



THE
JUBILEE CENTRE
FOR CHARACTER & VIRTUES

The Dangers of Oversimplifying
Explanations of Moral Behavior:
Robert Plomin's and Jonathan Haidt's
Excesses

Blaine J. Fowers
University of Miami

UNIVERSITY OF
BIRMINGHAM

For over 150 years, scholars have been avidly seeking scientific explanations of human behavior, especially moral behavior. Two attempts to explain moral behavior are of particular interest to readers of this series: Robert Plomin's genetic behavior theory and Jonathan Haidt's Social Intuitionist Model. Both accounts are systematic and rely on considerable data, which are key benchmarks for qualifying as scientific explanations. As everyone who has taken a science class knows, a third criterion for science is that the explanation should be as simple as possible. William of Occam first stated this principle and it has become known as Occam's razor. Let's examine Plomin's and Haidt's views in terms of these three principles of science: systematic theory, data support, and Occam's razor.

Robert Plomin's Blueprint Thesis

Plomin's genetic behavioral account is a very good example of the application of Occam's razor. He discusses a great deal of data demonstrating that many aspects of human behavior, including personality, intelligence, some forms of mental illness, and achievement test performance can be partly explained by inherited capacities. Twin studies are one of the most compelling and primary methods genetic researchers use. They study identical twins, who share 100% of the genetic make-up, fraternal twins and siblings, both of whom share 50% of their genetic inheritance. They also compare twins and siblings that were raised together to those who were raised in different families. If identical twins raised in different families are more similar to one another than fraternal twins or siblings raised together on a given characteristic, this provides powerful evidence that shared genes are important in explaining the characteristic. Plomin (2018) summarizes this in "the first law of behavioural genetics: All psychological traits show significant and substantial genetic influence" (p. 11).

At this point, Plomin hews to a version of Occam's razor that favors the simplest explanation. If genes can explain much of human behavior, perhaps we don't need to take environmental factors (e.g., parenting, peer influence, community settings) very seriously. Although Plomin (2018) states that "both genes and environment contribute to the psychological differences between people...environmental effects...are mostly random—unsystematic and unstable" (p. xii). He consistently claims that genetic explanations are the only effects worth knowing about because environmental factors amount to little more than "random noise" (p. 80) that can be safely ignored.

The trouble here is that there are two ways to go wrong with Occam's razor, not one. The lesson we were all taught is to avoid needlessly complex explanations that include superfluous concepts and excess causal inferences. That is fair enough, but it is also possible to oversimplify matters by ignoring data or ruling out some potentially important explanatory factors. This seems to be just what Plomin has done. He is an extremely accomplished genetic researcher with a tremendous body of work, and he is greatly respected for that work. When he speaks about genetics, he does so with well-deserved authority. When he tells us that genetics can explain 40-70% of various sets of behaviors (e.g., verbal ability), it is impressive.

Yet it is not at all clear what makes him authoritative about the so-called "noise" of environmental influence, that could, in principle, explain 30-60% of the same behaviors he discusses. The first thing that should give us pause about his ability to discuss environmental explanations is that environmental factors are routinely relegated to a noise category in his research without seriously trying to seriously tease out their import. The "environment" is a catch-all category for him, including everything from income to agency, with scant effort to differentiate these vastly different factors. He describes genetic influence in detail through

adoption and twin studies and through painstaking studies of tens of thousands of genes with samples up to one million people. But Plomin has not devoted the same care to identifying and mapping environmental influences because he includes very few, quite simple environmental variables such as comparing students from private and public schools and never assessing participants' choices or goals. This means that he is making very broad assertions without good data to support his strong claims about the trivial influence of environmental factors.

The lack of evidence about his assertions means that we cannot simply accept them, but the absence of evidence does not mean that he is necessarily wrong. Is there evidence that leads us to think that Plomin is wrong about the triviality of non-genetic explanations? Unfortunately for him, there is abundant evidence that many factors (e.g., such as parenting, community environment, school environment, trauma, and so forth) influence key areas of human behavior (e.g., school achievement, crime, drug use, income). Of course, Plomin could say that whatever influence can be ascribed to these environmental factors is due to shared genetics (as in parent-child interactions) or gene-environment selection (as in a person selecting environments that are congenial to his or her genetic predispositions). But again, these would be speculative assertions rather than data-based conclusions. Plomin's strong desire to reduce explanations of behavior to genetic factors leads him to oversimplify those explanations, violating Occam's razor by applying it excessively.

This excessive simplifying zeal leads to another very common problem in attempts to explain human behavior biologically: the problem of reductionism. Many biologically-minded scholars want all behavior to be ultimately explained by some features of biology, whether those are based on genetic expression or concurrent biochemical processes. Reducing all explanation to biology suggests that the external world of the organism is

essentially irrelevant or only relevant in how it activates genetic expression or biochemical processes. We are light years away from having the kind of evidence that could justify such reductionism. Until that evidence is clearly available (if it ever is), oversimplification is the best way to understand this reductionism. Even worse for Plomin's claims, there are thousands of studies that clarify the many ways that situational factors systematically influence human behavior in many, many domains, such as helping behavior, perception, and social conformity. It is simply incorrect to assert, as Plomin does, that situational influences are small and random.

Plomin's thesis has an even larger credibility problem. To assert that genetic expression is the only important explanation of human behavior is to claim that human choice or agency is irrelevant. Plomin states that genes make certain outcomes more probable rather than determining them. He very occasionally tells us that we can decide how to respond to our genetic makeup, but there is no place in his views for human choice. Is the degree of choice genetically caused? Is choice also random and of negligible importance? Can one choose to go against the grain of one's genetics? Plomin is curiously silent on all of this. Of course, the debate about human agency has been raging without conclusion at least since Socrates' time. Although Plomin and others cite enormous quantities of causally-related data, the question remains inconclusive because no demonstration has been sufficiently clear to convince thinkers who see choice as important for understanding human behavior.

I cannot resolve the determinism-agency debate here, but I can point to many aspects of our humanity that would have to be sacrificed if our genetic make-up (or any combination of causal forces) is taken to be the only interesting explanation of our behavior. We would have to give up the idea that we plan and work for a future of our choosing. Our

aspirations and ideals would become just the outputs of causal forces. It would be impossible to sustain the idea that our lives amount to something that can be expressed in a meaningful life story. The concept of personal growth would have to be sacrificed because any change in an individual would be due to a genetically programmed development that could not be dignified by a term like “personal growth”. In a fully causal world, there would be no human capacity to reflect on one’s goals and actions and to alter one’s direction in life because that reflection leads to the recognition that there is a better way to live. For many people, the world of meaning, value, and growth is as real as the world of physical objects. Is it the world of meaning that is an illusion or is it the worldview that accepts only causal explanation? It seems rather foolish to sacrifice the world of meaning to the idea of full causation, which has not been demonstrated and very likely cannot be. As such, full causation of the human world is more an assumption or an article of faith than a documented scientific fact, although it is frequently asserted as if it were scientific fact.

Therefore, it seems that Plomin’s reduction of the human world to biology is an oversimplification and his fully causal account is more an expression of faith than of science. There is, of course, no doubting that our genes influence us. That much is scientific fact, and it is an impressive piece of knowledge. But, Plomin’s leap from being able to explain a portion of the variation in human behavior to the claim that by gene expression is the only useful explanation for human action is simply not credible.

Jonathan Haidt’s Social Intuitionist Model

The problem of reductionism similarly haunts the theory that fueled Jonathan Haidt’s rise to prominence. He is best known for his Social Intuitionist Model (SIM), which suggests that morality is primarily an intuitive process. His thinking is based on the dual

process model of psychology, which specifies that we process our circumstances in two ways: in a fast, automatic, unreflective cognitive-emotional way and in a slower, deliberative cognitive-emotional way. There is a lot of evidence for this dual process model, although some psychologists still question its accuracy. It is easy to find everyday examples of automatic processes, such as driving an automobile, especially when one automatically drives toward one's workplace when the desired destination is in another direction. Many simple tasks and daily decisions are automatized because this simplifies our lives and reduces the time and energy needed for routine activities. In support of the SIM, Haidt and others have presented evidence that many of our moral judgments are quick and automatic. We generally don't have to think about whether undeserved rudeness or theft is wrong. It is difficult to argue against the automaticity thesis for many actions.

There are two additional, important elements in the SIM. The first is that Haidt and his colleagues have argued for at least five domains of moral intuition: care/harm, fairness/cheating, loyalty/betrayal, authority/subversion, and sanctity/degradation. That is, our moral behavior is directed by intuitions about caring for or not harming others of by intuitions about what is fair, and so forth. One of their most interesting claims is that each of these domains or "moral foundations" represents an adaption to evolutionary challenges. For example, humans have strong negative reactions to potential contaminants such as waste products, and this is adaptive because it reduces exposure to disease and pollution. Both the evolutionary account and the plurality of moral concerns strike me as strong elements of the SIM. From an evolutionary perspective, human morality is obviously a very important set of adaptations to facilitate social coordination and harmony. It seems equally obvious that human morality cannot be boiled down to a single set of concerns, such as pleasure or fairness, although both are important.

The second additional element of Haidt's SIM is that he prioritizes automatic emotional and intuitive responses over rational thought. In his famous article entitled "The Emotional Dog and Its Rational Tail" he argued that rational moral thought occurs primarily after one has already made a moral judgment and taken action. For Haidt, deliberative thought is mainly a rationalization that people use to justify their actions and convince others that they acted rightly. This strong prioritization has attracted significant criticism of the SIM. The most fundamental criticism is that intuitionism reduces moral reasoning to a social appeasement or influence process that has no other value. The SIM portrays people as acting on impulse and seeking to avoid social difficulties through verbal justifications. As Haidt (2013) has said, "The center of gravity [in research on morality] is now the study of automatic, intuitive and affective processes" (p. 294). Clearly, Haidt has significant scientific credibility due to his systematic theory, and a considerable amount of data that support his views, but questions about his reductionism remain.

If we take Haidt seriously, this means that the real work of morality occurs without conscious, deliberative activity. He adds that although post-action moral reasoning can affect others' moral intuitions, it rarely affects one's own intuitions or judgments. This social element in Haidt's thinking is often overlooked, but it is more than a little odd. According to his account, when individual A acts, it is based on an intuition, and A's reasoning plays little or no role until that action needs to be explained or justified to others. Thereupon, A makes up a story to justify the action, and the story that did not influence A's action can, nevertheless, influence B's intuitions and subsequent actions. Oddly, Haidt claims that I can influence you morally and you can influence me morally, but it is quite unlikely that either of us can guide our own decision making and actions with the rational grounds that can influence other people.

Clearly, there is some truth to the idea that moral judgments and actions often occur very quickly and do not require any deliberation. The question is whether *all* our moral actions can be described with this simple model. If Haidt is correct, there are three important moral functions that humans would be incapable of fulfilling.

The first incapacity would be that we could not intelligently judge how to act when more than one moral foundation is activated. There are many circumstances that would activate more than one moral foundation, and it is unclear how a person without the capacity to deliberate about the moral questions at issue could decide how to act. For example, how should one respond when parenting (authority) a toddler who has been playing with feces (sanctity/purity) and needs to be cleaned up (care)? Obviously, the first concern is to clean the child up, but to what degree should one express disgust? And how is authority involved? Will the child be able to understand and make use of direction and discipline regarding cleanliness? How does one best blend these three moral impulses? It seems likely that such complex moral situations would be at least as common as circumstances in which only one moral domain is evoked. It is easy to see how one could act based on any one of the three automatic processes alone, with a flash of authority, disgust, or care, but it's unlikely that fast, automatic, emotional processes can help us much when all three appear at the same time. Humans must be able to choose among and blend the moral considerations that show up in everyday circumstances. Does one just go with the strongest impulse? Or simply express 1/3 authority, 1/3 disgust, and 1/3 care, whatever that may mean? Clearly, we need to be able to sort things with a rational process or we could do no more than plunk for a course of action based on some emotional feel. In simple situations, that will work well enough, but it falls far short of mature morality in complex situations.

Second, Haidt's SIM has a sort of explanation for how important historical changes have emerged in human morality, such as the emergence of the prohibition of slavery, the equality of women and men, and equal treatment before the law. His model tells us that people convinced one another to move in these directions by telling post-action stories to justify their intuitive actions. The problem here is that these are monumental shifts from millennia of ingrained tribal loyalties, gender inequality, and status-based considerations of merit and transgression. It is easy enough to see how individuals could blunder into acting in ways that could be justified in terms of anti-slavery or equality rationalizations, but it is not clear why the listeners would be persuaded to change their judgments and behavior by such rationalizations. How did the ideas of human freedom and dignity take hold? What provided the leverage needed to fundamentally alter the narrow human interests typical of a small forager band mindset toward the concept of universal human rights and dignity? Do we really believe that made-up, self-justificatory stories could convince slave owners, chiefs, kings, and male heads of households to give up their power and privileges? I suggest, in contrast, that such fundamental shifts in what counts as moral required substantial, rational deliberation and argument, as well as genuine conviction. These elements are all dramatically absent from Haidt's model.

The third way that SIM is a reductive oversimplification is that it neglects the process by which many automatic intuitions become established. I think Haidt is correct in noting that humans have some basic moral instincts, whether those are evolved characteristics or the endowment of a creator. There is now a great deal of evidence, for example, that humans are deeply interested in justice, beginning as early as six months of age. (That is, before infants can talk.) But many cognitive, emotional, and behavioral processes become automatic intentionally, even following deliberation. Simple examples include driving or

showering, and a more complex example is cultivating inclusiveness toward people different from oneself. People decide they want to become better humans all the time. Often enough, we fail to improve ourselves, but success is not at all unusual in cultivating greater fairness, courage, generosity, or patience because that individual believes that it would be better to have these characteristics. With practice (frequently intentional), those qualities can become automatic responses to appropriate circumstances. One of Aristotle's fundamental ideas is that a characteristic (e.g., generosity) becomes a virtue when it becomes largely automatic and spontaneous through intentional practice and an understanding of the value of the characteristic. By leaving this intentional cultivation of moral qualities out of his model of morality, Haidt places an unreasonable and inhumane limit of what is morally possible for us. It is ironic that he believes that a moral theory should tell us that we are really less morally capable and intentional than we generally take ourselves to be.

Conclusion

The bottom line is that both Plomin and Haidt want to discredit the importance of reasoning and intentionality in human behavior, albeit in different ways. Unfortunately, they are in good company in attempting to degrade human rationality by arguing that our actions are guided by a genetic blueprint or by automatic moral instincts. Part of what makes these accounts credible is that there is some truth to both. Heredity is abundantly evident not just in eye color, but also in temperament and inclinations. Moral judgments and actions frequently occur quickly and without reflection because it is obvious to the actor what is at stake and what ought to be done. The simple, but deep mistake these theorists make is to overestimate the role of their favored mechanism and to attempt to rule out human

rationality and intentionality as constituents of human action. It is appropriate and important to recognize that there are limits to rationality and intentionality, but if we accept the virtually complete discrediting of human reasoning that these thinkers advocate, we will impoverish ourselves morally. (We might also note the irony of their carefully reasoned debunking of reason.)

Perhaps the most serious consequence of the attempt to discredit moral reasoning is that it can undermine our faith in and reliance on intellectual virtues such as practical wisdom. Aristotle placed practical wisdom (*phronesis*) at the center of moral action because he believed that acting morally was not primarily a matter of following rules. Rather, acting morally is a matter of promoting the goodness that is possible in the present situation. In other words, practical wisdom, or moral judgment, is part of every moral action. If, according to Plomin and Haidt human behavior is due either to inherited DNA or to automatic moral intuitions, there is no role for practical wisdom.

Ruling out the moral reasoning that practical wisdom makes possible is a critical problem because practical wisdom is the key ability for resolving two common and important moral difficulties. The first is that although most situations we encounter do not require deliberation, we occasionally find ourselves in confusing or very fraught circumstances that make it difficult to know what to do. When the best course of action is difficult to recognize, we must be able to reflect on what is most important in the situation and how we can do the most good. This reflection is one important function of practical wisdom. The second necessary function of practical wisdom is resolving the inevitable conflicts among moral inclinations noted previously. Practically wise individuals are better at recognizing which of several moral considerations is most important in a situation or at blending several moral concerns. In the example of the messy child, every parent would

know that care is called for in cleaning up the child, but only wise parents will know how to blend authority/discipline and disgust expression in a way that makes it a teachable moment (in contrast to a shaming or punitive moment). The proper blend will depend on situational factors and the age and temperament of the child. Both Plomin's and Haidt's views have no room for this kind of moral reasoning, which would unreasonably handicap us in responding well to complex or emotionally troubling situations.

We cannot afford to discredit the role of reason in human affairs, especially in a time where nationalism, tribalism, and intolerance are on the rise. Wisely resisting morally unacceptable social and political trends requires being able to reflect on what is wrong and why it is wrong and being able to systematically plan how we will combat these trends. Knee-jerk opposition to these trends will only inflame them, so we need to attend carefully to the moral concerns at issue and articulate them clearly. Plomin and Haidt do not offer any resources for this effort. We must look elsewhere. My favorite inspiration is Aristotle and the many contemporary theorists who are appropriating and expanding his ethical theory. There are others who can inspire us to be our best selves as well. However, clever debunking theories seem to be, we must keep in mind that moral guidance and inspiration are the real purposes of moral theory.

References

- Haidt, J. (2013). Moral psychology for the twenty-first century. *Journal of Moral Education, 42*, 281-297.
- Plomin, R. (2018). *Blueprint: How DNA makes us who we are*. Cambridge, MA: MIT Press.



THE
JUBILEE CENTRE
FOR CHARACTER & VIRTUES

www.jubileecentre.ac.uk

Insight Series

The Jubilee Centre for Character and Virtues

- ◆ *Pioneering interdisciplinary research of international standing focussing on character, virtues and values in the interest of human flourishing.*
- ◆ *Promoting a moral concept of character in order to explore the importance of virtue for public and professional life.*
- ◆ *A leading informant on policy and practice in this area through an extensive range of research and development projects contributing to a renewal of character and values in both individuals and societies.*

The views expressed in this article are those of the authors and do not necessarily reflect the official policy or position of The Jubilee Centre for Character and Virtues or The University of Birmingham.

Jubilee Centre for Character and Virtues

University of Birmingham | Edgbaston | | Birmingham | B15 2TT

www.jubileecentre.ac.uk