This week we explored the relationship between moral reasoning and social norms. We examined whether emotions are the consequence of a rational process or the source of the premises on which rationality operates. It seems that the moral emotions are a form of self-regulation, and possibly constitute a learning process whereby we anticipate and respond to social feedback. It seems further that while the emotional foundations of morality may be consistent across cultures, these moral foundations will find various expressions across different societies.

Jonathan Haidt argues that intuitive moral emotions drive morality, and that moral reasoning is a form of post-hoc justification that we use to convince others (but, crucially, not ourselves). The moral emotions, on this understanding, would be a mechanism to prevent individuals from realizing short-term benefits inimical to long-term success. Haidt concludes that moral reasoning is an epiphenomenon, and he proposes a "social intuitionist" model whereby moral emotions are primary but individuals deploy moral reasoning to persuade others.

Some experimental evidence seems to bear out Haidt's argument. Daniel Fessler (2004) conducted studies in Bengkulu (Indonesia) and California (USA), determining that self-reported shame was more common in Indonesia and that self-reported guilt was more common in California.

Haidt's assertion that our moral emotions are hypocritically self-justifying sits uneasily with his view that the same moral emotions prod us towards prosocial actions that redound to our long-term benefit. Surely if moral emotions enable group cooperation they must consist of something more than mere ratiocinated self-interest. His distinction between "Western" and non-"Western" cultures also seems spurious.

Elliot Turiel vigorously disputes Haidt's conclusion that moral reasoning has an exclusively external function. He argues that moral judgments begin at a very young age, and that they are distinct from social and personal judgments. On this account, emotion and reason are intimately intertwined and analytically inseparable. Turiel rejects a relativistic account of morality across cultures, arguing that the distinction between so-called individualistic and collectivist cultures is unfounded (p.487).

He also rebuts Haidt's proposed theory of moral intuitions, invoking research that questions response speed as a measure of reason. Turiel takes issue with moral-psychological research along the lines of the infamous "trolley problem", arguing that in fact such scenarios produce conclusions that are not generalizable. Finally, Turiel disputes accounts that emphasize the primacy of emotion over reason, pointing out that children as young as three are able to differentiate easily between issues of convention and morality, and (contra Piaget) they readily label

harmful acts as wrong even when told by authority figures that the act in question is permitted.

Turiel's critique of Haidt seems decisive, and his reminder to not reify concepts like "the West" is instructive. However, the precise definition of reason at work in the arguments of Turiel and Haidt seems inconsistent. Turiel's proposed authority-independent foundation of morality, if true, implies that cultural difference occludes significant commonality regarding the most freighted moral questions. On this point, Piazza and Sousa (2016) reexamined a prior cross-cultural study (Fessler et al. 2015) that had purported to show moral parochialism. After reanalyzing the results, Piazza and Sousa find that in cases involving harm or injustice, the parochialism effect disappears and judgments are highly correlated across cultures.

Sousa and Piazza (2014) extend Turiel's framework by proposing that harmful transgressions are seen as authority-independent and general in scope if the causation of harm is interpreted as involving "basic-rights violation and injustice." It is this injustice, rather than the harm itself, that makes the transgression a *moral* transgression. In a related study, Finger and her coauthors (2006) find that brain regions associated with moral reasoning were activated in different ways by prompts involving moral transgressions and social transgressions.

The role of anticipation in norm construction and maintenance appears to be underemphasized. Mackie et al. (2015) argue that social norms can be maintained by approval or disapproval within a reference group, and that this approval or disapproval is often conveyed by facial expressions. Anticipation of (positive or negative) sanction can lead us to change our behavior. The face itself appears to be crucial in generating these effects. Liu et al. (2019) note that face-to-face interactions are more effective in inducing compliance than other forms of interaction.

Geoffrey Brennan and Philip Pettit (2004) propose a widespread market mechanism for the exchange of esteem. The authors argue that esteem is an evaluative, comparative and directive attitude, which is to say that esteem is given or withheld on the basis of specific actions taken and the success or failure of those actions relative to the performance of others. Esteem, in other words, includes a core element of interpersonal competition.

Probing the relationship between esteem and social norms more deeply, Richard McAdams argues that as long as people seek esteem as an end in itself, then norm formation is inevitable. His proposed mechanism is that particular behaviors will cause many people to grant or withhold esteem, and that this coordination is well-known. If these conditions hold, even a weak concern for esteem can create significant costs for acting against the consensus, which can lead to norm emergence.

We now have some of the tools to better understand moral reasoning. The anticipatory role of emotion may actually be its most important feature. Baumeister et al. (2007) propose understanding emotion as a feedback system with indirect influence on behavior. Rather than directly influencing behavior (which the authors argue would be maladaptive), emotion retrospectively associates strong affect with past experience, thereby making particular patterns of behavior either more or less likely. Limited experimental evidence bears out this view. In a meta-analysis of studies of emotion, DeWall et al. found that direct causation of behavior was only significant in 22% of tests, while the emotion-as-feedback perspective received support in 87% of tests.

We can now give a much more satisfying account of the emotional basis of norm accretion. Emotions seem to exist to guide future behavior, providing anticipatory guidance and moving people towards outcomes associated with positive affect. Mutual anticipation in strategic contexts can result in the emergence of behavior-guiding norms. Our moral emotions appear to rely on distinct brain systems from those underpinning social conventions, but these consistent moral emotions can be refracted by a multiplicity of arbitrary conventional expectation and situational context, giving rise to considerable diversity in social norms.

986 words.