and most obvious retort, is that we have compelling reasons to reject the empirical claim upon which the criticism rests. As mentioned previously, it seems that when leading people to engage in a certain sort of deliberative process (i.e., thinking through the problem and how their solution solves it), people seem capable of overcoming their intuitive positions. Indeed, Greene's own research seems to show that people with certain sorts of training – the specific case he appeals to is public health officials – seem to adopt a more deliberative decision procedure. We see this evidenced in their willingness to take courses of action, both in their area and in more general cases, that our intuitions would ordinarily reject (Greene 2013, 128–131). The criticism, moreover, suffers from a deeper problem. I would suggest that it falsely conflates skepticism about the current viability of Greene's project with skepticism about his project being worth pursuing. Whether Greene's prescription that we should default away from our intuitions in certain cases fails because it isn't currently possible (and as I have shown this can be questioned), it doesn't follow that we shouldn't try to act with this goal in mind, even if we are currently liable to fail. In short, in addition to the criticism being empirically questionable, it strikes me as unnecessarily defeatist.

## **6. Conclusion**

To conclude, both of these accounts not only provide an invaluable insight into human cognitive architecture but also raise normative concerns that we ought to take seriously. However, what I have sought to demonstrate is that Gigerenzer's account, however successful in the non-moral domain, fails to provide an acceptable defence of our intuitions. It fails, insofar as it becomes trivial when applied to morally important cases, and the concept upon which it depends to resist this triviality, ecological rationality, is a sub-optimal and indeed implausible approach to these problems. Additionally, I have sought to show that Greene's account is useful, plausible, and capable of resisting various criticisms to the contrary. Thus, I conclude that Greene's account is superior and we should indeed distrust our moral intuitions in many significant cases.

## **Bibliography**

Bennis, Will M., Konstantinos V. Katsikopoulos, Daniel G. Goldstein, Anja Dieckmann, and Nathan Berg. 2012. "Designed to Fit Minds: Institutions and Ecological

- Rationality" In *Ecological Rationality: Intelligence In The World*, edited by Peter M. Todd and Gerd Gigerenzer. New York: Oxford University Press.
- Fernbach, Philip M., Todd Rogers, Craig R. Fox, and Steven A. Sloman. 2013. "Political Extremism Is Supported by an Illusion of Understanding." *Psychological Science* 24 (6): 939–46.
- Foot, Phillipa. 1967. "The Problem of Abortion and the Doctrine of Double Effect." *Oxford Review* 5: 5–15.
- Gigerenzer, Gerd. 2015. "On the Supposed Evidence for Libertarian Paternalism." *Review of Philosophy and Psychology* 6 (3): 361–383.
- Gigerenzer, Gerd. 2010. "Moral Satisficing: Rethinking Moral Behavior as Bounded Rationality." *Topics in Cognitive Science* 2 (3): 528–554.
- Gigerenzer, Gerd. 2008. "Moral Intuition = Fast and Frugal Heuristics?" In *Moral Psychology, Vol 2, The Cognitive Science Of Morality: Intuition And Diversity*, edited by Walter Sinnott-Armstrong, 1–26. Cambridge: MIT Press.
- Gigerenzer, Gerd, and Henry Brighton. 2009. "Homo Heuristicus: Why Biased Minds Make Better Inferences." *Topics in Cognitive Science* 1 (1) 107–143.
- Gigerenzer, Gerd, and Wolfgang Gaissmaier. 2011. "Heuristic Decision Making." *Annual Review Of Psychology* 62 (1): 451–482.
- Greene, Joshua D. 2014. "Beyond Point-and-Shoot Morality: Why Cognitive (Neuro) Science Matters for Ethics." *Ethics* 124 (4): 695–726.
- Greene, Joshua D. 2013. *Moral Tribes: Emotion, Reason, And The Gap Between Us And Them*. New York: Penguin Books.
- Greene, Joshua D., Fiery A. Cushman, Lisa E. Stewart, Kelly Lowenberg, Leigh E. Nystrom, and Jonathan D. Cohen. 2009. "Pushing moral buttons: The interaction between personal force and intention in moral judgment." *Cognition* 111 (3): 364–371.
- Greene, Joshua D., LE Nystrom, AD Engell, JM Darley, and JD Cohen. 2004. "The Neural Bases of Cognitive Conflict and Control in Moral Judgment." *Neuron* 44 (2): 389–400.
- Hume, David. 1738. *Treatise On Human Nature*. New York: Oxford University Press.
- Johnston, Christopher D., and Andrew O. Ballard. 2016. "Economists and Public Opinion: Expert Consensus and Economic Policy Judgments." *The Journal of Politics* 78 (2): 443–456.
- Kahneman, Daniel. 2013. *Thinking Fast And Slow*. New York: Farrar, Straus and Giroux.

- Kahneman, Daniel. 2003. "A Perspective on Judgment and Choice: Mapping Bounded Rationality." *American Psychologist* 58 (9): 697–720.
- McKenzie, Tom, and Cathy Pharoah. 2011. "How generous is the UK? Charitable giving in the context of household spending." *CGAP Briefing Note 7*. London: CGAP.
- Plato. 1966. *Plato in Twelve Volumes, Vol. 1*. Translated by Harold North Fowler. Cambridge: Harvard University Press.
- Schwitzgebel, E., and F. Cushman. 2012. "Expertise in Moral Reasoning? Order Effects on Moral Judgment in Professional Philosophers and Non-Philosophers." *Mind & Language* 27: 135–153.
- Simon, Herbert. 1990. "Invariants Of Human Behaviour." *Annual Review of Psychology* 41: 1–19.
- Singer, Peter. 1972. "Famine, Affluence And Morality." *Philosophy And Public Affairs* 1 (3): 229–243.
- Sommers, Samuel R. 2007. "Race And The Decision Making Of Juries." *Legal And Criminological Psychology* 12 (2): 171–187.
- Sunshine, Cass. R. 2005. "Moral Heuristics." *Behavioral and Brain Sciences* 28 (4): 531–573.
- Swann, WB, A Gómez, JF Dovidio, S Hart, and J Jetten. 2010. "Dying and Killing for One's Group: Identity Fusion Moderates Response to Intergroup Versions of the Trolley Problem." *Psychological Science* 21 (8): 1176–1183.
- Tversky, Amos, and Daniel Kahneman. 1981. "The Framing Of Decisions And The Psychology Of Choice." *Science, New Series* 211 (4481): 453–458.
- Tversky, Amos and Daniel Kahneman. 1974. "Judgment under Uncertainty: Heuristics and Biases." *Science, New Series* 185 (4157): 1124–1131.