<table>
<thead>
<tr>
<th>Title</th>
<th>The definition of morality: threading the needle</th>
</tr>
</thead>
<tbody>
<tr>
<td>Author(s)</td>
<td>Luco, Andrés Carlos</td>
</tr>
<tr>
<td>Date</td>
<td>2014</td>
</tr>
<tr>
<td>URL</td>
<td><a href="http://hdl.handle.net/10220/19853">http://hdl.handle.net/10220/19853</a></td>
</tr>
</tbody>
</table>

**Rights**

© 2014 Social Theory and Practice. This is the author created version of a work that has been peer reviewed and accepted for publication by Social Theory and Practice: An International and Interdisciplinary Journal of Social Philosophy, Social Theory and Practice. It incorporates referee's comments but changes resulting from the publishing process, such as copyediting, structural formatting, may not be reflected in this document. The published version is available at: [DOI: http://dx.doi.org/10.5840/soctheorpract201440324].
The Definition of Morality: Threading the Needle

Abstract

This essay proposes and defends a descriptive definition of morality. Under this definition, a moral system is a system of rules, psychological states, and modes of character development which performs the function of enabling mutually beneficial social cooperation. I shall argue that the methodologies employed by two prominent moral psychologists rely upon to establish a descriptive definition of morality only serve to track patterns in people’s uses of moral terms. However, these methods at best reveal a nominal definition of morality, since moral appraisers may be ignorant about the references of their moral terms. I propose a real definition of morality which characterizes moral systems as a natural kind—more precisely, a copied kind. I explain what it takes for a moral system to satisfy this definition, and I identify the sorts of evidence needed to distinguish moral systems from value systems of other kinds.
1 “Morality” in the Descriptive Sense

Philosophers and moral psychologists have long been interested in articulating a *descriptive* definition of morality. A descriptive definition of morality would supply an answer to the question, “What is morality?” But in this question, the term “morality” is to be understood strictly in a descriptive sense. As Bernard Gert observes, “morality” is associated with both a descriptive and a normative sense.¹ “Morality” in the descriptive sense refers to a code of conduct actually endorsed by an individual or group. By contrast, “morality” in the normative sense refers to “a code of conduct that, given specified conditions, would be put forward by all rational persons.”² Different individuals and groups have endorsed different and incompatible codes of conduct. Accordingly, the term “morality” can be used in its descriptive sense to speak of the moralities of various persons or groups—e.g., Hitler’s morality, Nazi morality, the Pope’s morality, Catholic morality, Plato’s morality, ancient Greek morality, and so forth. These are descriptive rather than normative uses of the term “morality,” because they are not intended to imply that the codes of conduct endorsed by Hitler, the Pope, Plato, Nazis, Catholics, Greeks, etc. are endorsed *rationally*. Nor do descriptive uses of “morality” imply that anyone *should* or *should not* follow any particular code of conduct.

We should be interested in a descriptive definition of morality for two reasons. First, moral discourse appears throughout recorded human history.³ A better understanding of

---

² Gert, emphasis added.
human nature demands inquiry into what morality is, and why it is so ubiquitous. Second, it is widely believed that morality has authority over our lives. Morality is supposed to be a source of normative reasons as to why we should act, think, or feel one way rather than another. Of course, there is controversy about what exactly morality demands of us. A descriptive definition of morality would help to settle such controversy by specifying what sorts of requirements can be *candidates* for authoritative moral requirements in the first place. For instance: the normative ethical theory known as “ethical egoism” holds that a person’s only moral duty is to promote his or her own self-interest. But if a descriptive definition of morality reveals that *moral* duties are the sorts of requirements that demand the suppression of self-interest, then ethical egoism is ineligible for candidacy as a theory of the *moral* duties of persons.⁴ Indeed, according to the proposal defended in this article, one defining property of moral systems is that they cause people to transcend a single-minded concern with self-interest.⁵

In this essay, I argue that two prominent moral psychologists classify moral systems in a way that is either too broad or too narrow. Elliot Turiel and his students put forward a descriptive definition of morality that fails to count some genuine moral systems as such, making his approach to classifying morality too narrow. Jonathan Haidt and his students propose a descriptive definition of morality very similar to the one I endorse here. However, Haidt tends to misapply his own definition. He misapplies it in a way that classifies some value systems as moral systems, when they ought to be designated as *non-moral* value systems. This

---

⁴ Gert, §1.
makes Haidt’s method for classifying morality too broad. I propose that a correct descriptive
definition of morality must cite the proper function performed by moral systems—namely, to
promote mutually beneficial social cooperation. This functionalist definition is very similar to
Haidt’s. However, I argue that when properly applied, the functionalist definition “threads the
needle” in the sense that it is neither too broad nor too narrow in the way it distinguishes the
moral from the non-moral.

2 The Domain of Morality

A typical strategy for advancing a descriptive definition of morality is to delineate the
domain of morality. The domain of morality encompasses the social and psychological
phenomena that can be classified as moral, where this common classification is warranted by
certain shared features or some non-accidental tendency to occur together. Undoubtedly, as
Gert emphasizes, the domain of morality includes codes of conduct. Codes of conduct are sets
of behavior-governing rules that require or forbid various actions. Compliance with codes of
conduct is maintained by social practices involving reward, punishment, and internalization.6
Practices of reward, punishment, and internalization are, in turn, sustained by certain types of
psychological states—e.g., attitudes, judgments, motivations—and certain modes of character
development. That which falls into the domain of morality includes moral (as opposed to non-
moral) codes of conduct, moral (as opposed to non-moral) states of psychology, and moral (as
opposed to non-moral) modes of character development. Hereafter, I shall use the term “moral
system” to refer to an assemblage of moral codes of conduct, moral states of psychology, and

---

6 Chandra Sekhar Sripada and Stephen Stich, “A Framework for the Psychology of Norms,” in The Innate Mind:
Culture and Cognition, eds. Peter Carruthers, Stephen Lawrence, & Stephen Stich (Oxford: Oxford University Press,
2006), 280 – 301.
moral modes of character development. Further inquiry may reveal that moral systems include additional elements; the ones described here are merely salient in existing descriptive accounts of morality.

Not all systems of behavior-governing rules, psychological states, and modes of character development are moral systems. As we shall see, Elliot Turiel’s research on the distinction between morality and convention reveals that a substantial number of people regard moral systems to be a special case of a more general type. I will call this more general type a “value system.” Although all moral systems are value systems, not all value systems are moral systems. I argue below that the differences between moral systems and value systems are not grounded merely in intuition; ultimately, those differences have to do with a real difference in causal history.

Perhaps the best-known approach to demarcating the domain of morality relies on the moral/conventional task. This task is designed to test whether subjects distinguish between conventional rules and moral rules. Paradigmatic conventional rules include the rule that people should wear gender-specific clothing (e.g., men should not wear dresses), the rule that students should not chew gum in class, and the rule that all motorists should drive on the right instead of the left. Paradigmatic moral rules include rules that forbid harming other people, stealing, lying, or breaking promises. In the moral/conventional task, subjects are presented with vignettes describing transgressions of paradigmatic conventional rules and moral rules. They are then asked a series of questions intended to reveal whether or not they recognize any

---

7 Elliot Turiel, The Development of Knowledge (Cambridge: Cambridge University Press, 1983).
differences between the moral and conventional transgressions. After numerous trials with subjects of various ages and nationalities, psychologist Elliot Turiel and colleagues reported that people distinguish between moral and conventional rules in the following four ways:

(T1) moral transgressions are considered more serious than conventional transgressions;

(T2) moral rules, unlike conventional rules, are considered authority-independent—i.e., it would be wrong to violate a moral rule, even if an authority figure allowed it;

(T3) moral rules are regarded as universally applicable, whereas conventional rules only apply in specific times and places; and,

(T4) moral rules are thought to be justified on the grounds that they prevent harm, injustice, or the violation of someone's rights.

Responses (T1) through (T4) displayed a consistent and non-accidental tendency to occur together in subjects as young as 3 years old, and from diverse cultural backgrounds. For this reason, Turiel concludes that “the moral domain refers to prescriptive judgments of justice, rights, and welfare pertaining to how people ought to relate to each other.”

---

9 Ibid., 118 – 119.
10 Ibid., 119.
12 Turiel, 3.
However, other studies using the moral/conventional task have reported contradictory findings. In particular, Jonathan Haidt and colleagues showed that among low-socioeconomic subjects in Brazil and the United States, actions perceived to be harmless, but disgusting and disrespectful, were still judged to be wrong.\textsuperscript{13} The low-socioeconomic groups regarded actions such as masturbating with a dead chicken and cleaning a toilet bowl with the national flag to be \textit{moral} offenses. Indeed, these transgressions were viewed as serious, and the moral rules they violated were thought to be authority-independent and universally applicable.\textsuperscript{14}

These and other findings led Haidt to postulate that the domain of morality is not limited to concerns for fairness/justice, rights, and welfare.\textsuperscript{15} Haidt and his co-authors advance what they call a \textit{moral foundations theory}, according to which five “foundational concerns” constitute the domain of morality.\textsuperscript{16} These concerns are “foundational” in the sense that survey respondents consistently rate them as being highly relevant to their moral judgments and decisions. Also, respondents display a tendency to agree with statements that express the foundational concerns.\textsuperscript{17}

The five foundational concerns enumerated by Haidt are these:\textsuperscript{18}

\begin{itemize}
  \item \textit{Harm}: Actions that cause harm to others are considered immoral.
  \item \textit{Harm-avoidance}: Actions that are intended to avoid harm to oneself or others are considered moral.
  \item \textit{Purity}: Actions that promote or maintain purity are considered moral.
  \item \textit{Authority and Ingroup Favoritism}: Actions that support authority or group loyalties are considered moral.
  \item \textit{Liberty}: Actions that promote individual autonomy and freedom are considered moral.
\end{itemize}

\textsuperscript{14} Haidt, Koller, and Dias, 613; Machery and Mallon, 34.
\textsuperscript{18} Haidt and Graham, 385 – 386.
\textsuperscript{19} Haidt and Kesebir, 822; Haidt and Joseph, 382.
(H1) *Harm/care*: Concerns for the suffering and well-being of others.

(H2) *Fairness/reciprocity*: Concerns about unfair treatment, cheating, and more abstract notions of justice and rights.

(H3) *Ingroup/loyalty*: Concerns related to obligations of group membership, such as loyalty, self-sacrifice, and vigilance against betrayal.

(H4) *Authority/respect*: Concerns related to social order and obligations involved in hierarchical relationships, such as obedience, respect, and the fulfillment of role-based duties.

(H5) *Purity/sanctity*: Concerns about physical and spiritual contagion, including virtues of chastity, wholesomeness, and control of desires.

I concur with Haidt’s argument, *contra* Turiel, that the domain of morality is not limited to concerns for harm and fairness. However, while Turiel’s method of classifying moral systems errs in being too narrow, Haidt errs by classifying moral systems too broadly. Contrary to what Haidt suggests, moral systems are not indiscriminately associated with *any* concern for loyalty, deference, or purity. In sections 3 and 4, I detail what I take to be the proper method for formulating a descriptive definition of morality. I criticize both Turiel and Haidt for employing methods that are inadequate to the task of arriving at such a definition.

3 Searching for a Real Definition of Morality
Stephen Stich outlines two approaches to formulating a descriptive definition of morality.\(^\text{19}\) The first is linguistic or *conceptual analysis*. Conceptual analysis involves the attempt to understand how speakers use moral terms (or concepts) like “moral,” “immoral,” “right,” “wrong,” etc. Speakers have intuitions about the conditions under which these terms (concepts) do and do not apply. A successful definition of morality would give a complete and coherent account of these intuitive conditions of application.

The second approach to a descriptive definition of morality is the attempt to specify a *natural kind* to which speakers at least intend to refer by means of moral terms (concepts).\(^\text{20}\) A natural kind is a group of objects whose members share essential properties that display a non-accidental tendency to occur together.\(^\text{21}\) The essential properties of a natural kind are the properties which explain the existence of that kind, and which distinguish members of the natural kind from non-members. Moreover, natural kinds have *mind-independent existences*. This means that a natural kind exists irrespective of whether or not any particular individual believes it exists.

There is a crucial difference between conceptual analysis and the search for natural kinds. To see this, consider some garden-variety examples familiar to discussions of natural kind terms in the philosophy of language. Some terms and concepts do not refer to any natural kind. Gremlins do not constitute a natural kind; they are figments of the human imagination. But there is still the concept of a *gremlin*, along with the term “gremlin.” In some cases, terms and

\(^\text{19}\) Nado, Kelly, and Stich, 621 – 622; Stephen Stich, “Reply to Prinz,” in *Stich and His Critics*, ed. Dominic Murphy and Michael Bishop (Malden, MA: Blackwell Publishing Ltd., 2009), 221 – 222.


concepts are initially believed to refer to a natural kind, but subsequent investigations reveal that there is no such kind. These include the terms of outmoded scientific theories, such as “phlogiston” and “caloric.” Lavoisier thought that heat consists of caloric—a subtle fluid or gas capable of penetrating matter. It turned out, though, that there is no such thing as caloric, and that heat is really mean molecular kinetic energy. Hence caloric is not a natural kind, although it was once believed to be. In other cases, terms and concepts do refer to at least one natural kind, but subsequent investigation reveals that speakers have some mistaken beliefs about those kinds. In Ancient Greece and China, water was thought to be a fundamental element of nature, rather than a compound composed of more basic elements. The 19th century chemist John Dalton erroneously believed water to be the molecular compound HO, rather than H₂O. Nevertheless, Dalton and the ancients had terms and concepts that referred to water. Similarly, for thousands of years, human societies around the world possessed concepts and terms which they applied to gemstones that English-speakers still call “jade.” However, it was not until the 19th century when mineralogists discovered that jade was really made up of two different natural kinds of stone with different chemical compositions—jadeite and nephrite. Contrary to what speakers once believed, the term “jade” and its counterparts in other languages refers to not one, but two natural kinds.

The point, here, is that terms and concepts are elements of linguistic and mental activity. Natural kinds exist as part of a mind-independent reality. Though we often intend for our terms and concepts to refer to this mind-independent reality, they may fail to do so, or they may only be partially successful in the attempt. For this reason, conceptual analysis alone can at best be a source of nominal definitions. A nominal definition of a term (or concept) is
discovered by investigating the conditions under which speakers use and apply that term (or concept). On the other hand, inquiry into natural kinds can furnish real definitions. A real definition of a term (or concept) is discovered by investigating the nature of the things to which that term (or concept) refers. Specifying a natural kind designated by a term is one form of real definition. Tracking the ways in which speakers use and apply terms is not an effective strategy for finding real definitions, because speakers can be ignorant about the references of their terms. At the same time, discoveries of real definitions for terms like “jade” and “water” have proven to be interesting and informative. It would likewise be interesting and informative to discern whether morality (in the descriptive sense) is a mind-independent natural kind. Accordingly, psychologists and philosophers interested in a descriptive definition of morality should seek a real definition.

Just what sort of approach to a definition of morality is pursued by Turiel and Haidt? Stich suggests that Turiel’s project can be interpreted as an attempt to establish a real definition of morality—one which characterizes moral systems as a natural kind. The responses elicited by the moral/conventional task—i.e., T1 through T4 above—are postulated to be the essential properties of this natural kind. In other words, moral systems are said to be the natural kind consisting of rules that “are authority-independent, universally applicable, and justified by appeal to harm, justice, or rights.”

---

23 Brink, 159 – 161.
25 Ibid., 223.
However, I doubt that Turiel’s methodology can ground a real definition of morality. At best, the moral/conventional task reveals patterns in moral appraisers’ applications of terms like “wrong.” But even if it were a consistent pattern that experimental subjects apply the term “wrong” to instances of rule-transgressions they regard as relatively serious, authority-independent, universally applicable, and productive of harm or injustice, there is still a possibility that no single natural kind is designated by the term “wrong.”

It is likewise doubtful that Haidt’s methods will illuminate a real definition of morality. Haidt draws his main findings from a survey he devised: the “Moral Foundations Questionnaire.”26 Like the moral/conventional task, this survey only warrants classifications of what factors people believe to be relevant in making their moral judgments. But, once again, even if a handful of factors—factors like, harm, purity, loyalty, etc.—consistently shape the way that people use moral concepts, it doesn’t follow from this that there is any (single) natural kind represented by people’s moral judgments.

On the other hand, Haidt does at least propose a real definition of morality. In what follows, I will argue that this definition is correct. Moreover, Haidt’s survey findings have been indispensable clues toward the discovery of moral systems qua natural kind. By consulting people’s intuitions about what sorts of actions are right versus wrong, we can acquire information about the referential intentions of speakers—that is, we learn what it is that speakers intend to refer to when they use and apply moral terms.27 Once the referential intentions which govern the use of moral terms are accounted for, investigators could then

---

26 Haidt and Graham, 385.
27 Brink, 170 – 172; Stich, “Reply to Prinz,” 222.
search for a natural kind that would qualify as the intended reference of moral terms. Haidt’s Moral Foundations Questionnaire offers a rich source of data about the referential intentions associated with moral terms. And indeed, it was a significant finding on his part that people in some cultures or subcultures regularly apply moral terms not just to rules pertaining to fairness and the prevention of harm, but also to rules about loyalty, authority, and purity/sanctity. Those five types of rules must be counted among the intended references of moral terms. And we should expect from any adequate characterization of morality as a natural kind that it be able to explain why people’s moral judgments are sensitive to those rule-types. The main reason that I broadly endorse Haidt’s proposed real definition of morality is that it can plausibly explain why so many people’s moral judgments are concerned with loyalty, authority, and purity/sanctity.

That being said, Haidt’s moral foundations theory suffers from two major limitations, which I elucidate in sections 4 through 7. First, Haidt misapplies his own definition of morality by being too permissive in the way he classifies subjects’ reported normative judgments as moral judgments. Indeed, Haidt’s own definition of morality requires narrower classifications. Second, much of Haidt’s empirical data—particularly the data drawn from the Moral Foundations Questionnaire—are insufficient to support his own proposed real definition of morality.

4 Haidt’s Evolutionary-Functionalist Account of Morality

Haidt’s moral foundations theory does offer a picture of morality as a natural kind. Therefore, it proposes a real definition of morality. A major attraction of this definition is that it
offers a plausible and testable explanation for the existence of moral systems. Haidt’s theory holds that moral systems evolved because they performed the function of securing mutually beneficial social cooperation. Accordingly, Haidt writes that humanity’s “moral nature” consists of

ancient ‘inside the head’ mechanisms...that co-evolved in more recent times with ‘outside the head’ cultural creations (such as law, religion, and political institutions), to serve the function of suppressing selfishness and increasing group cohesion, trust, and coordinated action.\(^{28}\)

On the basis of this evolutionary-functionalist account of morality, Haidt offers the following real definition of moral systems:

Moral systems are interlocking sets of values, virtues, norms, practices, identities, institutions, technologies, and evolved psychological mechanisms that work together to suppress or regulate selfishness and make cooperative societies possible.\(^{29}\)

According to Haidt, a moral system is a natural kind constituted by codes of conduct, psychological states, and modes of character development that are reproduced and transmitted because they perform a function. Haidt argues that moral systems have the

\(^{28}\) Haidt and Kesebir, 815, emphasis added.

\(^{29}\) Haidt, *The Righteous Mind*, 314.
function of suppressing selfishness and enabling mutually beneficial social cooperation. Moral systems perform this function when they successfully resolve the five foundational concerns—H1 through H5 listed above. The foundational concerns are described as innate psychological predispositions, or “modules,” that motivate agents to resolve recurrent “adaptive challenges” to the longevity of both individuals and social groups. According to Haidt and Graham, the harm/care concern motivates individuals to meet the adaptive challenge of protecting and caring for vulnerable kin. The fairness/reciprocity concern motivates individuals to acquire the benefits of cooperation with non-kin, while guarding against free-riders. The ingroup/loyalty concern motivates individuals to acquire the benefits of cooperation in large groups, particularly in the context of intergroup competition. The authority/respect concern motivates individuals to acquire the benefits of living in a stable and orderly social hierarchy. And, finally, the purity/sanctity concern motivates individuals to avoid potentially deadly microbes and parasites.

When a social group institutes a moral system that successfully addresses the foundational concerns, it will be more effective at engaging in collectively beneficial cooperation than social groups that lack such a system. Inter alia, groups with well-functioning moral systems would be able to out-compete other groups in military conflict or in economic performance. They would be able to sustain a common resource—e.g., a publicly available food source or dwelling—while protecting it from being usurped by selfish free-riders.

---

30 Haidt and Joseph, 381 – 385; Haidt and Kesebir, 816 – 818; Haidt and Graham, 381.
And, they would be able to share knowledge of strategies for coping with threats in the surrounding environment, like predators, diseases, and dangerous substances.

5 What a Moral System is Not

At this stage, I argue that Haidt’s classifications of what counts as a moral system are too broad and permissive. Indeed, I suggest that the way Haidt marks out the boundaries of morality is inconsistent with the real definition of morality that he himself proposes. In particular, Haidt often suggests that any value system that can be interpreted as pertaining to any one of the five foundational concerns is a moral system. But even if a value system can be interpreted in this way, it may not have the function of enabling mutually beneficial social cooperation. And if a value system doesn’t have this function, then under Haidt’s own real definition of morality, it should not qualify as a moral system.

Haidt suggests that taboos are moral systems designed to address the purity/sanctity concern. Taboos can motivate people to stay away from foods or people that carry harmful contagions. On the other hand, anthropologists have found that some taboos have no such function. As a case in point, the anthropologist L.L. Langness reported that the Bena Bena people of highland Papua New Guinea refused to eat chicken and eggs because, they said, chickens eat feces. This was so despite the fact the Bena Bena’s favorite food—pig—also eat feces. Langness could find no rationale for the taboo that was related to the avoidance of

32 Haidt and Joseph, 382, 384.
contagion. Indeed, the taboo existed in spite of having positively harmful health effects, since the Bena Bena were chemically deficient in protein.\textsuperscript{34}

Next, consider the ingroup/loyalty foundation. Again, Haidt indiscriminately counts any rule or judgment which mandates the virtue of loyalty to one’s social group as a \textit{moral} rule or judgment.\textsuperscript{35} But some demands for ingroup loyalty can be a cause of social \textit{breakdown}, not cooperation. Between 1900 and 1950 the Mae Enga—another group in Papua New Guinea—killed each other in inter-clan warfare at the astonishing rate of about 300 deaths per 100,000 living people per year.\textsuperscript{36} Beginning in the late 1930s, the Australian government was able to reduce violence among the Enga. But in 1975, when Papua New Guinea gained independence from Australia, warring among the Enga returned. Officials in the newly independent government doled out land and favors to their clansmen. Other clans resented this favoritism and exacted vengeance. According to anthropologist Polly Wiessner,

[w]arring clans employed mercenaries skilled in the use of semiautomatic weapons, called ‘Rambos’ or ‘hiremen,’ who fought largely for brotherhood and fame and to ‘carry out justice where the government had failed.’ In the late 1990s, warriors formed ‘teams’ of peers led by Rambos that cross-cut clan boundaries, similar to some inner-city gangs. Men who had never cooperated before were bonded through oaths of loyalty, rituals of confession, traditional fight magic, and victory celebrations.\textsuperscript{37}

\textsuperscript{34} Edgerton, 56.
\textsuperscript{35} Haidt and Joseph, 382 – 383.
Wiessner and her colleagues conducted extensive interviews with these “Rambos,” in which they were asked to explain what motivated them to practice such rampant warfare:

All of them [i.e., the Rambos] began their careers by fighting for their own clans with bows and arrows or shotguns to avenge injustices to clan members...

Most Rambos said that they became caught up in the excitement of fighting in their initial year, reveling in the praise they received from kin...From there, they sought other wars, whether as allied kin or local mercenaries...They used most of their earnings to help clan members with school fees, hospital bills, and other needs. 38

The Mae Enga “Rambos” were driven to the infliction wanton bloodshed. Their motives were diverse, but among those motives was a sense of loyalty to the in-group—be it clan or gang. (Eventually, beginning in the early 2000s, members of the community in the Enga Province—often women, elders, and clergy—started campaigning to restore peace and security. Deaths by homicide in Papua New Guinea subsequently declined between 2001 and 2009. 39)

As a last case in point, consider Haidt’s foundational concern for authority/respect:

---

39 Wiessner & Pupu, 1652.
The current triggers of the Authority/subversion foundation, therefore, include *anything that is construed as* an act of obedience, disobedience, respect, disrespect, submission, or rebellion, *with regard to authorities perceived to be legitimate.* Current triggers also include acts that are seen to subvert the traditions, institutions, or values that are *perceived to provide stability.*

Under Haidt’s moral foundations theory, moral systems would include any rules that impart obligations of obedience, respect, and deference toward higher-status individuals regarded as having legitimate authority.

However, some rules that require obedience to authority contravene the function Haidt ascribes to moral systems. The Hindu caste system in India obligates lower-caste individuals to subordinate themselves to the higher castes. The caste system is widely perceived to be legitimate by members of upper castes. But even members of relatively low-ranking castes invoke the caste ideology to explain their superiority over castes ranked beneath *them.* In this way, remarkably, the various low-caste “Untouchable” groups accept the caste system as a whole—particularly the station of other Untouchable groups deemed lower than they are in the hierarchical order. Untouchables only object to their own group’s lowly status. Robert Deliège describes the tragically ambivalent attitudes of Untouchables toward the caste structure:

---

40 Haidt, *The Righteous Mind,* 168 (emphasis added).
...an Untouchable will refer to caste ideology for the purpose of explaining the degradation of the castes below him, but he will never use these values to explain his own status or to legitimize the position of his caste.42

...Untouchables, then, do not form a homogeneous category capable of opposing the system...their resentment is not directed against the [caste] system as such, but rather against their own position within it. They therefore do not try so much to eliminate the system as to improve their place within it.43

Although the caste system is generally perceived to be legitimate, even by those most disadvantaged by it, it is also a cause of great social strife. Although the Indian constitution guarantees equality before the law to all citizens, efforts by Untouchables to assert their legal rights have provoked violent backlashes from people in the higher castes. According to a 1999 report published by Human Rights Watch, “between 1994 and 1996, a total of 98,349 cases were registered with the police nationwide as crimes and atrocities against scheduled castes.”44 In the Indian state of Bihar, caste militias have embroiled themselves in bloody feuds for 30 years. Hundreds of non-combatant Untouchables have been caught in the line of fire, or killed in massacres.45 One high-caste militia, the Ranvir Sena, massacred more than 400 Untouchable villagers in Bihar between 1995 and 1999.46

42 Ibid., 62.
43 Ibid., 63.
46 Human Rights Watch, Broken People, §1.
Caste-based violence is triggered by the authority/respect foundation, as Haidt describes it. It is a concerted effort by members of the upper castes to maintain traditional social hierarchies that allow the upper castes to monopolize land, education, and political power, while denying these goods to Untouchables. Violence is only the most extreme form of intolerance toward government measures to eliminate caste discrimination. Many people in the upper castes feel entitled to the submission of the Untouchables, and are willing to resort to brutality as a means of enforcing their privilege.47

Haidt emphasizes that moral systems function to enable social groups to engage in collectively beneficial forms of economic cooperation related to agriculture, trade, and infrastructure development.48 From this perspective, it becomes even harder to fathom how the Hindu caste system could qualify as a moral system. Economic historian E.L. Jones enumerates several economic inefficiencies inherent in the caste system:

The result of this immensely complicated anthropological pattern was to push the living-standards of the lowest caste to a level that presumably reduced the capacity to work; allocate functions on a basis of heredity, not aptitude; instill ritualistic attitudes to work; restrict the market through caste-defined sumptuary rules; and divide the community and reduce the chances of mutual evasion or resistance to exploitation...this inserting of

47 Ibid., §III.
48 Haidt and Kesebir, 809.
an artificial ‘individual distance’ between persons suppressed social interaction and competition that in the European case, at any rate, has proved energizing.\textsuperscript{49}

Here we find yet another respect in which the caste system, despite being a value system consisting of obligations to defer to higher-status individuals, militates against the cooperative function characteristic of moral systems. Although it has been suggested that the caste system may have afforded some economic benefits—for instance, as a means of contract enforcement\textsuperscript{50}—it’s extremely unlikely that these benefits would ever have offset the many costs outlined by Jones. First of all, contracts can be effectively enforced in societies not stratified by hereditary castes. Moreover, economists have learned much about the conditions under which economies prosper or lapse into poverty. In a wide-ranging comparative-historical analysis of the causes of national wealth and poverty, Daron Acemoglu and James A. Robinson argue that wealthy countries around the world owe their success to “inclusive” institutions\textsuperscript{51}. These institutions include the impartial rule of law, which denies any single group the privilege of using the power of the state to further its own narrow interests, and which ensures equal opportunities for people to participate in economic activity. In short, inclusive institutions—the keys to economic success—permit exactly the social equality and freedom that the Hindu caste system stifles. Meanwhile, Acemoglu and Robinson explain the economic stagnation experienced by many impoverished societies to be the result of “extractive” economic and


\textsuperscript{50} Kripa Freitas, “Essays on the Institutions of Caste and Dowry” (PhD diss., Northwestern University, 2008), 48 – 147.

political institutions. Under these institutions, most people in society lack the freedom to choose their vocation and the resources to develop their talents. Entrepreneurs are blocked from establishing new businesses and entering the market, and arbitrary power is concentrated into the hands a small elite. Acemoglu and Robinson explicitly identify the Hindu caste system as an extractive institution.\textsuperscript{52}

We have surveyed a series of value systems that pertain to one of Haidt’s five foundational moral concerns. In Haidt’s view, they would therefore qualify as moral systems. However, Haidt does not correctly apply his own real definition of moral systems. According to Haidt’s definition, a moral system is a set of “values, virtues, norms,” etc. that has the function of enabling mutually beneficial social cooperation. But some food taboos, norms of group loyalty, and social hierarchies exist in spite of going against that function. Such value systems cannot be moral systems.

6 Interdependencies among the Five Foundations

While Haidt often includes too much in the domain of morality, he is occasionally more circumspect. In particular, he is relatively strict about what he counts as moral judgments that can be based in the foundational concern for authority/respect. This strictness, I argue, is in fact required by Haidt’s real definition of morality, and it should be applied consistently in classifying moral systems. To exercise the necessary strictness, we must appreciate that nothing can count as an element of a moral system if it systematically undermines any one of the five foundational concerns. Moral rules, moral judgments, moral virtues, and other elements of a genuine moral

\textsuperscript{52} Ibid., 117 – 118.
system must address the foundational concerns in a way that is sensitive to interdependencies among those concerns. Moreover, what explains the interdependencies among the foundational concerns is that they all aim at securing a common outcome: namely, mutually beneficial cooperation. Thus, Haidt’s real definition of morality requires that something be counted as an element of a moral system only if it has some effect in promoting mutually beneficial cooperation. This is the claim for which I argue in the present section and section 7.

According to Haidt, moral judgments grounded in the authority/respect foundation are triggered by disrespect or insubordination toward an authority perceived by the appraiser to be legitimate. For instance, in the Moral Foundations Questionnaire, Haidt gauged the extent to which respondents’ moral judgments were based on the authority/respect concern by asking them if they agreed with this statement: “If I were a soldier and disagreed with my commanding officer’s orders, I would obey anyway because that is my duty.” Presumably Haidt took an “agree” response to indicate a moral judgment elicited by the authority/respect concern. The idea seems to be that the respondent would not agree it was his or her duty to comply with the commander’s order, unless the respondent regarded the commander’s order to be legitimate.

But Haidt introduces another set of moral judgments that are also motivated by the authority/respect concern. These include moral judgments to the effect that a person wields authority illegitimately, as a consequence of violating the duties attached to his or her superior position. Haidt explains that moral judgments based on the authority/respect foundation may apply to superiors whose conduct is deemed to have violated conditions for legitimate authority.

53 Haidt & Graham, 385 – 386.
...Authority Ranking is a two-way street: Subordinates must show respect and deference, but superiors must then protect them from external threats and maintain order within the group...In societies that value authority...norms and related virtues govern the behavior of superiors (e.g., impartiality, magnanimity, fatherliness) and subordinates (e.g., respect, deference).54

Above, Haidt implies that the perceived legitimacy of authorities depends on their success in meeting some specific expectations: protecting subordinates, maintaining order, and exercising virtues of impartiality, “magnanimity,” and “fatherliness.” Notice, however, that these conditions for perceived legitimacy sound closely connected to other foundational concerns in Haidt’s theory. In particular, to be perceived as legitimate, authorities must also be thought to protect subordinates from harm, to practice impartiality, and to be magnanimous. Accordingly, moral judgments about the legitimacy of authorities are dependent upon additional foundational concerns, including the harm/care concern and the fairness/reciprocity concern. Moral judgments motivated by one foundational concern (e.g., authority/respect) may also be motivated by other foundational concerns (e.g., harm/care and fairness/reciprocity).

Now why would such interdependencies among the foundational concerns exist? Why, for instance, would impartiality, magnanimity, and protecting subordinates from harm be conditions for perceived legitimate authority? Haidt explicitly suggests that this is no accident:

54 Haidt and Joseph, 384 (emphases added).
Human authority, then, is not just raw power backed by the threat of force. Human authorities take on responsibility for maintaining order and justice...if we want to understand how civilizations burst forth and covered the Earth in just a few thousand years, we’ll have to look closely at the role of authority in creating moral order.55

The Authority foundation, as I describe it, is...more complex than the other foundations because its modules must look in two directions—up toward superiors and down toward subordinates. These modules work together to help individuals meet the adaptive challenge of forging beneficial relationships within hierarchies. We are the descendants of the individuals who were best able to...rise in status while cultivating the protection of superiors and the allegiance of subordinates.56

It’s just been noted that, in Haidt’s view, the “upward looking” module of the authority/respect concern is dependent on other foundational concerns for care/harm prevention and fairness/reciprocity. The above passages give a suitable explanation for this interdependency: hierarchical relationships between authorities and subordinates tend to be more mutually beneficial when authority figures maintain order, provide protection, and treat subordinates fairly. Moreover, modern humans inherited the sensibilities of ancestors who had the best success in creating hierarchical relationships that were mutually beneficial. This is precisely why

55 Haidt, The Righteous Mind, 167 (emphasis added).
56 Ibid., 168 (emphasis added).
Haidt takes the five foundational concerns to be innate modules within human moral psychology.

The interdependency between authority/respect concerns and other foundational concerns reflect a constraint on what can be classified as a moral system—at least under Haidt’s moral foundations theory. The theory allows that moral systems can comprise rules, judgments, virtues, etc. that address authority/respect concerns related to the maintenance of a hierarchical social order. But from the passages just surveyed, it is also clear that authority/respect concerns are not arbitrarily related to any sort of hierarchical social order. Instead, authority/respect concerns strictly pertain to hierarchies that are, to some extent, mutually beneficial for both superiors and subordinates. Recall from section 4 that, according to Haidt’s definition of morality, moral systems have the function of fostering mutually beneficial social cooperation in human groups. Furthermore, moral systems perform this function by addressing the foundational concerns; indeed, moral systems exist because they execute this function. On the other hand, if a value system does not have the function of enabling mutual cooperation, then under Haidt’s definition of morality, it cannot qualify as a moral system. And surely some value systems lack this function, since they do not exist in virtue of facilitating mutually beneficial cooperation. The Hindu caste system, for instance, is a value system which exists in spite of systematically militating against mutually beneficial social cooperation. This can be observed in the civil violence and economic inefficiencies generated by the caste system. Additionally, some power asymmetries can block responses to the adaptive challenges which Haidt associates with the harm/care and justice/reciprocity concerns. Moral systems that address the harm/care concerns enable the protection and care of kin, while moral systems
that address the justice/reciprocity concern enable us to reap the benefits of cooperation with non-kin.\textsuperscript{57} The violence and discrimination wrought by the caste system thwarts solutions to both these adaptive challenges.

In sum, the constraints Haidt places on what counts as a moral system grounded in authority/respect concerns are in fact \textit{required} by Haidt’s proposed real definition of morality. Haidt should have been more consistent by imposing similar constraints on what is to count as a moral system grounded in \textit{the other} foundational concerns, as well. He should have affirmed that no value system can count as a moral system if it has the effect of frustrating \textit{any one} of the five foundational concerns.

\textbf{7 Moral Systems as Copied Kinds}

I have criticized Haidt for interpreting his survey data in a way that is inconsistent with the real definition of morality proposed in his own theory. My second criticism of Haidt is that his survey data do not adequately support this proposed definition. As we saw in section 4, Haidt proposes an evolutionary-functionalist account of morality. This account characterizes moral systems as entities that have been reproduced and transmitted because of their tendency to produce certain effects—namely, enabling mutually beneficial social cooperation. But in order to support this hypothesis, Haidt cannot merely rely on survey data which record what factors people now consider relevant to their moral judgments. Rather, Haidt must also draw on historical evidence showing that moral systems really have emerged and persisted in virtue of enabling mutually beneficial cooperation. To his credit, Haidt \textit{has} provided some

\footnote{Haidt and Joseph, 382.}
evidence in this regard, though I expect he would agree that more is needed. In this section, I will take note of the historical evidence Haidt has managed to adduce in favor of his real definition of morality. I will also identify additional sources of evidence, and I will explain how an examination of the history of moral systems would support Haidt’s real definition of morality.

Proving a real definition of morality requires demonstrating that moral systems are a natural kind. I contend that moral systems are what Crawford Elder has called a “copied kind.”\(^{58}\)

**Copied kinds** are a subset of natural kinds. Members of copied kinds have properties in common that non-accidentally occur together due to a *common history of function*. According to Elder, copied kinds “*include* many kinds of artifacts, but more besides: kinds of biological devices, kinds of naturally selected behaviors (e.g., mating dances), kinds of customary performances (e.g., rain dances), and kinds of linguistic structure.”\(^ {59}\) Three essential properties are distinctive of copied kinds.\(^ {60}\) First, members of copied kinds all have a *qualitative “shape”*—i.e., a structure or pattern. Bodily organs like hearts and eyes have a common structure. Animal mating dances and cultural rituals have a common pattern that makes instances of them recognizable as instances of the same dance or ritual. Second, the members of a copied kind have a *proper function*. The proper function of members of a copied kind is an effect of those members’ properties which causes new members of the same kind to be copied (i.e., designed, selected, or reproduced) over time. The new members of the copied kind will bear the same properties as previously existing members they were copied from, and the effect (i.e., the

---

\(^{58}\) Elder, “On the Place of Artifacts in Ontology.”

\(^{59}\) Ibid., 35.

\(^{60}\) Ibid., 38 – 39.
proper function) of those properties subsequently causes the copying of more members of the same copied kind. To take some classic examples, hearts have the proper function of circulating blood, as this is the effect of hearts which caused them to be reproduced in organisms over time. The proper function of watches is to indicate the time, since that is the effect which historically caused watches to be continually designed, manufactured, and sold. Third, the members of a copied kind have a *historically proper placement*—they are copied or reproduced over time in conjunction with other copied kinds. Thus, the eye was reproduced in seeing organisms in conjunction with brains that were capable of interpreting light stimuli. Automobiles were designed, manufactured, and sold in conjunction with roads on which they could be driven.

In sum, the essential properties of a copied kind are fixed by specific combinations of shape, proper function, and historical placement. Haidt’s real definition of morality is consistent with an account of morality as a copied kind. First, moral systems have a qualitative “shape.” They are all composed of practices of reward and punishment, behavior-governing rules, prescriptive judgments, and modes of character development. Second, moral systems have a proper function. That function has been identified by Haidt and other evolutionary theorists as enabling mutually beneficial social cooperation. It is because moral systems have made possible mutually beneficial cooperation in the past that human beings continue to institute such systems. Furthermore, the benefits of cooperation also explain why human beings have

---

developed innate psychological mechanisms that make it easier for us to design and comply with moral systems. Haidt argues that the five foundational concerns are among such innate mechanisms. Finally, moral systems have a historically proper placement in human evolution. For instance, Haidt suggests that moral systems could not have evolved without the human capacity for shared intentionality. Shared intentionality is the ability to share mental representations of tasks with others. For instance, a team of hunters can share a plan for bringing down a large prey, and they all can understand each team member’s role in executing the plan. Shared intentionality would have been necessary for our ancestors to cooperate and coordinate their activities for a common purpose.\footnote{Haidt, \textit{The Righteous Mind}, 237 – 243.}

The foregoing account of morality as a copied kind provides a basis on which to distinguish moral systems from other value systems. I’ve maintained (in sections 5 and 6) that nothing can count as a moral system, nor as an element of a moral system, if it does not have the proper function of enabling mutually beneficial social cooperation. In order to tell whether a system of social activity is a moral system, we need historical evidence indicating what caused that system to persist (i.e., be copied or reproduced) over time. If the cause of persistence was not the promotion of mutually beneficial cooperation, then we do not have a \textit{moral} system on our hands.

In order to identify a moral system by means of the method I’m proposing, we must pin down what is meant by “mutually beneficial cooperation.” At a minimum, cooperation happens whenever disparate individuals work together towards a shared end. Cooperation can occur among humans, animals, cells, and even computers. It can take many forms. Sometimes,
cooperation yields a net benefit for the cooperating individual—this is called *mutualism*. For instance, when a team of hunters kills an elephant, each gets more meat than he could have acquired by hunting alone. In other situations, cooperation yields a net cost for the individual—this is called *altruism*. People who cooperate to advance a charitable cause sacrifice time, energy, and money that they could have used for their own benefit. The “benefits” and “costs” of cooperation can be denominated in different currencies, depending on the nature of the cooperating organisms and the theoretical interests of the researcher. One form of “benefit” associated with cooperation is biological in nature, and measured in terms of differential reproductive success. In humans and other sentient animals, there are psychological benefits and costs associated with cooperation—e.g., pleasure/pain, or the satisfaction/dissatisfaction of desire.

*Mutually beneficial* cooperation is cooperation that benefits all cooperating agents, in some sense. One of the great paradoxical findings of modern evolutionary theory is that altruism can be mutually beneficial. Although altruism is, by definition, costly for the individual, groups of altruists are better off than groups of individuals that do not engage in altruism. Consequently, groups of altruists may outcompete, outlast, or out-reproduce groups of non-altruists. This is a process known as *group selection*. Haidt explicitly identifies group selection

---

as a mechanism that explains the evolution of moral systems, and he makes the case that group selection really was a driving force in human evolutionary history.\(^67\)

In human morality, individuals cooperate by maintaining compliance with the rules of the moral system. They teach the rules to children, and they punish those who do not abide by the rules. This cooperation can be either mutualistic or altruistic. But what sorts of mutual benefits come from cooperatively participating in moral systems? One possibility is success in military competition.\(^68\) Another is success in economic trade and production.\(^69\) Yet another is success in creating a store of knowledge relevant to coping in the surrounding environment—taboos forbidding contact with genuinely harmful substances would be a case in point.\(^70\)

Haidt suggests that all these forms of mutually beneficial cooperation explain the historical emergence and persistence of moral systems.\(^71\) Moral systems are instituted to motivate cooperative action in response to recurring “adaptive challenges” faced by all human societies. When cooperative action succeeds in resolving the adaptive challenges, the mutual benefits of cooperation are reaped. When a moral system succeeds in fostering cooperative solutions to adaptive challenges, it is more likely to persist over time. Human groups with successful moral systems are in a position to out-compete other groups. They may also be imitated by other groups.

However, some value systems are not moral systems because they do not have the same history of function as moral systems. As noted in section 5, some food taboos do not

\(^{67}\) Haidt, The Righteous Mind, chapter 9.  
\(^{68}\) Sober and Wilson, 186 – 191.  
\(^{69}\) Kitcher, §19; Haidt and Kesebir, 809.  
perform the function of enabling mutually beneficial cooperation in response to the adaptive challenge of avoiding harmful microbes. Indeed, some food taboos—such as those of the Bena Bena—are harmful because they forbid the eating of beneficial foods.

What then, could explain the historical persistence of harmful food taboos? To answer this question, we must appreciate that food taboos, moral systems, and value systems in general, are all cultural variants. Culture is information stored in human minds and transmitted from one person’s mind to another through social learning. Social learning is the process by which a mental state in one person causes a similar mental state in another person. A cultural variant, then, is a unit of cultural information. According to anthropologists Robert Boyd and Peter J. Richerson, the social learning of cultural variants can be biased “when people preferentially adopt some cultural variants rather than others.” Biases on social learning may be innate preferences, or they may be preferences acquired from an earlier episode of social learning. Some biases are direct, while others are indirect. In direct bias, individuals deliberately evaluate different cultural variants, and select the one that they most prefer. In indirect bias, individuals select a cultural variant in a relatively automatic way, based on the presence of some characteristic. One type of indirect bias is called prestige bias. This bias occurs when a cultural variant is selected by an individual because it was selected by another individual who enjoys great social prestige. Another type of indirect bias is called conformist

---

74 Richerson and Boyd, 63.
75 Ibid., 68 – 69.
76 Ibid., 66.
bias. This occurs when a cultural variant is selected by an individual because it is the one that is most common or popular in one’s surrounding environment.

Chandra Sekhar Sripada and Stephen Stich point out that indirect biases on social learning can result in “maladaptive” or harmful cultural variants. To substantiate this point, Sripada and Stich concentrate on food taboos that attach to beneficial foods. Such a taboo could easily originate when a prestigious individual falls ill after ingesting what is normally a beneficial food. This prestigious individual could develop a strong, permanent aversion (a “Garcia aversion”) to the food. Prestige bias could then take effect, as other individuals decide to imitate the prestigious individual’s avoidance of the beneficial food. As the aversion to the beneficial food becomes more common within a society, conformist bias could cause even more people to adopt it. Eventually, a taboo which codes the food as a profane and forbidden may be established. When hardly anyone dares to transgress the taboo, the benefits of the prohibited food may be forgotten.

It is not that taboo systems that emerge and persist in this fashion have the function of facilitating cooperation, but just perform it in a suboptimal way. Rather, these taboos do not have the proper function of enabling beneficial cooperation at all. In contrast, the human birth canal clearly has the function of allowing newborns to pass from the uterus of a mother to the world outside. Ancestor mothers with birth canals that had this effect would have out-reproduced mothers who lacked them. Nevertheless, the normal average birth canal is exceedingly narrow, causing mortal risks to both mother and child. Thus, the human birth canal performs its function sub-optimally. On the other hand, if Sripada and Stich’s account of taboos

78 Ibid., 151, 153 – 155.
against beneficial foods is correct, these taboos do not even persist by sub-optimally causing us to avoid harmful microbes. Rather, the explanation of their persistence has nothing whatsoever to do with avoidance of harmful microbes. Hence, taboos of this kind lack any beneficial cooperative function, and cannot be elements of moral systems.

The main lesson, here, is that moral systems can be distinguished from non-moral value systems by examining the processes that caused them to emerge and persist over time. Moral systems are a copied kind, and one of their essential properties is a common history of function. Moral systems have the proper function of enabling mutually beneficial cooperation, because the performance of this function is what explains why moral systems emerged and persisted. Non-moral value systems, such as harmful food taboos, do not emerge and persist in virtue of enabling beneficial cooperation. Therefore, they do not have the same history of function as moral systems, and they cannot be moral systems.

It's been suggested that the correct method for classifying value systems as moral systems should begin with histories of function. The experimental and survey data that Haidt adduces in favor of his moral foundations theory do not include any specific histories of function for any value system that Haidt counts as a moral system. This is a major omission.

In sum, an accurate delineation of the moral domain must take histories of function into account. However, when using histories of function to distinguish moral systems from other value systems, a serious ambiguity must be dealt with. In the remainder of this section, I suggest a way that this ambiguity can be overcome.

It may be that, in some past period, a value system once did perform the function of enabling beneficial social cooperation, but does not do so anymore. In that case, a value system
may be a moral system for some time, but eventually ceases to be a moral system. This sort of transformation also occurs in biological organisms. For instance, an appendage of one type may gradually become an appendage of another type as it loses or acquires functions. Millions of years ago, the ancestors of penguins had wings adapted for flight. But over time, the creatures’ appendages became flippers adapted for efficient diving. It was recently discovered that appendages adapted for diving tend not to be suitable for flight. Modern penguins have dense, thick bones which makes them less buoyant. This is helpful for diving. However, penguins’ bones are much too dense for flight. Thus, while ancestors of penguins had wings, the function of which was to fly, penguins have flippers, the function of which is not to fly but rather to dive. As the appendages of this lineage of organisms lost one function and took on a new one, the appendages changed in kind. Again, I suggest that a similar change can take place when it comes to moral systems. At some point in time, a value system may have the function of enabling beneficial cooperation, in which case the value system was a moral system during that time. But if a value system loses this function, it cannot be a moral system.

Now we confront a source of ambiguity in distinguishing moral systems from other value systems. At what point can a value system stop or start to be a moral system? I suggest that this ambiguity can be cleared up through historical analysis. Such investigation should aim at understanding the mechanisms that cause the emergence and persistence of a value system. If that mechanism does not currently, or at least recently, involve the performance of the function of enabling beneficial social cooperation, then the value system in question is not a moral system.

---

moral system. A case in point would be the “culture of honor” in the American South. Cultures of honor have arisen in many parts of the world. Their central feature is that males protect a personal reputation for strength by resorting to violence.\footnote{Sripada and Stich, “Evolution, Culture, and the Irrationality of the Emotions,” 147.} Cultures of honor tend to appear in environments where a person’s resources are vulnerable to being stolen, and no government apparatus can be relied upon to deter theft. Psychologists Richard Nisbett and Dov Cohen suggest that a culture of honor continues to exist in the American South.\footnote{Richard E. Nisbett and Dov Cohen, \textit{Culture of Honor: The Psychology of Violence in the South}, New Directions in Social Psychology (Boulder, CO: Westview Press, 1996).} They base this contention on experiments in which they found young Southern men to be substantially more angered by verbal insults than young Northern men.\footnote{Sripada and Stich, “Evolution, Culture, and the Irrationality of the Emotions,” 147 – 149.} In the Southern males, verbal insult produced increased levels of cortisol and testosterone—increases which were significantly larger than increases exhibited by Northern males. Cortisol is a hormone associated with stress and anxiety, while testosterone is a hormone associated with aggression and dominance behavior. Nisbett and Cohen also conducted surveys revealing that Southern respondents were more likely than Northerners to say that a violent response would be justified in situations where a man’s honor is challenged.\footnote{Richerson and Boyd, \textit{Not By Genes Alone}, 2.} At the macro-social level, the homicide rate in the American South was ten times the rate for the whole United States during the period 1865 – 1915, and a similar asymmetry exists today.\footnote{Ibid., 1 – 2.}

Nisbett and Cohen explain these differences as symptoms of a culture of honor that continues to hold sway in the South. Interestingly, however, Nisbett and Cohen do not suggest that the culture of honor exists because it currently performs a beneficial function. Instead, the
Southern culture of honor was likely brought to the region centuries ago by predominantly Scottish-Irish immigrants, who had subsisted as herders for generations. The high value and portability of herd animals made them easily susceptible to theft. In addition, the Southern region was, for some time, sparsely populated with a minimal governmental presence. These conditions made it beneficial for herders (who were typically men) to deter theft by cultivating a reputation for reacting violently to personal affronts.

If Nisbett and Cohen’s account is correct, it would seem that the Southern culture of honor once served the characteristic proper function of moral systems. Settlers cooperated to establish a code of conduct according to which men were expected and encouraged to protect their reputations with violence. General compliance with the code yielded the benefit of deterring theft in an environment where no other institution could furnish the same benefit. So, during this period, the Southern culture of honor qualified as a moral system.

However, the benefits associated with the culture of honor disappeared as the South came to be reliably policed by the state, and as the vast majority of Southerners have taken on non-agricultural occupations. Today, the persistence of the culture of honor is no longer explained by its effect in bringing about beneficial cooperation. Instead, the culture of honor’s continued presence is more likely a product of conformist social learning. Cohen and colleagues found that Southern men today generally do not think that not responding to an insult is a reason to view someone as less manly, but they still believe that others will view him this way. As a result, men continue to follow honor norms because they think others would have a low opinion of them if they didn’t, not because those norms facilitate mutually beneficial
Thus, the culture of honor no longer performs the proper function of a moral system—that of enabling beneficial social cooperation. And for this reason, I take the Southern culture of honor to be an example of a value system that once was a moral system, but no longer is.

8 Conclusion

Researchers in search of a descriptive definition of morality should pursue a real definition, rather than a nominal definition. A real definition of morality identifies the properties that constitute moral systems as a natural kind—or more precisely, a *copied kind*. Such a definition should cite the properties that explain the emergence and persistence of moral systems over time. One essential property is the function which moral systems perform in enabling mutually beneficial social cooperation. Under the definition proposed here, morality should not be understood as pertaining solely to concerns with harm, justice, and rights. This is because moral systems also perform their function by addressing concerns related to loyalty, authority, and purity. On this score, I am in agreement Jonathan Haidt’s criticism of Turiel. However, contrary to what Haidt suggests, some value systems that demand loyalty, deference to authority, and purity are not in fact moral systems. For some of these systems do not exist in virtue of enabling beneficial social cooperation.

---