

## Some Questions About *The Evolution of Morality*<sup>1</sup>

Stephen Stich  
Rutgers University

Richard Joyce has written an admirable book, brimming over with fascinating findings, bold empirical hypotheses and philosophical arguments that are both innovative and provocative, all set out in a straightforward, engaging style. One of the virtues of this journal's book symposia that they give commentators an opportunity to ask questions that authors can address in their responses. But symposium articles must also be short, and by the time I had finished my second reading of Joyce's book, I had a list of questions that would fill many more pages than I am allowed. So, for want of a better strategy for narrowing down the list, I'll focus on questions that were suggested by apparent differences between Joyce's account of our "moral sense" and the account of the psychology of norms that Chandra Sripada and I have defended in a recent paper (Sripada & Stich, 2006). To fill in the necessary background, I'll begin with a very brief overview of the Sripada & Stich (S&S) model.

Figure 1 is a sketch of the psychological mechanisms which, Sripada and I argue, underlie the acquisition and implementation of norms. The job of the Acquisition Mechanism is to identify the norms in the surrounding culture whose violation is typically met with punishment, to infer the content of those norms, and to pass that information to the Execution Mechanism, where it is stored in the Norm Data Base. The Execution Mechanism has the job of inferring that some actual or contemplated behavior violates (or is required by) a norm, and generating intrinsic (i.e. non-instrumental) motivation to comply and to punish those who do not comply. There is good reason to believe that the emotion system is involved in punitive motivation and it may also play a role in compliance motivation, though the evidence for that is less persuasive. Influenced by the remarkable findings reported in Wheatley and Haidt (2005), Figure 1 portrays the making of moral judgments to be downstream from the emotion system. In Wheatley and Haidt's study, participants who were hypnotized to feel disgust when they read the word 'often' or 'take' made much more severe moral judgments about behavior described using one of these words than they made when the behavior was described without using the words. However, following Greene (2004, Green et al. 2001), who has demonstrated very different patterns of brain activity in response to different sorts of moral dilemmas, we also included a second pathway leading to moral judgment which involves the explicit reasoning system and may not involve the norm and emotion systems at all. While Greene's account of the sorts of dilemmas that do not

---

<sup>1</sup> I am grateful to Edouard Machery and Chandra Sripada for helpful comments on an earlier draft of this paper.

engage the emotion centers in the brain has been evolving steadily as new data become available, the rough idea is that they are relatively impersonal cases rather than those in which the interactions among agents are (as Greene used to say) “up close and personal.” For present purposes, that’s all we’ll need about the S&S model, so let me turn to Joyce’s book.

The central question in the first four chapters of *The Evolution of Morality* is “Is human morality innate?” (1)<sup>2</sup>, and Joyce does an admirable job of saying how he will interpret the question. To ask whether morality is *innate* is to ask whether it “can be given an adaptive explanation in genetic terms: whether the present-day existence of the trait is to be explained by reference to a genotype having granted ancestors reproductive advantage.” (2) To ask whether *morality* is innate is to ask “whether the human capacity to make *moral judgments* is innate.” (4, emphasis added) As Joyce wisely notes, in order to address that question seriously, we need an account of what moral judgments *are*. Thus much of chapter 2 is devoted to a detailed account of the nature of moral judgments.

One crucial feature of moral judgments, on Joyce’s account, is that they are imbued with a kind of “practical clout” (or “oomph” as Joyce sometimes says) – they “draw attention to a deliberative consideration that cannot be legitimately be ignored or evaded.” (58) Moreover, this practical oomph “doesn’t have its source in internal or external sanctions, nor in some institution’s inviolable rules, nor in the desires or goals of the person to whom it is addressed. In this respect ordinary thought distinguishes moral requirements from conventional and prudential requirements.” Joyce goes on to note that “[t]here is a large body of empirical evidence ... demonstrating that even very young children make these distinctions.” (63) The empirical literature that Joyce is alluding to here is the work by Eliot Turiel and others that utilizes the “moral / conventional task”. (Turiel 1983; Nucci 2001)

I am inclined to think that the sort of architecture sketched in Figure 1 can go a long way toward explaining the “oomph” that looms large in Joyce’s account of moral judgment. For if a judgment is generated by the norm execution mechanism, then those who make the judgment are intrinsically motivated to comply with that judgment and to punish those who do not. Also, as Daniel Kelly and I have argued (Kelly & Stich, forthcoming), judgments generated by the norm execution mechanism will strike those who make them as “authority independent” in the sorts of experiments that Turiel and his associates typically employ. When participants in these experiments are asked to suppose that an authority figure has decreed that there is no rule prohibiting a transgressive action which violates a norm stored in the data base, this will have no impact on their motivation to comply with the rule and to punish the transgressions.

---

<sup>2</sup> All references to Joyce’s book will be given in parentheses in the text. And, in case you were wondering, Joyce believes the answer to the question, as he interprets it, is yes.

There are, however, other features of the S&S model which comport less well with Joyce's account of moral judgment. One of these is the "second pathway" to moral judgment, the one which does not involve the norm execution mechanism or the emotion system. If there are moral judgments generated in this way, they pose a pair of problems for Joyce. First, it is far from clear where *these* judgments get their "practical clout" since there is no intrinsic motivation to comply with them or to punish those who don't. Second, moral judgments generated in this way would pose a problem for Joyce's projectivist account of moral phenomenology. According to Joyce, "moral attributes seem to be 'in the world'" but "moral appearances are in fact caused largely by emotional activity. A corollary is that appearances are to some extent deceptive; though our judgments are in fact prompted by emotional activity, our phenomenology is one as of the emotional activity being a response to attributes instantiated in the world." (128-9) There is much to commend in Joyce's discussion of projectivism; it makes a promising start at analyzing and explaining important aspects of the phenomenology of those moral judgments that are "caused largely by emotional activity." But, of course, the projectivist account does not apply to judgments generated via the second route – the one in which the emotion system plays little role.

There are a number of ways in which Joyce might respond to these problems. Perhaps the simplest and boldest way would be to deny that there is "second route" to moral judgment which does not involve the emotion system. Another option would be to offer some non-projectivist explanation of the objectivist phenomenology and practical clout of moral judgments generated via the second route. These are not the only options, but rather than continuing to speculate, let me ask the author: What *is* your view about second route moral judgments? Do you think that they don't exist? If they do exist, what sort of account would you offer of their phenomenology and their clout?

These questions turn on a feature of the S&S model that seems to find no place in Joyce's account. Let me turn now to a feature of Joyce's account that plays no role in the S&S model. According to Joyce, the emotion of guilt "surely lies at the core of the moral conscience" (122-3), and conscience is unpacked as "a repertoire of judgments and emotions (most notably guilt) that motivate behavior in accordance with accepted standards of conduct even when external sanctions are absent." (120) So, for Joyce, guilt plays a central role in motivating moral behavior. On the S&S model, by contrast, guilt is accorded no special role. Since the model allows that the emotion system *might* be involved in compliance motivation, it is not incompatible with the claim that guilt is important in moral motivation. But I am rather skeptical of the proposal, since I find it hard to see how it is supposed to work. Guilt, after all, is an emotion one has typically has *after* one has committed some transgression. As Joyce puts it, "[g]uilt seems most naturally to associate with the judgment that the person *has performed* a wrongful action for which amends might be made." (102, emphasis added) But if guilt is an emotion one feels after one has performed a wrongful action, how,

exactly, does it “motivate behavior in accordance with accepted standards of conduct?”

One familiar idea is that people believe that they will feel guilty if they violate one of the norms they have internalized, and that they are motivated not to violate the norm since they also believe that guilt is a very unpleasant emotion, and they want to avoid having that unpleasant experience. This is, however, a singularly implausible account of the phenomenology of *my* moral motivation when, for example, I decide to return a lost wallet or not to tell a convenient lie. And informal surveys among my students confirm that I am not unique. Indeed, these surveys suggest that concern about future guilt plays almost no role deciding what to do, except when the student has been raised in a religious family and the behavior being contemplated is sexual behavior. Even if I am quite wrong about the phenomenology of moral decision making – or if I am right about the phenomenology and the thoughts about future guilt are typically unconscious – it still would not support Joyce’s contention that guilt plays a major role in moral motivation. For on the account we are considering, *the emotion of guilt* is playing no role in the generating compliance motivation. Rather it is the *belief that one will feel guilty* and the desire to avoid this feeling that are doing all the work. Joyce might, I suppose, suggest that the emotion of guilt plays a crucial role in *producing and sustaining* that belief, because people have learned via inductive inference that transgressions lead to guilt. But I know of no evidence that even begins to suggest that people learn the link between transgressing and feeling guilty in this way. Rather than speculating further, let me ask Joyce: Do you think that the emotion of guilt (rather than beliefs about the emotion) plays an important role in motivating people to act in accordance with prevailing norms even though guilt is typically experienced after a transgression has taken place? If so, can you provide some further details on how this works?

Joyce’s account of moral judgment is rich and complex, and while most of the details are compatible with the S&S model, few of them would be predicted by that model. For example, according to Joyce, in order for an utterance, S, to count as a moral judgment there must be a “linguistic convention that decrees that when S is uttered [in an appropriate context] the speaker thereby expresses *two* mental states” (57, cf. 53); one of these mental states is a belief, the other is “a connotative attitude” “such as approval, contempt, or, more generally, *subscription to standards.*” (70, emphasis added) Thus, he maintains, a pair of sentences like:

(1) The Elgin Marbles morally ought to be returned to Greece. But I subscribe to no moral standard that commends their return to Greece.

“would be challenged if uttered....”(56) There are, I suspect, many philosophers who would take issue with this (and other) features of Joyce’s account of moral judgment. Moral particularists, for example, might well balk at Joyce’s insistence

that making moral judgments requires “subscription to standards”. (Dancy, 2005) But even if we grant that Joyce’s characterization of moral judgments is correct, its richness and complexity pose a problem. For if moral judgment requires *all of that*, what reason is there to think that people in cultures very different from ours *make* moral judgments? Why should we think that making moral judgments is a pan-cultural phenomenon? The question is an important one for a project like Joyce’s since, as Joyce recognizes, if he is right that human morality is innate, we should expect it to be present in all cultures, with the exception, perhaps, of those that are so stressed that normal psychological and social processes break down. Joyce clearly believes that “morality (by which I here mean *the tendency to make moral judgments*) exists in all human societies we have ever heard of.” (134, emphasis in the original). But once we realize how much Joyce has built into the notion of a moral judgment, the evidence he offers for this claim seems far from convincing. “Moral precepts,” he tells us, “are mentioned in the Egyptian Book of the Dead and in the Mesopotamian epic of Gilgamesh.... Moreover, morality exists in virtually every human individual. It develops without formal instruction, with no deliberate effort, and with no conscious awareness of its special features.” (134-5) And, lest we mistakenly interpret him as talking loosely here, Joyce adds: “When I talk here of ‘moral development’ I don’t just mean prosocial behavior or even simply prosocial emotions; I mean genuine cognitive ... moral judgments.” (135)

As I see it, these considerations (and those that Joyce offers in the next few pages) don’t come close to supporting his claim that the tendency to make the sort of rich and complex moral judgments that he has gone to such pains to characterize exists in all human societies. To the best of my knowledge, we have no serious information about the details of the linguistic conventions that prevailed in the communities that produced the epic of Gilgamesh or the Book of the Dead. To support his contention that “morality exists in virtually every human individual,” Joyce appeals to work in the Turiel tradition. Researchers in that tradition have maintained that the capacity to draw the moral/conventional distinction is pan-cultural and emerges early in development. But there is a growing body of literature indicating that it is simply false that there is a pan-cultural ability to draw the “moral/conventional” distinction as characterized by Turiel and his associates. Indeed, as I read that literature, the right conclusion to draw is that the moral/conventional distinction, as characterized by Turiel and his followers, is a myth.<sup>3</sup> Moreover, I suspect that the practice of making moral judgments of the sort that Joyce describes is a culturally and temporally local one restricted to Western (and Western-influenced) cultural groups in relatively recent times. Of course, this suspicion would be substantially undermined if there was evidence that folks in a number of cultures very different from our own really do make Joyce-style moral judgments. Richard, do you know of any such evidence?

---

<sup>3</sup> For more on this admittedly controversial claim, see Kelly et al. (2007), Kelly and Stich (forthcoming), Nado, Kelly and Stich (forthcoming).

Though the S&S model says nothing about the evolution of the mechanisms it posits, the model does pose a puzzle for Joyce's account of the evolution of morality. Though that account is complex and nuanced, two ideas are quite central. The first is that the core evolutionary function of moral judgment is to get people to *behave* in appropriate ways. "My thinking on this matter," Joyce tells us, "is dominated by the natural assumption that an individual sincerely judging some available action in a morally positive light increases the probability that the individual will perform that action...." (109) The second idea, and the one I propose to question, is that the primary sort of behavior moral judgment was selected to motivate is *cooperative or prosocial* behavior. Here is how Joyce makes the point.

[S]elf-directed moral judgment may enhance reproductive fitness so long as it is attached to the appropriate actions. We have already seen that the "appropriate actions" – that is, the fitness enhancing actions – will in many circumstances include helpful and cooperative behaviors. Therefore it may serve an individual's fitness to judge certain prosocial behaviors – *her own* prosocial behaviors – in moral terms. (109)

The benefits that may come from cooperation ... are typically long term values, and merely to be aware of and desire these long term desires does not guarantee that the goal will be effectively pursued.... The hypothesis, then, is that natural selection opted for a special motivational mechanism for this realm: moral conscience. (111)

On the S&S model, the norm acquisition system is designed to internalize whatever norms prevail in the surrounding environment. So if there are prosocial norms or norms of cooperation, they will be acquired. And, as Joyce rightly notes, "all human moral systems give a leading role to *reciprocal relations*." (140). But, as Sripada and I note, norms of cooperation are just one among many sorts of norms that are to be found in just about every culture.

[M]ost societies have rules that prohibit killing, physical assault and incest (or sexual activity with one's kin).... Most societies have rules regulating sexual behavior among various members of society, and especially among adolescents.... Examples like these could be multiplied easily in domains such as social justice, kinship [and] marriage .... [Most societies also have norms] governing what food can be eaten, how to dispose of the dead, how to show deference to high ranking people, and a host of other matters. (Sripada & Stich, 2006)

Since norms governing all of these matters are as ubiquitous as norms governing reciprocity, it strikes me as rather implausible that reciprocity and prosocial norms should have pride of place in an account of the evolution of morality. Moreover, there are other suggestions about the evolution of norms in which prosocial and cooperative norms play no special role. (Boyd forthcoming;

Sripada forthcoming) Joyce does not deny that other processes may have played a role in the evolution of morality. Indeed, he suggests that “[g]roup selection – most probably at the cultural level – may well have been a major factor.” His “hunch” however, “is that reciprocity, broadly construed, is what got the ball rolling.” (141) Since Joyce offers no argument for his hunch, my last question is: Why does he think that an account which gives reciprocity a central role in the evolution of morality is a better bet than competing accounts in which reciprocity plays no special role?

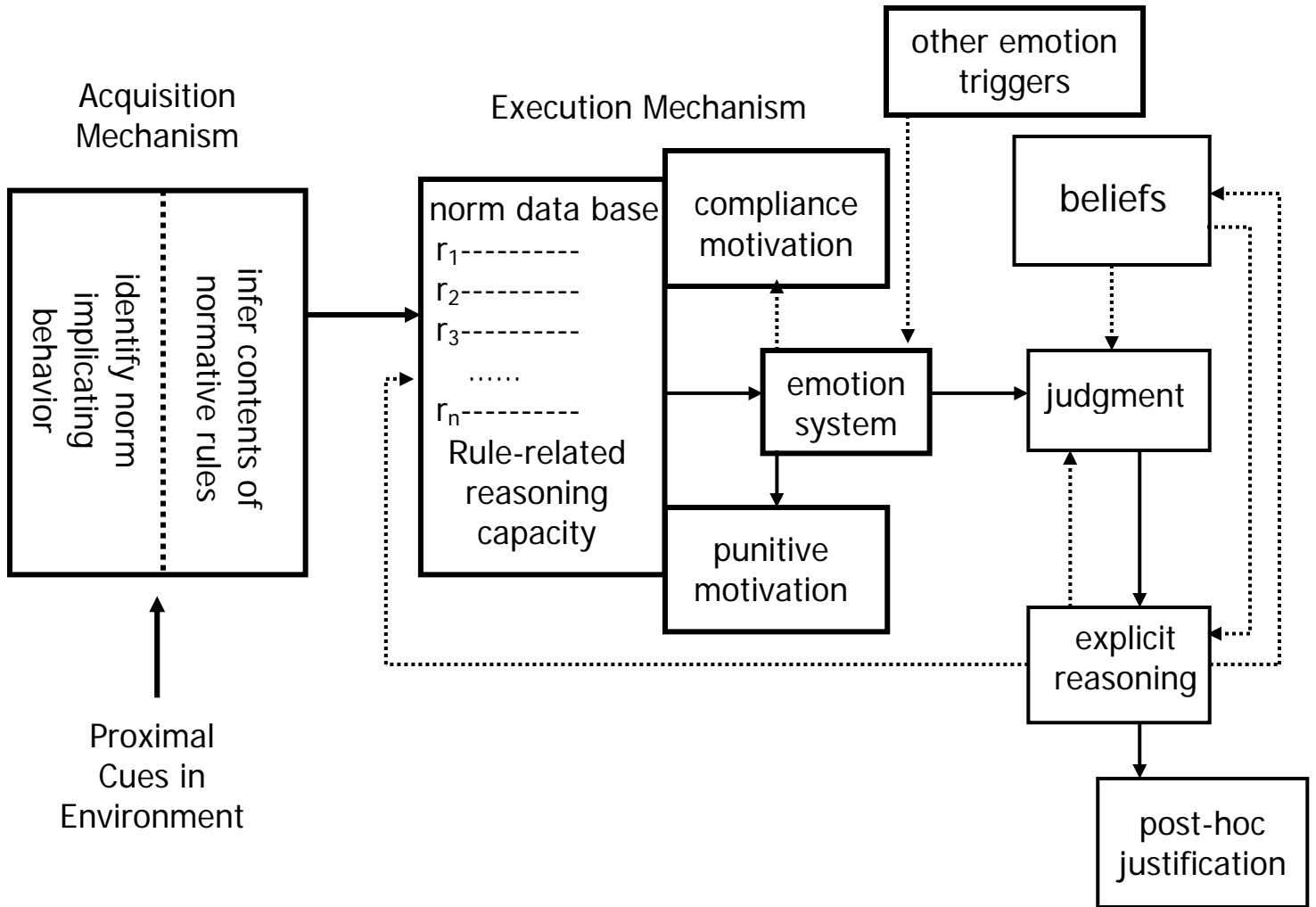


FIGURE 1

A sketch of the mechanisms underlying the acquisition and implementation of norms set out in Sripada & Stich (2006). Solid lines indicate links that we take to be well supported by evidence; dotted lines indicate more speculative links.

## REFERENCES

- Boyd, R. (forthcoming). Population structure, equilibrium selection and the evolution of norms. To appear in *Economics and Evolution*, Ugo Pagano ed., Cambridge University Press.
- Dancy, J. (2005). Moral particularism. In *The Stanford Encyclopedia of Philosophy (Summer 2005 Edition)*, Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/sum2005/entries/moral-particularism/>.
- Greene, G. (2004). fMRI studies of moral judgment. Unpublished lecture given at the Dartmouth College Conference on The Psychology & Biology of Morality.
- Greene, J., Sommerville, R., Nystrom, L., Darley, J., & Cohen, J. (2001). An fMRI investigation of emotional engagement in moral judgment. *Science*, Vol. 293, Sept. 14, 2001, 2105-2108.
- Kelly, D., Stich, S., Haley, K., Eng, S. & Fessler, D. (2007). Harm, affect and the moral / conventional distinction. *Mind and Language*, 22, 2, 117-131.
- Kelly, D. & Stich, S. (forthcoming). Two theories about the cognitive architecture underlying morality. To appear in P. Carruthers, S. Laurence & S. Stich, eds., *Innateness and the Structure of the Mind: Foundations and the Future*. (New York: Oxford University Press) 2007.
- Nado, J., Kelly, D. & Stich, S. (forthcoming). Moral judgment. To appear in the *Routledge Companion to the Philosophy of Psychology*, ed. by John Symons & Paco Calvo.
- Nucci, L. 2001. *Education in the Moral Domain*. Cambridge: Cambridge University Press.
- Sripada, C. & Stich, S (2006). A framework for the psychology of norms. In P. Carruthers, S. Laurence & S. Stich, eds., *The Innate Mind: Culture and Cognition*. (New York: Oxford University Press) 2006. Pp. 280-301.
- Sripada, C. (forthcoming). *Adaptationism, culture and the malleability of human nature*. To appear in P. Carruthers, S. Laurence & S. Stich, eds., *Innateness and the Structure of the Mind: Foundations and the Future*. (New York: Oxford University Press) 2007.
- Turiel, E. 1983: *The Development of Social Knowledge*. Cambridge: Cambridge University Press.

Wheatley, T., & Haidt, J. (2005). Hypnotically induced disgust makes moral judgments more severe. *Psychological Science*, 16, 780-784.