To appear in *Proceedings & Addresses of the American Philosophical Association* (2024)

The Cognitive Science of Moral Judgment and the Tower of Babel Problem

Stephen Stich
Rutgers University

1. What's The Tower of Babel Problem?

Though "The Tower of Babel Problem" is not a widely used label in philosophy[1], the problem I use the label for is widely recognized. It arises when a single term is used (within one discipline or in neighboring disciplines) to denote importantly different phenomena. I became vividly aware of the problem several decades ago when I began working on altruism.[2] During the last quarter of the 20th century, it was common for philosophers and biologists to use the term 'altruism' in dramatically different ways, though there was often little awareness of the ambiguity. The problem was significantly reduced, though certainly not eliminated, with the publication of Sober and Wilson's *Unto Others* (1998), where they analyzed and emphasized the importance of the distinction between psychological and evolutionary interpretations of 'altruism'.

Moral philosophers, from Hobbes[3] onward, have most often been concerned with psychological altruism, though Bedhwar (1993), Schramme (2017) and others remind us that some moral philosophers have had something very different in mind. For these philosophers "altruism is … basically identical with taking the moral point of view, i.e. an individual appreciation of the normative force of morality." (Schramme, 2017, 203-4) Psychologists too have primarily been concerned with psychological altruism, and some heated debates have been fueled by different accounts of psychological altruism.[4] In biology, the focus is usually on evolutionary altruism, though biologists and philosophers of biology have distinguished many different kinds of evolutionary altruism.[5]

In the social sciences, however, 'altruism' is still used in a bewildering variety of ways. Clavien & Chapuisat (2013) review the literature in experimental economics, evolutionary anthropology and evolutionary game theory, and note, with more than a hint of understatement, that "the nature of the altruism implicated in these studies is not fully clear" (130). Ramsey (2016) argues that a number of eminent primatologists, including Frans de Waal and Michael Tomasello, invoke still other accounts of altruism. To make their case, Clavien & Chapuisat and Ramsey don't simply collect explicit definitions of "altruism" since many researchers do not provide them, and those that are provided are often hard to interpret. Rather they focus on the tests or procedures that researchers use to determine whether the humans or primates they are studying behave altruistically. In light of this bristling diversity, Clavien & Chapuisat lament

[1] I wish I could say that I am the first to use it. But I'm not. See Iliadis (2019).
[2] See, for example, Stich, Doris & Roedder (2010).
[3] Hobbes never used the *term* "altruism," which was a neologism initially introduced by August Comte in the 19th century. For a fascinating history of the term, see Dixon (2008).
[4] Grant (1997), Batson (2011), 20-30.
[5] See, for example, Kerr, Godfrey-Smith and Feldman (2004).

that "the notion of altruism has become so plastic that it is often hard to understand what is really meant by authors using the term, and even harder to evaluate the degree to which results from one research field – e.g., experimental economics – may facilitate the resolution of debates in another research field – e.g., evolutionary biology or philosophy." (2013, 134) Though the existence of a Tower of Babel Problem in the large literature on altruism is well documented and widely acknowledged, it's my contention that there is an equally serious, and largely unrecognized, Tower of Babel Problem in the even larger literature on the cognitive science of moral judgment.[6]

2. Turiel vs. Haidt: An Example of The Tower of Babel Problem in the Cognitive Science of Moral Judgment

The literature on the cognitive science of moral judgment is enormous and the full story about the Tower of Babel Problem that besets this literature is long and complicated. But since space is limited, I'm going to start by focusing on the often-ignored details of one of the most important and widely known disputes in this area. Then, much more quickly, in §3, I'll review a wide variety of other studies to make a *prima facie* case that the Tower of Babel Problem is widespread in the cognitive of moral judgment. Though there are many researchers involved in the dispute I'll start with, the two central figures are Elliot Turiel and Jonathan Haidt.

2.1. Turiel's Achievement

Turiel was a student of Lawrence Kohlberg (1927-1987), and Kohlberg was influenced by the pioneering work of Jean Piaget (1896-1980). Both Piaget & Kohlberg were advocates of what Turiel has called "the differentiation model" which maintains that "moral reasoning emerges through its differentiation from nonmoral processes". In young children "convention & morality are presumed to be undifferentiated, while in older children the two are differentiated."[7] Turiel was skeptical of the claim that there is only one sort of normative cognition in young children. He was convinced that moral cognition is distinct from cognition about social conventions and that the distinction is present quite early in development.

In order to defend this claim, Turiel needed an empirical test that would indicate whether normative judgments made by experimental participants (including very young participants) were moral judgments or judgments about a conventional matter. In constructing his test, Turiel drew inspiration from the large philosophical literature on the definition of morality that had emerged beginning in the early 1950s. That literature has largely disappeared from the philosophical curriculum. But from 1950 thru 1980 it generated hundreds of papers including some by philosophers who were widely recognized as leading figures in the field.[8] The philosophers who contributed to this large literature disagreed on just about everything, and after

[6] In January, 2024, a Google Scholar search found 376,000 publications on "cognitive science of altruism" and 1,140,000 publications on "cognitive science of moral judgment".
[7] Turiel (1977), p. 78.
[8] Wallace and Walker (1970) is a valuable collection of papers in this area. Among the authors included are Elizabeth Anscombe, Philippa Foot, William Frankena, Alasdair MacIntyre, Peter Strawson and Charles Taylor.

a third of a century of vigorous philosophical debate, they reached no consensus on any of the issues that divided them.[9]

Turiel's test incorporated several features, each of which had been proposed by a number of philosophers in this literature as a necessary feature of moral judgments, though those proposals had been disputed by other philosophers. The first feature was universalizability. "Moral prescriptions" Turiel tells us, "are *universally applicable* in that they apply to everyone in similar circumstances" (1983, 36). So if a young experimental participant judges that it is wrong for a child in her school to push someone off a swing, and if that judgment is a moral judgment we should expect the participant to judge that it is wrong for a child in a different school in a different location to push someone off a swing under similar circumstances. We should also expect the participant to judge that the same action would be wrong if it happened sometime in the past or if it were to happen sometime in the future. A second feature that was discussed in the philosophical literature and adopted by Turiel is often referred to as "authority independence." The basic idea is that if an action is morally wrong, then it would still be wrong even if there were no explicit rule against it, and it would still be wrong even if some recognized authority said that the action was not prohibited. Since Turiel thought that young children could distinguish transgressions of moral rules from transgressions of conventional rules, he also needed an empirical test that would indicate that a participant considered an action to be a transgression of a conventional rule. Here again, he consulted the philosophical literature. Relying principally on his interpretation of the work of his Berkeley colleague, John Searle, he claimed that if a participant took an action to be a transgression of a conventional rule she will not insist that the same action would be wrong at other times and other places (so it is not "universalizable") and she will think that the rule might be altered by a suitably authoritative person or group (so it is not "authority independent").

Using these putative features of moral and conventional judgments, Turiel constructed an empirical test to determine whether an experimental participant's judgment about a transgression is a moral judgment or a conventional judgment. The test has 4 steps:

1. It begins with a brief, age-appropriate, vignette describing a hypothetical transgression.

2. Then the participant is asked whether she thinks the transgression was wrong. The question does not use terms like 'moral'. Rather the participant may be asked whether the action is "OK," or some similar locution may be used.

3. In the third step the participant is asked age-appropriate questions about whether the transgression generalizes in space and time. Questions might include: "Would it be OK in [a distant town or a foreign country known to the participant]?" "Would it be OK when your grandparents were kids?" "Would it be OK 100 years in the future?"

4. In the fourth step, the participant is asked one or more age-appropriate questions about whether the wrongness of the action is authority independent: "Would it be OK if your

[9] For a more detailed and nuanced discussion of this literature, see Stich (2019).

> teacher [or the principal of your school] said there was no rule against it?" "Would it be OK if the priest in your church said it was OK?")

This experimental procedure is going to play an important role in the pages to follow, so we would do well to give it a distinctive label. Turiel and his collaborators refer to participants' responses to the questions asked as "criterion judgments". So I propose to label the procedure the "Turiel Criterion Judgment Test". This would also be a good place to introduce the some of Turiel's collaborators. Though he has many co-authors, three of the most important are Melanie Killen, Larry Nucci and Judith Smetana. For reasons that needn't concern us, Turiel and his followers are often called "social domain theorists".

Starting in the mid-1970s Turiel and colleagues used versions of the Turiel Criterion Judgment Test with a variety of participant groups. This work used transgressions that Turiel and colleagues took to be obviously moral – like one child pushing another off a swing because he wanted to use the swing – and transgressions that they took to be obviously conventional – like a child wearing pajamas to school. The results clearly confirmed Turiel's prediction. Transgressions that Turiel and colleagues took to be obviously moral were typically judged to be wrong, authority independent and generalizable to other places and times. Transgressions that Turiel and colleagues judged to be obviously conventional were typically judged to be wrong, authority dependent and not generalized to other times or places. Moreover, these distinctions emerged very early in development. By their 4th birthday, and often earlier, kids had systematically different reactions to the sorts of transgressions used. This was an overwhelming victory for Turiel in his disagreement with Kohlberg and Piaget. It is not the case that young children think of all normative rules in the same way; their normative cognition is not "undifferentiated". Normative cognition in young children is much more subtle, more varied and more complex than Kohlberg and Piaget had portrayed it. These findings were a major achievement in the study of child development.

2.2. Turiel's Definition of Morality and Haidt's Critique

I've lured you to read this far by promising to discuss Haidt's challenge to Turiel. But bear with me a little further. A bit more background is needed. In the work of Turiel and his collaborators there are frequent discussions of "the definition of morality." However, these definitions are very puzzling. It is often not clear what the definitions are, what their status is, or what role they play in the research of Turiel and his collaborators. At one point Haidt claims that the definitions are "stipulative" (Haidt & Joseph, 2007, 371). But this is clearly mistaken since Shweder, Turiel and Much (1981, 289) insist that "the meaning of 'morality' is something discovered, not stipulated". What's going on here? Why do Turiel and colleagues maintain that the meaning of 'morality' is something to be discovered? One obvious suggestion is that they sought to discover the meaning of the ordinary English terms "morality," "moral" etc. But that, too, is clearly mistaken. "The terminology", Turiel tells us,

> does not necessarily correspond to general, nonsocial scientific usage of the labels "convention" and "morality". Clear or systematic patterns of general use of the labels would be difficult to discern. As labels, the terms are often used interchangeably or even

> inconsistently; sometimes they correspond to the definitions provided here and sometimes they are inconsistent with them. (Turiel 1983: 34)

Machery and I (2022) have argued that the best interpretation of Turiel's definitions is that after having constructed the Turiel Criterion Judgment Test and shown that it reliably identifies many judgments that are intuitively moral, the proposed "definitions of morality" are attempts to characterize the class of judgments that the Criterion Judgment Test classifies as moral. As we might expect on this account, the proposed definitions of morality evolved over the years as additional experiments were conducted. The earliest definition identified moral transgressions as transgressions the violated principles of justice (Turiel, 1977, 80; Nucci & Turiel, 1978, 400-401). A later, widely quoted, account maintained that "the moral domain refers to prescriptive judgments of justice, rights and welfare pertaining to how people ought to relate to each other" (Turiel, 1983, 3). And the surrounding text makes it clear that a central element in Turiel's conception of welfare is the avoidance of harm. In more recent publications, judgments involving fairness and equality have been added to the definition. (Killen & Smetana, 2015; Killen & Dahl, 2018)

2.3. Shweder's Challenge

These definitions lie at the center of the Haidt's critique of Turiel's work. The problem with the definitions, Haidt maintains, is that "they do not travel well" (2012, 22). Moreover, on Haidt's view, the problem is not restricted to Turiel and his collaborators.

> [T]he psychological study of morality… has been dominated by politically liberal researchers …. The lack of moral and political diversity among researchers has led to an inappropriate narrowing of the moral domain to issues of harm/care and fairness/ reciprocity/justice…. Morality in most cultures (and for social conservatives in Western cultures), is in fact much broader, including issues of ingroup / loyalty, authority / respect, and purity / sanctity …." (Haidt & Joseph, 2007, p. 367)

To understand what's going on here, we need to go back a few decades and look at the work of Haidt's post doc supervisor, Richard Shweder. Shweder was an early critic of Turiel's definition of morality. He claimed that in many non-Western societies many transgressions not involving harm (or justice or rights) are classified as moral, and that the class of transgressions viewed as conventional is vanishingly small. To support this claim Shweder and colleagues (1987) did an elaborate study comparing the judgments of participants in Chicago with those of orthodox Hindu participants in Bhubaneswar, India. Using his own test for distinguishing moral from conventional judgments, Shweder found a number of behaviors that the Indians took to be moral transgressions and the Americans did not. Two widely discussed examples were

> The day after his father's death, the eldest son had a haircut and ate chicken.

> A widow in your community eats fish two or three times a week.

Not surprisingly, using Shweder's test, the Americans judged these to be not morally wrong at all; but the Indians judged them to be quite serious moral transgressions. Unfortunately for

Shweder, Turiel and colleagues (1987) pointed out that the Indian participants had important factual beliefs that the Americans did not share. They believed that the son's behavior would prevent the father's soul from receiving salvation, and that the widow's fish eating might lead her to have sex, which would lead her husband's spirit to suffer. So the Indians believed that the son's and widow's behavior was harmful, and the Americans did not.

2.4. Haidt to the Rescue

In the early 1990's Haidt designed a now famous study aimed at providing a better test of Shweder's critique of Turiel. To do this he looked at people's judgments about what he described as "harmless taboo violations". "The basic research strategy", Haidt tells us, "is to present subjects with stories that are affectively loaded – disrespectful or disgusting actions that 'feel" wrong – yet that are harmless" (Haidt, Koller & Dias, 1993, 215). To be sure that the participants regarded the transgressions as harmless he *asked* them whether anyone was harmed.[10] Some of the scenarios Haidt used in this study have become very well known. But they are too much fun to pass over without reminding you of a few:

> *Dog*: A family's dog was killed by a car in front of their house. They had heard that dog meat was delicious, so they cut up the dog's body and cooked it and ate it for dinner.
>
> *Chicken*: A man goes to the super-market once a week and buys a dead chicken. But before cooking the chicken, he has sexual intercourse with it. (Haidt, Koller & Dias, 1993, 617)

Famously, using a modified and truncated version of Shweder's test, Haidt and colleagues found that Brazilians in the poor, less westernized city of Recife and low SES Americans tend to regard transgressions like these as moral transgressions even though they explicitly state that no one is harmed. It was this work that Haidt used to support his claim that Shweder was right about Turiel's definition of morality. *It does not travel well!*

2.5. Haidt and Shweder vs. Turiel: The Tower of Babel Problem

Though I have tucked it under rug to add a bit of drama to my story, there is a largely unnoticed Tower of Babel Problem lurking here. To determine whether a judgment is a *moral* judgment, Turiel and his colleagues always use a version of Turiel's Criterion Judgment Test. This, recall, begins with an age-appropriate vignette describing a hypothetical transgression. It then poses three sets of questions. The first set is aimed at determining whether the participant thinks the behavior described is wrong. The second set is aimed at determining whether a participant thinks the transgression generalizes in time and in space. The third set is aimed at determining whether the participant thinks the wrongness of the transgression is authority independent. Shweder and colleagues also used a questionnaire whose questions, they maintain, "can be viewed as criteria for distinguishing moral or objective obligations from conventional or consensus obligations" (Shweder, Malapatra & Miller, 1987, 42). But Shweder's questionnaire is dramatically different from Turiel's. After describing the behavior in question it asks a total of eight questions:

[10] For a lively account of this research project and the events that led up to it, see Haidt (2012), 14-26.

1. Is the behavior under consideration wrong?
2. How serious is the violation?
   (a) not a violation; (b) a minor offense; (c) a somewhat serious offense; (d) a very serious offense
3. Is it a sin?
4. What if no one knew that this had been done. It was done in private or secretly. Would it be wrong then?
5. Would it be best if everyone in the world followed (*the rule endorsed by the informant*)?
6. In (*name of a relevant society*) people do (*the opposite of the practice endorsed by the informant*) all the time. Would (*name of relevant society*) be a better place if they stopped doing that?
7. What if most people in (*name of informant's society*) wanted to (*change the practice*). Would it be OK to change it?
8. Do you think a person who does (*the practice under consideration*) should be stopped or punished in some way?
   (Shweder, Malapatra & Miller, 1987, 42)

Shweder's questionnaire includes *six* features that are not included on Turiel's Criterion Judgement Test:

- the seriousness of the transgression
- whether the transgression is a sin
- whether the transgression would still be wrong if it were done in private
- whether the transgression would still be wrong if it were done secretly
- whether the perpetrator should be stopped
- whether the perpetrator should be punished

Though Turiel and his collaborators rarely mention the relation between being a sin and the moral or conventional status of a transgression, there is a brief passage in Nucci & Turiel (1993, 1476) where they claim that some behaviors that the Catholic Church considers to be sins are moral transgressions while others are conventional transgressions. To the best of my knowledge, Turiel has have never suggested any link between performing action in private or performing it secretly and the moral or conventional status of the action. Nor has he addressed whether being a moral or conventional transgression is relevant to whether the perpetrator should be stopped. But Judith Smetana, one of Turiel's frequent collaborators, has made it clear that social domain theorists do not include judgments about whether or not a perpetrator should be punished as a "criterion judgment" because they do not think it is relevant to the moral or conventional status of a transgression.[11] Smetana also insists that the seriousness of a transgression is not taken to be relevant in deciding whether it is moral or conventional (Smetana, 1993, 117). And in their reply to Shweder, Turiel and colleagues report studies in which participants judge some

[11] [J]udgments of … the degree of censure that the transgressor deserves … are less informative than criterion judgments because all rule violations are, by definition, unacceptable, wrong, and punishable. (Smetana et al., 2018, 270)

transgressions that Turiel takes to be obviously conventional, like wearing pajamas to school, to be *more* serious than some transgressions, like stealing a pencil, that Turiel takes to be obviously moral (Turiel, Killen & Helwig, 1987, 175). So Shweder's questionnaire includes *lots* of features that Turiel and colleagues take to be irrelevant to the moral or conventional status of a transgression. Also, Shweder's questionnaire does *not* include any question about generalization in time, nor does it include a Turiel-style authority independence question, though Turiel considers both of these to be crucial in determining whether a participant takes a transgression to be moral or conventional. In light of these many differences, it is overwhelmingly likely that if Shweder's questionnaire and Turiel's Criterion Judgement Test were both used with a wide range of transgressions in the same population, they would diverge substantially in the transgressions that they classify as moral.

I noted earlier that Haidt used a modified and truncated version of Shweder's test. Here are the questions that Haidt asks after describing the behavior in question:

1. What do you think about this? Is it very wrong, a little wrong, or is it perfectly OK for [act specified]?
2. Can you tell me why?
3. Imagine that you actually saw someone [performing the act]. Would it bother you or would you not care?
4. Should [the actor] be stopped or punished in any way?
5. Suppose you learned about two different foreign countries. In country A, people [do that act] very often, and in country B, they never [do that act]. Are both of these customs OK, or is one of them bad or wrong?

Though *none* of these questions are exactly the same as questions on Shweder's questionnaire, Haidt's question 4 and Shweder's question 8 are almost identical, and Haidt's question 1 is similar to Shweder's question 2. Haidt's question 5 is clearly inspired by Shweder's question 6, though the judgments requested are notably different. Haidt's questions 2 and 3 have no analog on Shweder's questionnaire, and Shweder's questions 3, 4 and 7 have no analog on Haidt's questionnaire. Turiel and colleagues would classify questions like Haidt's 2 to be a "justification category" question which they consider to be importantly different from criterion judgment questions and are not used when they classify a judgment as moral or conventional (Turiel, 1983: 52–53, 66–68). Though authority independence and generalization in time and space are crucial components of Turiel's Criterion Judgment Test, Haidt asks nothing about authority independence and does not explicitly ask about generalization in time. So it is all but certain that if Haidt's questionnaire, Shweder's questionnaire and Turiel's Criterion Judgment Test were all used with a wide range of transgressions in the same population, the transgressions classified as moral by Haidt's questionnaire would differ from those Turiel's Task classified as moral *and* from those that Shweder's questionnaire classified as moral.

The upshot of all of this is that Haidt's critique of Turiel's definition of morality is undermined by a Tower of Babel Problem. Turiel's definition was an attempt to specify the sorts of transgressions that his Criterion Judgment Test would classify as moral. And even if Haidt's famous "…Is it Wrong to Eat Your Dog?" study is otherwise completely unproblematic, he has not shown that Brazilians in Recife and low SES participants in Philadelphia judge that

"harmless taboo violations" are morally wrong, *using Turiel's Criterion Judgment Test to identify moral judgments*. Nor has he used a questionnaire that is likely to agree with Turiel's Test about which judgments are moral judgments. So Haidt and Turiel are arguing at crossed purposes, as are Shweder and Turiel. They are using dramatically different tests to determine which judgements they will classify as *moral* judgments.[12]

3. The Rest of the Iceberg: A Brief Survey of the Tower of Babel Problems in the Cognitive Science of Moral Judgment

Though Haidt's critique of Turiel provides a clear and quite striking example of a Tower of Babel Problem, the cognitive science of moral judgment is awash in additional examples. During the last three decades, researchers have used dozens of strategies to determine which judgments made by experimental participants the researchers would classify as moral judgments. In this section I will offer a far from comprehensive survey of classification strategies. It divides them into three categories:

i. **"Turiel Style" Strategies** that borrow ideas, or claim inspiration, from Turiel in specifying what the researcher takes to be features of moral judgments

ii. **Ordinary Language Strategies** that ask participants whether a judgment was a moral judgment, using the English word 'moral' or related terms

iii. **Intuitive Strategies** that rely on the researcher's intuition to determine whether a participant's judgment is a moral judgment

3.1. "Turiel Style" Strategies

Starting around the turn of the century, the term "the moral/conventional task" became increasingly common in the experimental literature on moral judgment, though Turiel and other social domain theorists never use the locution. Researchers who do use the term typically acknowledge the importance of Turiel's work. But most versions of the task depart in important ways from Turiel's Criterion Judgement Test. Here are a few examples.

Shaun Nichols was among the first to use the term "the moral/conventional task" in discussing his experiment work. The version of the task used in Nichols (2002) was closely modeled on the procedure used by the psychologist James Blair (1995). Blair called it "the moral/conventional distinction task". Nichols' task asked four questions:

1. Was [the behavior described] OK?
2. If [the behavior] is not OK, then, on a scale of 1 to 10, how bad was [the behavior]?

[12] Though it is not relevant to the claims made in this paper, it is worth noting that 20+ years after Haidt's paper appeared, two studies (Berniūnas et al. 2016 & 2020) *did* use Turiel's Criterion Judgment Test and Haidt-style harmless taboo violations in non-Western societies, and they found that Haidt was right. A number of harmless taboo violations do evoke *moral* rather than *conventional* criterion judgments.

3) Why was the behavior bad?
4) [A question designed to determine whether the wrongness of the behavior is authority dependent]. (Nichols, 2002, 229)

The second item on this list is a seriousness question, and as noted earlier, Turiel and other social domain theorists explicitly reject including a seriousness question in their Criterion Judgment Test. The third question on Nichols' list is what Turiel and his colleagues call a "justification category" question, and these too are excluded in Turiel's Criterion Judgment Test. However, that Test always includes questions aimed at determining whether the participant views the wrongness of the transgression to be generalizable in time and space. Nichols' task does not ask about the generalizability of the transgression. So Nichols' moral/conventional task and Turiel's Criterion Judgment Test pose notably different questions.

Huebner et al. (2010) use a quite different 4-question task. Here is their account of the questions they ask after recounting a transgression:

1. BADNESS: [Name]'s behavior was: (1, very bad; 4, neither good nor bad; 7, very good)
2. PUNISHABILITY How much should [name] be punished: (1, severely punished; 7, not punished at all)
3. UNIVERSALITY: If [name] lived somewhere where everyone else did this, would it be wrong for [name] to do this (Yes; No)
4. AUTHORITY: If the government passed a law that said it was ok to do what [name] did, would that make [name's] action OK? (Yes; No).

(Huebner et al. 2010, 5)

Question 1 is a seriousness question, which plays no role in Turiel's Criterion Judgment Test. Question 2 is a question about punishment, and as noted in our discussion of Haidt, social domain theorists reject the inclusion of a punishment question. The third question asks about universalizability, but it does so in a way quite different from the universalizability questions Turiel uses. Turiel asks whether the behavior would be wrong in different places and different times, but Huebner and colleagues ask whether it would be wrong in a place where everyone did it. So Huebner et al.'s moral/conventional task questions differ substantially from both Nichols' task and Turiel's.

Edward Royzman and colleagues have been frequent defenders of Turiel's general approach to the study of moral cognition,[13] and they have noted with approval that the moral/conventional task is "one of the most widely used measures of mature moral judgment" (Royzman, Landy & Goodwin, 2014: 178). However, in a number of studies Royzman and colleagues use versions of the moral/conventional task that depart quite dramatically from Turiel's Criterion Judgment Test. Perhaps the most radical departure is the single question "Moral-Conventional Distinction Task (MCDT)" used by Royzman, Landy and Goodwin (2014).

[13] See, for example, Royzman, Leeman & Baron (2009), 172-172, and Royzman, Goodwin & Leeman (2011), 102.

> The key MCDT probe[14] asked participants to "suppose that there were foreign country A where some time ago everyone came together and decided that a behavior such as this was OK. In your view, would it be wrong or not wrong for [the protagonist's name] to do what he/she did, assuming he/she was raised and lived in country A?" In keeping with the traditional MCDT format, this question was framed as a dichotomous "wrong"/"not wrong" choice, with "wrong" and "not wrong" responses being categorized as "moralizing" and "non-moralizing" responses, respectively." (Royzman, Landy & Goodwin, 2014, 181)

It would be easy to add additional examples of experimental studies that identify moral judgments using procedures inspired by Turiel's Criterion Judgment Test, but differing from it in a variety of significant ways. And it would be an all but impossible task to assemble a complete list of such studies, since new ones continue to appear. But perhaps I have already said enough to support the claim that there is likely to be a massive Tower of Babel Problem in this literature. Researchers who identify moral judgments using a moral/conventional task inspired by Turiel's work are often using substantially different procedures to identify moral judgments. These researchers offer no reason to think that *their* procedure and Turiel's Test will classify judgments in the same way, and that's not surprising, since it is very implausible that they would.

3.2. Ordinary Language Strategies

Turiel warned that in everyday discourse the term "morality" was used in a variety of ways. And it is very likely that much the same is true of the terms "moral" [15] and "immoral", and of expressions like "morally wrong", "morally acceptable", "morally permissible," "morally good," "morally bad," "moral belief", "moral conviction", and "moral dilemma". However, these terms are frequently used to elicit what researchers take to be participants' moral judgments. Here are a few examples:

- Young and Saxe (2011, p. 206) ask participants "How morally wrong was the action? (1 = not at all, 7 = very much)."
- In Huebner, Hauser and Pettit (2011. p. 214), "[a]fter reading the text of a dilemma, each participant was asked to judge whether the protagonist's action was morally permissible; and, participants responded with either a 'Yes' or a 'No'."
- Schwitzgebel and Rust (2014, p. 298) "asked the respondent to rate 'the degree to which the action described is morally good or morally bad' by checking one circle on a nine-point scale from 'very morally bad' (which we coded as 1) to 'very morally good' (coded as 9), with the midpoint labeled 'morally neutral' and the 3 and 7 points labeled 'somewhat morally bad' and 'somewhat morally good' respectively."
- In Bernhard et al. (2016, p. 1875), "Screen 3 prompts participants for a yes/no response to the question, 'Is it morally acceptable for you to [perform action described in Screen 2]?'"

[14] It wasn't just the key MCDT probe, it was the *only* MCDT probe!
[15] The online Oxford English Dictionary (OED.com, 2023) offers 17 meanings for the adjective "moral".

- In Gray et al. (2022, p. 12), "[p]articipants rated the immorality (our operationalization of moral judgment) … of each scenario with the following question …— all using 6-point Likert scales from 1 to 6: How immoral is this act? 1 (*Not immoral*) to 6 (*Extremely immoral*)"

None of these studies explore how participants understand these probes, nor do they offer any evidence that most participants in a study understand the probe used in the same way. And it is far from clear that all 5 probes would yield the same results. More important, for our purposes, none of these studies offers any reason to think that Turiel's Criterion Judgment Test (or other "Turiel Style" tests) would agree with the probe used in classifying judgments as moral judgments. Of course, whether a specific probe would agree with Turiel's Test over a substantial range of cases is an empirical question. But, to the best of my knowledge, it has never been explored. If, as I suspect, Turiel's Test and "ordinary language" probes like those assembled above would *not* classify the same judgments as moral, the use of these probes generates a quite massive Tower of Babel Problem in the moral psychology literature. Moreover, as noted earlier, typically no attempt is made to determine how participants *in a given study* interpret a putative moral judgment question or whether all participants in that study interpret a question in the same way. If they don't, then there is a Tower of Babel Problem lurking *within* many moral psychology studies.

During the last two decades, some of the most interesting and influential work in moral psychology has been done by Linda Skitka and her colleagues. Skitka has shown that a wide range of phenomena including intolerance of disagreement, difficulty in resolving conflicts, political engagement, reluctance to sit near someone, and willingness to accept lying, cheating and violence to achieve one's goals are correlated with the extent to which participants classify their normative belief as a "core moral conviction" (Skitka et al. 2005, 2009, 2011, 2017, 2021).

To determine whether participants take their view on an issue to be a core moral conviction, Skitka asks them questions like:

> "How much are your feelings about [*the issue*] connected to your core moral beliefs or convictions?"

and

> "How much are your feelings about [*the issue*] based on fundamental beliefs about right and wrong?"

Skitka describes these questions as "transparent and face-valid self-report measures" and maintains that while "people may not always be skilled at explaining why they believe a given attitude is moral, they have little problem recognizing whether and to what degree a given attitude reflects a moral conviction" (Skitka et al. 2021, 351). Though Skitka suggests that the work of Turiel and other social domain theorists supports her strategy for identifying moral judgments (2021, 349), Skitka's questions are, obviously, very different from the questions posed in the Turiel Criterion Judgment Test, and from all the other Turiel Style procedures we've considered for identifying moral judgments. It's far from obvious that participants would reply

that their feelings about one child pushing another off a swing were "very much connected to their core moral beliefs or convictions". But the sorts of issues that are the focus of Skitka's work – issues like abortion, gay marriage, and capital punishment – are far removed from the sorts of examples typically used by social domain theorists, and to the best of my knowledge no one has ever used Skitka's questions with Turiel's schoolyard examples.[16]

Though Skitka was confident that participants would find her questions unproblematic and easy to interpret, some researchers influenced by Skitka's work were not convinced that participants would understand her questions, or that all participants would interpret them in the same way. So they instructed participants on how the term "moral" should be construed. A notable example is Jennifer Cole Wright. Wright and colleagues instructed participants "to identify an issue as *moral* if they believed the issue's rightness or wrongness to be non-negotiable and objectively grounded and *nonmoral* if they believed the issue's rightness or wrongness to be dependent on an individual or social decision…. Potential examples of each category were given (nonmoral, listening to classical music or driving on the right side of the road; moral, torturing innocent children for pleasure)" (Wright et al., 2008, 1464). Wright and colleagues offered participants no explanation of the philosophically laden terms "non-negotiable" and "objectively grounded," nor did they attempt to determine how participants interpreted these terms. But when they asked participants to judge whether 40 examples were moral or nonmoral issues, the results were rather startling:

> 10% said rape was a nonmoral issue
> 23% said "children with handicaps being put to death" was a nonmoral issue
> 64% said euthanasia was a nonmoral issue
> 72% said the death penalty was a nonmoral issue
> (Wright et al., 2008, 1465)

Clearly, the procedure that Wright and colleagues used to identify which issues a participant took to be moral is notably different from Turiel's procedure for identifying moral judgments (and from Haidt's and Shweder's) and from the "Turiel Style" strategies surveyed in §3.1. It's hard to believe that Wright's procedure and Turiel's Criterion Judgment Test would agree about whether a participant considered a wide range of judgments to be moral. But we can't be sure since no one has ever tried.

Before moving on, a pair of comments comparing Ordinary Language Strategies with Turiel's Criterion Judgment Test are in order. First, unlike Turiel's Criterion Judgment Test, the Ordinary Language Strategies we considered can't be used on young children since they have not yet acquired the word "moral" or expressions like "morally wrong," and "morally permissible" nor do they understand expressions like "fundamental beliefs" or "objectively grounded". Second, though the Turiel's Criterion Judgement Test is relatively easy to translate, there may be many languages in which the Ordinary Language Strategies we've reviewed can't be used at all, because those languages do not have good translations for the English word "moral" and similar terms. Citing the work of Buchtel et al. (2015), Schein, Ritter & Gray (2016, 870) worry that there may be no good Chinese translation for the word "morality," and Sachdeva, et al. (2011, 174) claim that there is no word for morality in Hindi.

[16] I'm grateful to Professor Skitka for timely and helpful responses to my inquiries about her work.

3.3. Intuitive Strategies

Let's turn now to Intuitive Strategies that rely on *researchers' intuition* to determine whether a participant's judgment is a moral judgment. Though there are many studies that rely on the researchers' intuition, during the last decade and a half an enormous literature has emerged that relies on the intuition of a small group of researchers. The two best known members of the group are Jonathan Haidt and Jesse Graham, and the research program that they launched is Moral Foundations Theory.[17] These researchers have made their primary research tool, the Moral Foundations Questionnaire, freely available online and other researchers are welcome to use it.[18] It is now available in 39 languages! The group has also published a series of papers detailing how the Questionnaire was developed and validated. Two of the most important of these are Graham, Haidt & Nosek (2009) and Graham et al. (2011). Here are a pair of quotes explaining how the widely used second version of the questionnaire was constructed:

> For the second version [of the Moral Foundations Questionnaire] we added a new section that assessed levels of agreement with … moral judgment statements…. [W]e wanted to supplement the abstract relevance assessments [used in the first version] … with contextualized items that could more directly gauge *actual moral judgments*. [Graham et al., 2011, 369, 371, emphasis added]

> In Study 2, we retained the abstract moral relevance assessments from Study 1 and added more contextualized and concrete items that could more strongly trigger the sorts of moral intuitions that are said play an important role in moral judgment…. This approach *requires participants to make moral judgments* about cases that instantiate or violate the abstract principles they rated in response to our "relevance" questions." ... Moral judgment statements were rated on a 6-point scale, from *strongly disagree* to *strongly agree*. [Graham, Haidt & Nosek (2009), 1033-4, first emphasis added]

What's important for our purposes is that the authors were designing the revised questionnaire to insure that in expressing their agreement or disagreement with the statements in the new section, participants would be making "actual moral judgments." And what are the statements? Here is the complete list.

[17] A Google Scholar search for "Moral Foundations Theory" in December, 2023, yielded an astounding 4.3 million citations!

[18] https://moralfoundations.org/questionnaires/

Appendix B

Moral Judgment Items, Study 2

Harm:
If I saw a mother slapping her child, I would be outraged.
It can never be right to kill a human being.
Compassion for those who are suffering is the most crucial virtue.
The government must first and foremost protect all people from harm.

Fairness:
If a friend wanted to cut in with me on a long line, I would feel uncomfortable because it wouldn't be fair to those behind me.
In the fight against terrorism, some people's rights will have to be violated [reverse scored].
Justice, fairness and equality are the most important requirements for a society.
When the government makes laws, the number one principle should be ensuring that everyone is treated fairly.

Ingroup:
If I knew that my brother had committed a murder, and the police were looking for him, I would turn him in [reverse scored].
When it comes to close friendships and romantic relationships, it is okay for people to seek out only members of their own ethnic or religious group.
Loyalty to one's group is more important than one's individual concerns.
The government should strive to improve the well-being of people in our nation, even if it sometimes happens at the expense of people in other nations.

Authority:
Men and women each have different roles to play in society.
If I were a soldier and disagreed with my commanding officer's orders, I would obey anyway because that is my duty.
Respect for authority is something all children need to learn.
When the government makes laws, those laws should always respect the traditions and heritage of the country.

Purity:
People should not do things that are revolting to others, even if no one is harmed.
I would call some acts wrong on the grounds that they are unnatural or disgusting.
Chastity is still an important virtue for teenagers today, even if many don't think it is.
The government should try to help people live virtuously and avoid sin.

(Graham, Haidt & Nosek, 2009, 1044)

So according to Graham and colleagues, a participant who responds to

> People should not do things that are revolting to others, even if no one is harmed.

with "strongly disagree" (on a 6 point scale that runs from strongly agree to strongly disagree) is making a moral judgment. And a participant who responds to

> The government should strive to improve the well-being of people in our nation, even if it sometimes happens at the expense of people in other nations.

with "strongly agree" (on a 6 point scale that runs from strongly agree to strongly disagree) is making a moral judgment. But Graham & colleagues offer no argument that these participants are making moral judgments. Indeed, they say nothing at all to defend that claim. Presumably, that's because they take it to be intuitively obvious. Would participants who offer these responses also classify these responses as moral judgments if we used Turiel's Criterion Judgment Test or other Turiel Style strategies? Would they classify these responses as moral judgments if we used one of the Ordinary Language Strategies discussed in §3.2. Though I can't claim to have read the 4.3 million papers on Moral Foundations Theory, I have read dozens of them, and I have yet to find one that attempts to answer these questions. So the literature on Moral Foundations Theory generates another massive Tower of Babel Problem. The strategy that Moral Foundations Theorists use to identify moral judgments is dramatically different from the strategies used by researchers inspired by Turiel and by researchers who rely on Ordinary Language probes.

4. Strategies for Dealing With the Tower of Babel Problem

Though there are many more examples that might be offered, I trust I've already done enough to make it plausible that researchers studying moral judgment have indeed used a wide variety of different strategies to determine which judgments they will count as moral judgments. So there is a prima facie case that the cognitive science of moral judgment has a Tower of Babel Problem. The question I want to consider in this final section is "So what?" A bit less flippantly, my question is: How should researchers interested in moral judgment deal with the Tower of Babel Problem? There are many different answers that might be offered, but since space is limited I'll only consider three.

4.1. The Alfred E. Neuman Response

The first response borrows the name of the beloved *Mad Magazine* character, who famously asked "What, me worry?" The central claim of this response is that, apart from rather special circumstances, the fact that researchers use different strategies to identify moral judgments may pose no problem at all, since (perhaps with a few exceptions) all the strategies for identifying moral judgments may actually pick out the same judgments. At a number of places in the preceding sections I have noted that no one has actually studied whether the strategy being discussed picks out the same judgments as a strategy discussed earlier. On other occasions, I have been less cautious and simply asserted that it is overwhelmingly likely that a pair of strategies would diverge substantially in the judgments they classify as moral. I think that both of these observations generalize. If we pick *any* pair of procedures discussed in earlier sections it is indeed overwhelmingly likely that they will differ in their classification of many judgments. But it is also the case that no one has ever done the sort of careful comparative study that would be needed to make a strong empirical case for this claim for any pair of procedures. And it is unlikely that anyone ever will, because designing a convincing experiment would be methodologically challenging, running the experiment would be costly and time consuming, and the scientific payoff would be minimal, since the likely result would surprise no one. The bottom line is that I'm not prepared to take the Alfred E. Neuman Response seriously, though I concede that I have no convincing response to someone who is.

4.2. Build Better Tests

A second response to the realization that researchers are using many different procedures to determine which judgments are moral judgments is to note that much the same situation often obtains in many parts of science. In medicine, for example, there are often a number of different tests for an illness, and those tests sometimes disagree. The Covid epidemic provides an all too salient example. Early on in the epidemic, patients with flu-like respiratory symptoms accompanied by the loss of taste and smell were typically diagnosed with Covid. It was soon recognized that low blood oxygen was an additional symptom of serious infection. Later, rapid antigen tests and more accurate PCR tests became available, and both of these were fine tuned and improved. What we should do in moral psychology, this second response urges, is analogous to what is typically done in medicine. We should try to construct better tests.

I am inclined to think that this analogy may be fatally flawed. Covid tests, and many other tests in medicine and in other areas of science are aimed at detecting a natural kind – a phenomenon characterized by a nomologically linked cluster of properties. So the Covid test analogy, and analogies with the development of test procedures in many other areas of science, is most plausible if moral judgment is a natural kind. There are, however, two lines of research that cast doubt on the hypothesis that moral judgment is a natural kind.

The first of these is the work of Walter Sinnott-Armstrong and Thalia Wheatley (2012, 2014). Using their own intuition to classify judgements as *moral* judgements, they have made persuasive case that these judgments exhibit a wide range of different psychological, neurological, functional and evolutionary properties. Moral judgments – intuitively characterized – don't cluster in a single natural kind but in many quite different natural kinds.

The second line of research picks up the other end of the stick. Rather than relying on intuition to identify moral judgments, it focuses on a specific identification procedure and asks whether that procedure picks out a nomologically linked cluster of properties. Most of the work that pursues this strategy has focused on the Turiel Criterion Judgment Test, and has found that the cluster of psychological properties that that test relies on comes unglued when we attend to the judgments of people in non-Western cultures and transgressions beyond the schoolyard. There has been no shortage of criticism of this work. My own view is that none of that criticism is persuasive. But since I am a co-author of many of the studies being criticized, you would be well advised to read the stuff and make your own decision.[19] Of course, even if I am right that the intuitive strategy picks out too many natural kinds and the Turiel Test doesn't pick our any, it is still possible that some other procedure for identifying moral judgments will be found to identify a substantial class of normative judgments that do indeed form a nomological cluster. And were this happen, it might (or might not!) be plausible to identify these judgments as the only genuine moral judgments. But it is also possible – and I'm guessing likely – that there is no nomological cluster of properties that pick out anything that could be plausibly identified with the class of moral judgments. What should we do then?

4.3. 'Moral Judgment' Eliminativism

The most plausible answer, I think, is that we should give up the idea that 'moral judgment' is a useful theoretical term in cognitive science and in neighboring disciplines like neuroscience and evolutionary biology. We should recognize that researchers who use different strategies for identifying what they call 'moral judgments' are often studying importantly different phenomena. The response to the Tower of Babel Problem that plagues the term 'altruism' can serve as a model for the response I am recommending. Most researchers in the relevant fields no longer use the term 'altruism' without a tacit or explicit acknowledgement that they are interested in just one of the many phenomena that that term has been used to denote. There are no serious debates about whether the phenomena that Daniel Batson studies[20], or the phenomena that Robert Trivers and David Sloan Wilson study[21] really is altruism. Nothing

[19] See Kelly et al. (2007), Sousa et al. (2009), Stich et al. (2009), Fessler et al. (2015), Piazza and Sousa (2016), Fessler et al. (2016), Kumar (2015), Stich (2019).
[20] Batson (1991, 2011).
[21] Trivers (1971), Wilson (2015).

really is altruism. Similarly, I think the right thing to say about moral judgment is that researchers who use the term are talking about many different things, and that nothing *really is* a moral judgment. The take home message from the Tower of Babel Problem in the study of moral judgment is that *there is no such thing as moral judgment*.

References

Badhwar, Neera Kapur (1993). Altruism versus self-interest: Sometimes a false dichotomy, *Social Philosophy and Policy*, 10, 1, 90–117.

Batson, C. Daniel (1991). *The Altruism Question*: *Toward a Social-Psychological Answer*, Hillsdale, NJ: Lawrence Erlbaum Associates.

Batson, C. Daniel (2011). *Altruism in Humans*, Oxford: Oxford University Press.

Bernhard, Regan, Jonathan Chaponis, Richie Siburian, Patience Gallagher, Katherine Ransohoff, Daniel Wikler, Roy Perlis, and Joshua Greene (2016). Variation in the oxytocin receptor gene (OXTR) is associated with differences in moral judgment, *Social Cognitive and Affective Neuroscience*, 11, 12, 1872–1881.

Berniūnas, Renatas, Vilius Dranseika, and Paulo Sousa (2016). Are there different moral domains? Evidence from Mongolia, *Asian Journal of Social Psychology*, 19, 3, 275–282.

Berniūnas, Renatas, Vytis Silius, and Vilius Dranseika (2020). Beyond the moral domain: The normative sense among the Chinese, *Psichologija*, 60, 86–105.

Blair, James (1995). A cognitive developmental approach to morality: Investigating the psychopath, *Cognition*, 57, 1-29.

Buchtel, Emma, Yanjun Guan, Qin Peng, Yanjie Su, Biao Sang, Sylvia Xiaohua Chen, and Michael Bond (2015). Immorality east and west: Are immoral behaviors especially harmful, or especially uncivilized? *Personality and Social Psychology Bulletin*, 41, 10, 1382–1394.

Clavien, Christine, and Michel Chapuisat (2013). Altruism across disciplines: One word, multiple meanings, *Biology and Philosophy*, 28, 125–40.

Dixon, Thomas (2008). *The Invention of Altruism*: *Making Moral Meanings in Victorian Britain*, Oxford: Oxford University Press.

Fessler, Daniel, H. Clark Barrett, Martin Kanovsky, Stephen Stich, Colin Holbrook, Joseph Henrich, Alexander Bolyanatz, Matthew Gervais, Michael Gurven, Geoff Kushnick, Anne Pisor, Christopher von Rueden, and Stephen Laurence (2015). Moral parochialism and contextual contingency across seven disparate societies, *Proceedings of the Royal Society B*, 282.

Fessler, Daniel, H. Clark Barrett, Martin Kanovsky, Stephen Stich, Colin Holbrook, Joseph Henrich, Alexander Bolyanatz, Matthew Gervais, Michael Gurven, Geoff Kushnick, Anne Pisor, Christopher von Rueden, and Stephen Laurence (2016). Moral parochialism misunderstood: A reply to Piazza and Sousa, *Proceedings of the Royal Society B*, 283.

Graham, Jesse, Jonathan Haidt, and Brian Nosek (2009). Liberals and conservatives rely on different sets of moral foundations, *Journal of Personality and Social Psychology*, 96, 5, 1029–1046.

Graham, Jesse, Brian Nosek, Jonathan Haidt, Ravi Iyer, Spassena Koleva, and Peter Ditto (2011). Mapping the moral domain, *Journal of Personality and Social Psychology*, 101, 2, 366–385.

Grant, Colin (1997). Altruism: A social science chameleon, *Zygon*, 32,3, 321–40.

Gray, Kurt, Jennifer MacCormack, Teague Henry, Emmie Banks, Chelsea Schein, Emma Armstrong-Carter, Samantha Abrams, and Keely Muscatell (2022). The affective harm account (AHA) of moral judgment: Reconciling cognition and affect, dyadic morality and disgust, harm and purity, *Journal of Personality and Social Psychology: Attitudes and Social Cognition*, 123, 6, 1199-1222.

Greene, Joshua, R. Brian Sommerville, Leigh Nystrom, John Darley, and Jonathan Cohen (2001). An fMRI investigation of emotional engagement in moral judgment, *Science*, 293, 2105-2108.

Haidt, Jonathan, Silvia Koller, and Maria Dias (1993). Affect, culture, and morality, or is it wrong to eat your dog? *Journal of Personality and Social Psychology*, 65, 4, 613–628.

Haidt, Jonathan and Craig Joseph (2007). The moral mind: How 5 sets of innate intuitions guide the development of many culture-specific virtues, and perhaps even modules, in *The Innate Mind, Volume 3: Foundations and the Future*. Peter Carruthers, Stephen Laurence, and Stephen Stich (eds.), Oxford/New York: Oxford University Press, 367–391.

Haidt, Jonathan (2012). *The Righteous Mind: Why Good People Are Divided by Politics and Religion*, New York: Pantheon Books.

Huebner, Bryce, James Lee, and Marc Hauser (2010). The moral-conventional distinction in mature moral competence, *Journal of Cognition and Culture*, 10(1–2), 1–26.

Huebner, Bryce, Marc Hauser, and Phillip Pettit (2011). How the source, inevitability and means of bringing about harm interact in folk-moral judgments, *Mind & Language*, 26, 2, 210–233.

Iliadis, Andrew (2019). The tower of babel problem: Making data make sense with basic formal ontology, *Online Information Review*, *43*,6, 1021-1045.

Kagan, Jerome and Sharon Lamb (eds.) (1987). *The Emergence of Morality in Young Children*, Chicago: University of Chicago Press.

Kelly, Daniel, Stephen Stich, Kevin Haley, Serena Eng, and Daniel Fessler (2007). Harm, affect and the moral/conventional distinction, *Mind and Language*, 22, 2, 117-131.

Kerr, Benjamin, Peter Godfrey-Smith, and Marcus Feldman (2004). What is altruism? *Trends in Ecology and Evolution*, 19, 3, 135-140.

Killen, Melanie and Judith Smetana (2015). Origins and development of morality, *Handbook of Child Psychology, Volume 3: Social and Emotional Development*, seventh edition, M. Lamb (ed.), New York: Wiley/Blackwell Publishers, 701–749.

Killen, Melanie and Audun Dahl (2018). Moral judgment: reflective, interactive, spontaneous, challenging, and always evolving, *Atlas of Moral Psychology*, Kurt Gray and Jesse Graham (eds.), New York, NY: The Guilford Press, 20–30.

Kumar, Victor (2015). Moral judgment as a natural kind, *Philosophical Studies*, 172, 2887–2910.

Machery, Edouard and Stephen Stich (2022). The moral/conventional distinction, *The Stanford Encyclopedia of Philosophy*, Summer 2022 Edition, Edward N. Zalta (ed.), https://plato.stanford.edu/entries/moral-conventional/

Nichols, Shaun (2002). Norms with feeling: Towards a psychological account of moral judgment, *Cognition*, 84, 221–236.

Nucci, Larry and Elliot Turiel (1978). Social interactions and the development of social concepts in preschool children, *Child Development*, 49, 400–407.

Nucci, Larry and Elliot Turiel (1993). God's word, religious rules, and their relation to Christian and Jewish children's concepts of morality, *Child Development*, 64, 5, 1475–1491.

Parkinson, Carolyn, Walter Sinnott-Armstrong, Philipp Koralus, Angela Mendelovici, Victoria McGeer, and Thalia Wheatley (2011). Is morality unified? Evidence that distinct neural systems underlie moral judgments of harm, dishonesty, and disgust, *Journal of Cognitive Neuroscience*, 23, 10, 3162–3180.

Piazza, Jared and Paulo Sousa (2016). When injustice is at stake, moral judgments are not parochial, *Proceedings of the Royal Society B*, 283: 20152037.

Ramsey, Grant (2016). Can altruism be unified? *Studies in History and Philosophy of Biological and Biomedical Sciences*, 56, 32–38.

Royzman, Edward, Robert Leeman, and Jonathan Baron (2009). Unsentimental ethics: Towards a content-specific account of the moral–conventional distinction, *Cognition*, 112, 159–174.

Royzman, Edward, Geoffrey Goodwin, and Robert Leeman (2011). When sentimental rules collide: 'Norms with feelings' in the dilemmatic context, *Cognition*, 121, 1, 101–114.

Royzman, Edward, Justin Landy, and Geoffrey Goodwin (2014). Are good reasoners more incest-friendly? Trait cognitive reflection predicts selective moralization in a sample of American adults, *Judgment & Decision Making*, 9, 3, 176–190.

Sachdeva, Sonya, Purnima Singh, and Douglas Medin (2011). Culture and the quest for universal principles in moral reasoning, *International Journal of Psychology*, *46*, 161-176.

Schramme, Thomas (2017). Empathy and altruism, *The Routledge Handbook of Philosophy of Empathy*, Heidi Maibom (ed.), New York: Routledge, 203–214.

Schwitzgebel, Eric and Joshua Rust (2014). The moral behavior of ethics professors: Relationships among self-reported behavior, expressed normative attitude, and directly observed behavior, *Philosophical Psychology*, 27, 3, 293–327.

Shein, Chelsea, Ryan Ritter, and Kurt Gray (2016). Harm mediates the disgust-immorality link, *Emotion*, 16, 6, 862–876.

Shweder, Richard, Elliot Turiel, and Nancy Much (1981). The moral intuitions of the child, *Social Cognitive Development: Frontiers and Possible Futures*, John Flavell and Lee Ross (eds.), New York: Cambridge University Press, 288–305.

Shweder, Richard, Manamohan Mahapatra, and Joan Miller (1987). Culture and moral development," in Kagan and Lamb (1987), 1–83.

Sinnott-Armstrong, Walter and Thalia Wheatley (2012). The disunity of morality and why it matters to philosophy, *The Monist* 95, 3, 355–377.

Sinnott-Armstrong, Walter and Thalia Wheatley (2014). Are moral judgments unified? *Philosophical Psychology*, 27, 4, 451-474.

Skitka, Linda, Christopher Bauman, and Edward Sargis (2005). Moral conviction: Another contributor to attitude strength or something more? *Journal of Personality and Social Psychology*, 88, 6, 895–917.

Skitka, Linda, Christopher Bauman, and Brad Lytle (2009). The Limits of legitimacy: Moral and religious convictions as constraints on deference to authority, *Journal of Personality and Social Psychology*, 97, 4, 567–578.

Skitka, Linda and Daniel Wisneski (2011). Moral conviction and emotion, *Emotion Review*, 3, 3, 328–30.

Skitka, Linda, Brittany Hanson, and Daniel Wisneski (2017). Utopian hopes or dystopian fears? Understanding the motivational underpinnings of morally motivated political engagement, *Personality and Social Psychology Bulletin*, 43, 2, 177–190.

Skitka, Linda, Brittany Hanson, G. Scott Morgan, and Daniel Wisneski (2021). The psychology of moral conviction, *Annual Review of Psychology*. 72, 347–66.

Smetana, Judith (1993). Understanding of social rules, *The Child as Psychologist: An Introduction to the Development of Social Cognition*, Mark Bennett (ed.), New York: Harvester Wheatsheaf, 111–141.

Smetana, Judith, Marc Jambon, and Courtney Ball (2018). Normative changes and individual differences in early moral judgments: A constructivist developmental perspective, *Human Development*, 61, 4–5, 264–280.

Sober, Elliott and David Sloan Wilson (1998). *Unto Others: The Evolution and Psychology of Unselfish Behavior*, Cambridge, MA: Harvard University Press.

Sousa, Paulo, Colin Holbrook, and Jared Piazza (2009). The morality of harm, *Cognition*, 113, 80–92.

Stich, Stephen, Daniel Fessler, and Daniel Kelly (2009). On the morality of harm: A response to Sousa, Holbrook and Piazza, *Cognition*, 113, 93–97.

Stich, Stephen, John Doris & Erica Roedder (2010). Altruism, *The Moral Psychology Handbook*, John Doris and the Moral Psychology Research Group (eds.), Oxford: Oxford University Press, 147-205.

Stich, Stephen (2019). The quest for the boundaries of morality, *The Routledge Handbook of Moral Epistemology*, Aaron Zimmerman, Karen Jones, and Mark Timmons (eds.), New York: Routledge, 15–37.

Trivers, Robert (1971). The evolution of reciprocal altruism, *The Quarterly Review of Biology*, 46, 35–57.

Turiel, Elliot (1977). Distinct conceptual and developmental domains: Social convention and morality, *Nebraska Symposium on Motivation*, 25, 77–116.

Turiel, Elliot (1983). *The Development of Social Knowledge: Morality and Convention*, (Cambridge Studies in Social and Emotional Development), Cambridge/New York: Cambridge University Press.

Turiel, Elliot, Melanie Killen, and Charles Helwig (1987). Morality: Its structure, functions, and vagaries, in Kagan and Lamb 1987: 155–243.

Wallace, Gerald and Arthur Walker (eds.) (1970). *The Definition of Morality*, London: Methuen.

Wilson, David Sloan (2015). *Does Altruism Exist?* New Haven: Yale University Press.

Wright, Jennifer Cole, Jerry Cullum, and Nicholas Schwab (2008). The cognitive and affective dimensions of moral conviction: Implications for attitudinal and behavioral measures of interpersonal tolerance, *Personality and Social Psychology Bulletin*, 34, 11, 1461–1476.

Young, Liane and Rebecca Saxe (2011). Moral universals and individual differences, *Emotion Review*, 3, 3, 323–324.