

November 2010

APA Symposium on Passion and Impartiality: Passions and Emotions in Moral Judgment

Constructive Sentimentalism: Legal and Political Implications

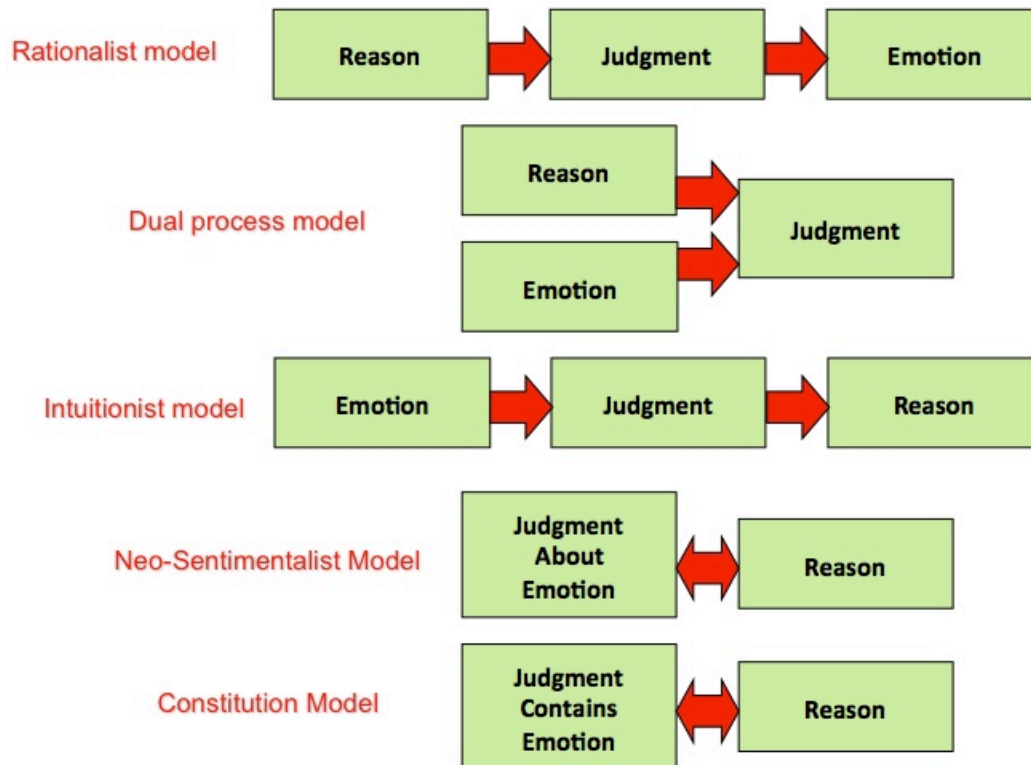
Jesse Prinz

There is mounting empirical evidence linking emotions to moral judgment. Though open to competing interpretations, this evidence is best interpreted as supporting the kind of sentimentalist theory associated with philosophers of the Scottish enlightenment. Recent findings also allow us to update sentimentalism by specifying which emotions contribute to moral judgment, and work in history, anthropology, and cross-cultural psychology is proving richer insights into the origins of our emotionally grounded values. If values are emotionally based and culturally diverse, there may be moral conflicts that have no rational resolution. This would have implications for normative ethics, for politics, and for law. Here I will review the empirical case for sentimentalism and then draw attention to some of these implications.

1. The Place of Emotions in Moral Judgment

The empirical turn in ethics has been fuelled, in part, by the emergence of moral neuroscience. In 2001, Josh Greene and his collaborators published a paper showing neural activations as people reflected on trolley dilemmas. Since then, scores of other studies have appeared. Brains have been scanned as people make judgments of wrongness, engage in reciprocal exchanges, engage in charitable giving, play morally significant video games, and look at morally meaningful photographs. Throughout the many studies, one common denominator has been emotion. Again and again, areas of the brain associated with emotional response are active when people engage in moral cognition. These areas include the posterior cingulate, temporal pole, insula, ventromedial prefrontal cortex, the amygdala, and the ventral striatum. Authors of the studies interpret these results in different ways, but one common refrain is that emotions contribute to moral judgment.

The exact nature of that contribution, however, is difficult to assess using fMRI, which is a correlational method. We can see this by considering the wide range of models that are compatible with the finding that emotional activations regularly co-occur with moral judgment. Figure 1 illustrates some of the possibilities.



The Rationalist Model says that some kind of reasoning – either conscious deliberation or unconscious rules – drives our moral judgments, with emotions arising as a consequence (Hauser). Dual Process Model says that emotion drives moral judgments some of the time, but reason can also, depending on the case (Greene). The Intuitionist Model says that emotions constitution intuitions about what is right or wrong, and we use these intuitions to make our judgments; reason then follows to provide post-hoc rationalizations (Haidt). Neo-sentimentalists claim that moral judgments are judgments about whether emotional responses are merited (I should feel guilty/angry about this), and we can imagine that emotions typically arise in conditions where we deem them appropriate (McDowell, Gibbard, D’Arms and Jacobson). The Constitution Model, which was endorsed by the sentimentalists of the Scottish Enlightenment, says that emotions are components of moral judgments: to think that something is morally wrong is to have a negative emotion towards it (Hutcheson, Hume, Smith, Ayer, Stevenson, Prinz). Both the Neo-Sentimentalists and the old-school Sentimentalists agree with Intuitionists that reason often plays a post-hoc role in the moral domain, but, as we will see in a moment, reason can also make a more substantive contribution.

Given the ambiguity of imaging evidence, how do we decide between these models? I think there are good reasons to be dubious about all but the Constitution model. The evidence is behavioral and philosophical. The Rationalist Model neither explains nor predicts the well established fact that emotions can influence moral judgements. For example, people assess vignettes as more wrong than the otherwise would if they are hypnotized to feel disgust (Haidt), exposed to noxious

smells (Schnall), or asked to imbibe bitter beverages (Eskine). People make more utilitarian judgments when amused (Valdesolo and DeSteno), and more deontological judgments when feeling elevated (Strominger). Clearly emotions are not just an effect of moral judgment.

The Dual-Process model has been supported by appeal to the fact that consequentialist judgments show less activation in emotion centers than deontological judgments in trolley dilemmas, and individuals with ventromedial brain injuries—known for disrupting emotion-based inferences—are more likely than others to make consequentialist judgments. These findings are intriguing, but the fMRI data also clearly show that consequentialist judgments show more emotional activation than non-moral judgments. Likewise, ventromedial patients do not lack emotions; the very fact that they engage in reward seeking behavior (Damasio) shows that they are at least capable of experiencing and acting on appetitive emotional states. Their deficit principally involves an inability to curtail reward-seeking in light of negative feedback. When confronted with a trolley dilemmas, we must normally decide between two moral injunctions: It would be good to help the people in need, and it would be bad to harm someone in the process. These can be regarded as a positive and a negative norm respectively. VM patients seem to motivated by the former, and indifferent to the latter. They don't lack emotions; they simply lack the ability to regulate positive in light of negative emotions. The Dual Process Model is under-motivated.

The Intuitionist Model advanced by Jon Haidt is an improvement, but remains, in an important way, obscure. Haidt suggests that emotions *precede* moral judgments. That implies that they are not components parts. What, then, are moral judgments? It can't just be that they are sentences of English, like "Cannibalism is wrong," because one can make a moral judgment without verbalizing it. Also, if moral judgments are the *effects* of emotions, then they should be able to occur without emotions, just because most effects can come about in different ways. But Haidt offers no evidence that we can make moral judgments without emotions, and that would go against the spirit of his approach. A further issue concerns Haidt's very strong skepticism about the role of reason in morality. He gives the impression that reasoning never contributes to moral deliberation. That seems unlike. We often need reason to determine whether something is morally significant. This is especially clear in policy decisions. In inheritance tax unjust? Is late-term abortion permissible? Should factory farming be regulated? Should we fight to stop vaginal circumcision? Should we assist in foreign wars? Haidt would have us believe that such cases are rare, or that that the reasoning here always involves some kind of blind social conformity, but there is little reason for such a cynical view.

Let's turn from these psychological theories to an account that has gained currency in philosophy: Neo-Sentimentalism. This turn covered a range of positions, but they share in common what can be called the meta-move. They say that moral judgments are judgments about the merit, warrant, or appropriateness of an affective response. The meta-move is designed to improve on traditional sentimentalism, which says that thinking something is wrong is a matter of having a negative emotion towards it. Clearly we sometimes have negative feelings towards things that we would not, on reflection, view as wrong. The wrong is not simply that

which causes our disapproval; it is that merits warrants it. Or so the story goes. But the view faces some serious worries. First, it seems to mislocate the object of moral judgments; when we say that killing is wrong, we are saying something about killing, not about our feelings (Schroeter). Second, it is hard to define merit without circularity. Depending on circumstances, a murder might merit fear (prudentially speaking) or forgiveness (if we aim for reconciliation). We say that killing merits anger is to say that it merits it morally, even if these other considerations make other emotions more appropriate overall. But it would be circular to define a moral judgment as a judgment that emotions are morally merited, because that it just another moral judgment (D'Arms and Jacobson). There is also an empirical worry: some individuals (children and some people with autism) make moral judgments easily but lack the capacity to form beliefs about emotional states (Nichols).

This brings us to the Constitution Model, which says that moral judgments *contain* emotions. To judge that something is wrong, on this view, is to have a negative emotion towards it. This seems to be the kind of view Hume and his contemporaries had in mind, and it overcomes all the difficulties facing the other models. The most obvious objection to the constitution model is that people often seem to have negative emotions while withholding moral judgments. A person raised in homophobic community might later in life experience disgust when seeing homosexual affection while insisting that the observed activity is morally acceptable. Doesn't this show that moral judgments are not constituted by emotions? An alternative explanation is that such an individual *does* in fact think homosexuality is wrong at some level, but also thinks it's permissible and identifies with the latter conviction. Compare the person who exhibits implicit racism, but also believes in racial equality. We should say in both cases, there is a bigoted automatic appraisal that happens to get outweighed by a considered appraisal. In a questionnaire study, I was able to show that this, in fact, is how ordinary people interpret such cases.

Another worry about the Constitution Model pushes in the opposite direction, pointing out that we can make moral judgments without emotions. Empirically, this has not been explored, but cases are easy enough to imagine. When we speak in generalities (cruelty is wrong), or about complex policies (sin tax is wrong), or during episodes of numb depression, we might not feel strong moral emotions. But here I would caution that we must recognize that there are at least dispositions to emote. Someone who did not shudder at a case of cruelty could not be credited with truly believing the generalization that cruelty is wrong. I like to distinguish emotions, which are occurrent states, from sentiments, which are emotional dispositions. I think moral *judgments* usually contain emotions, but our long-standing moral *values* are sentiments. In some cases, when we say that something is wrong, we are communicating that we have a certain value, not making a judgment based on that value. Compare: I can declare, "Sushi is delicious" at 6:00am, when I have no desire to eat it. But this statement of value would be empty, were I not disposed to experience sushi as delicious.

In summary, I think the Constitution Model can withstand objections and account for the data better than its alternatives. But the model still needs some fleshing out. Moral judgments contain negative feelings, but which ones? Clearly not every bad feeling is a case of moral judgment.

2. The Nature of Emotions in Moral Judgment

To read the literature on sentimentalism, one might think there is a single emotional state called disapproval. But the empirical literature suggests that moral disapproval is felt differently in different contexts. We can speak of a family of disapproval emotions (see Rozin's CAD model).

One important emotion in this family is anger, and its variants, indignation, and outrage. Anger arises most typically for cases in which learn that someone has intentionally harmed or sought to harm another. Another form of disapproval is disgust. This arises when we encounter crimes against nature, such as violations of sexual taboos, even when no one is harmed (necrophilia, for example). Corresponding to these other-directed emotions, we also experience self-directed disapproval when we ourselves misbehave. Guilt arises when we harm other, and shame arises when we violate sexual taboos.

I think all of these emotions originate outside the moral context. Anger is a feeling of the body's preparation to aggress, and that can occur when we are under threat, even if no norm have been violated. Consider the anger mustered by two boxers in a bout. Disgust is a feeling of the body's preparation to expel contaminants, and it can arise when seeing or tasting rotten foods. Even shame and guilt may have non-moral variants. Shame is a kind of unpleasant embarrassment, manifested as a feeling of the body as we try to conceal ourselves from others. Guilt may be a blend of fear and sadness—a feeling in the paradoxical state of flight preparation while also reaching out dolefully to those we have harmed. Sadness characteristically arises when we become separated from those we care about, and this is precisely the risk we incur when we harm someone.

Given that emotions of disapproval can arise in non-moral contexts, one might wonder what distinguishes the moral cases from these others. The answer, I think, has to do with the distinction between self- and other-directed emotions just adduced. If you see me eating rotten food, you will feel disgusted, but if you eat rotten food yourself, you will not feel ashamed, you will feel disgust—an outward emotion for both cases. If you are boxing you might feel aggressive irritation when your sparring partner hits you, but you won't feel guilt when you hit back. The moral domain is distinguished from the non-moral by the pairing of disgust and shame on the one hand and anger and guilt on the other. A judgment about an action qualifies as moral if it issues from a sentiment that disposes you to feel other-directed disapproval if someone performs the action.

In summary, to judge that an action is morally bad is to feel an emotion of disapproval issuing from a disposition to feel other-directed disapproval when others perform that action and self-directed disapproval when I do. That disposition is a basic moral value. Of course there are many actions towards which we have no basic values, such as unfamiliar tax policies, but we can morally assess them by seeing whether they would lead to a basic value violation if instituted. Extensive reasoning is often needed to go from novel cases to basic values.

This explains disapproval, but what about the judgment that something is morally good? Consider the trolley cases, in which the recognition that five people

are in danger motivates one to consider helping. In some cases, helping motivation is negative: we judge that it would be bad not to help, and this judgment consists in a feeling of (anticipatory) guilt at the thought of omission. But there might also be a judgment that it would be good to help. This can be understood in terms of a positively valence emotion, perhaps a kind of anticipatory pride. What distinguishes moral pride from other kinds, is that actions that make us morally proud would make us feel grateful if we were the beneficiary rather than the moral agent. Again, moral emotions are ordinary non-moral emotions that stem from dispositions that have characteristic self- and other-directed manifestations. The disposition to feel pride and gratitude can be regarded toward some action type can be regarded as a moral value of approval.

Notice that I have neglected the affective construct that is most conspicuous in 18th century sentimental theories: sympathy. I said the desire to help is driven by anticipatory pride, not a sympathetic experience of the suffering of others (what we now call empathy). In making this claim, I do not deny that prosocial motivation is sometimes driven empathetically. But, contra Hume and Smith, I deny that empathy is a component of moral judgments, and I think it is only a precursor on some occasions. We feel empathy when there is a single salient victim, especially if the victim is similar to us. But empathy is less prevalent when we judge that it is good to help a *collection* of people, or to join in a political cause (evidence suggests that empathy is not a main contributor to reasoning about justice). Elsewhere I have argued the empathy is not necessary for moral judgment in any way—neither developmentally, epistemically, motivationally, nor normatively. Here I just want to register, more modestly, that even if empathy sometimes compels us to see that some action is good or bad, other emotional states constitute that resulting moral insight.

3. The Source of Emotions in Moral Judgment

I have suggested that basic moral values are sentiments (dispositions to feel emotions), and moral judgments are emotions of disapproval or approval. Reasoning can help us see when basic values are at stake, but basic values themselves do not arise through reasoning. When I recognize that something is a case of calculated killing, rape, or theft, my feelings of disapproval arise immediately without any further inferential steps. Basic values do not arise through reasoning, and they are not acquired that way. Just as we cannot derive an ought from an is, no process of reasoning can entail a value from anything other than another value. Reasoning alone will not suffice.

Where, then, do basic values come from? A popular answer is evolution. I think evolutionary approaches to morality are fundamentally mistaken, but I won't rehearse that case here. For present purposes, I am content to point out that biological evolution does not provide as sufficient explanation of our basic values. There is one simple and decisive reason for this conclusion: basic values differ.

Cross-cultural research suggests that just about every value we cherish is rejected by some other group. We find cultures that practice cannibalism, bloodsports, slavery, and bride conquest through kidnapping and rape. Most

societies have some norms that protect members of the in-group, but what passes as protection varies greatly. Most state-scale societies have class structures in which some individuals are subjugated by others. Most have restrictions on who can own property, who can have sexual autonomy, who can compete on the labor market, and so on. In the contemporary industrialized worlds, we find value differences across East and West—with Eastern nations showing a higher degree of collectivist morality, which emphasizes interdependence and self-sacrifice for others. There are also differences between cultures of honor, like the American South or Sicily, and their northern counterparts. Within the many contemporary societies, we also find deep divergence between liberals and conservatives. Haidt has shown that American liberals and conservatives have different basic values, with conservatives showing more concern for purity, hierarchy, and tradition. Lakoff has argued that American liberals have a nurturant ethics, which focuses on the right to self-expression, and treats norm-violators with understanding, while conservatives have an authoritarian ethics, which focuses on self-reliance, and has little tolerance for violations from norms. Liberals don't see how conservatives can oppose abortion and favor capital punishment, but both can be seen efforts to hold people accountable for norm violation. Likewise, the pro-choice, anti-death penalty stance of liberals is consistent with the attitude that we should protect those who get into "trouble."

To understand the source of our values, it is not enough to point out that there is variation. We must also investigate why these differences arise. Here I think we should follow Nietzsche in interrogating the genealogy of morals. What we find is that each of our values is an historical artifact. American liberalism emerges from events such as the two major depressions in our history, and the civil rights movements. American conservatism issues from frontier libertarianism and the cold war. Among other sources. Genealogical analyses shed light on a wide range of values. Broadening strictures against incest seem to arise with social stratification, to prevent consolidation of wealth (Thornhill). Monogamy norms may have emerged in Christian Europe to reduce family size, increase hairlessness, a line the coffers of the Church (Goody). Male dominant values can be linked to the invention of the plow, which puts farming into the hand of men, making them the primary bread winners (Harris). Tolerance of homosexuality seems to go up, with economic disincentives to rear children. Slavery disappeared with the industrial revolution, when industrialists wanted to take power from the farming industry. Bloodsports are especially common in places imperialistic nations that want to encourage military prowess and instill fear in foreigners. Torture is condoned when there is a perceived enemy within, as in the case of heretics, revolutionaries, intellectuals, or terrorists. These equations are simplistic, of course, and more adequate explanations arise when one goes from generalities to specific details: why did Chinese footbinding come and go? How did American southerners become more violent than northern counterparts? What makes a member of Al Qaeda think it's okay to kill civilians? What is the genealogical link between capitalism and democracy? Social scientists have developed plausible answers to these questions.

For the moral psychologist the main interest of genealogy is the moral past sheds light on the moral present. Trivially, we all believe our values to be right. We

also believe that our basic values derive from some insight into the truth, rather than cultural inculcation. We think our moral opponents are confused or malicious. We bolster our basic values with elaborate arguments, which are no better than the arguments on the other side. Pretty much every reason any of us standardly give for our basic values is a poor one, meaning that someone more clever has already considered the argument and debunked it. Moreover, the arguments we deploy are often acquired long after the acquisition of the values they allegedly support. To this extent Haidt is right about post hocness. There are cases of persuasion through reason, as I suggested. Like those who come to see that a basic value extends to a surprising new case--one thinks of animal rights as an example. But often, I suspect, these cases of moral persuasion turn on something other than reason: the cuteness of animals, the allure of joining a liberal cause, the elegance of consequentialist theories. Without denying that much more discourse turns on rationally decidable question about which of our basic values applies in a given case, it is important to realize that those basic values are not products of personal reasoning, and value change often originates from material pressures that are outside our awareness and control. This is true even in the case of causes that we see as moral progress: transatlantic slavery could not have ended without the industrial revolution; the two major American women's movements followed on the heels of two world wars. Did these changes allow *improvement* or did they just change the context allowing for a new set of power relations, and, perhaps, new systems of oppression? These questions are difficult and bound up with historical and empirical facts (are we better off now? Did slavery hinge on false scientific theories?). Optimists might say moral change is progressive because more people are free and healthy now than in the past. But that is an expression of value. Since many of us share such sentiments, we can celebrate our success, but we should not be stupefied when other cultures view our achievements as retrograde.

4. Implications of Emotions in Moral Judgment

I've claimed that moral values are based on sentiments that dispose us to emotions of approval and disapproval, and that these sentiments are shaped by historical processes. I call this Constructive Sentimentalism. In this final section, I want to consider some implication, with special emphasis on legal and political domains. I will be sketchy, because there are others in this conversation who are better placed to see what follows. Let me make two main suggestions, corresponding to the sentimentalism and the constructionism respectively.

First, if morality as a basis in emotions, we should expect moral judgments to be very susceptible to the influence of emotions that are not necessarily relevant to a case at hand. In the context of courtroom deliberation, this means that emotional elicitation through the presentation of disturbing crime photographs, evocative testimony, or emotional expressiveness on the part of plaintiffs or victims is likely to influence jurors' assessments. Empirical evidence supports this conclusion. We should also expect legislators to have such vulnerabilities, and politicians can exploit the emotions of their constituencies. Mustering fear of an enemy within can increase tolerance for judicial torture.

We can guard against some forms of biasing influence, but it is often difficult, in principle, to decide what counts as bias and what counts as evidence. Consider a photo of a murder victim. Emotionally evocative, yes, but the disgust and outrage we experience may help us achieve moral clarity. If badness judgments are composed of such emotions, then this can be viewed as a direct apprehension of moral badness of the crime. On the other hand, we might think that badness should track actions abstractly characterized (taking a life), not characterized in terms of superficial details (the blood on the victim). But the question of which details matter may be impossible to resolve. By comparison, think of framing effects. Different ways to present the same facts can elicit different emotions, even when there is no way to settle which framing is more faithful to reality.

Second, some moral disputes are not rationally resolvable, because participants have different basic values. In the American political context, this is most evident with respect to the liberal/conservative divide. Liberals and conservatives are bewildered by each other, as if we reside in different moral universes. Curiously, public discourse often includes arguments, which are presented as if they provide reasons for adopting liberal or conservative policies. These arguments may influence some undecided voters, but they are largely inert, when appealing to the opposition. Liberals stay liberal and conservatives stay conservative. Personal and economic crises, changes in living situation, material resources, aging, world events, and other factors can shift values, as can powerful rhetoric, but the arguments offered probably don't exert rational influence. Why, then, do we deploy such arguments? I think they are a form of auto-persuasion. For each of us, the veneer or reason boosts confidence, and for the eloquent leader, a good argument can rally the base. Of course, arguments about arcane bits of policy may do some real work, but only relative to basic values, so these arguments will be less convincing to some than to others.

We can find similar issues of moral division with small group deliberations in the courtroom. A jury will find itself made up of liberals and conservatives, Jews and gentiles, blacks and white, men and women. Each of these variables, along with idiosyncrasies of biography has put us into contact with different sources of moral inculcation. Views about guilt and punishment are sure to vary. For a vivid example, recall the O.J. Simpson trial, in which whites were more inclined than blacks to reject the allegation that the police set Simpson up. Or, to take a literary case, think of the white jurors in *To Kill a Mockingbird*. In such cases, there is a truth about what happened, but value differences influence the interpretation of the evidence and beliefs what should be done when it comes to punishment. Given different inculcated standpoints and incomplete knowledge, there may be no way to settle some disputes rationally.

This may look like a cynical view. It would be nice to think that there is a single true morality, which could be invoked to settle moral disagreements in sociopolitical contexts. Much work in normative ethics aims at such universals, and this is not the place to assess those efforts. I'll settle for the weaker claim that these noble projects have limited bearing on actual moral disputes. Academic insights are usually moot in the courtroom, and the voting public regards our activities with suspicion and contempt. On the optimistic side, one can hope that balanced juries

and a balance of political power can insure that different opinions, equally biases, can weigh in.

These are the kinds of solutions we've devised, but we must remember their limitations. Democracies wage war and even elect dictators. Courtrooms draw juries from specific districts, and women and minorities often capitulate to white men. In any case, it's not clear that deliberation and democracy are adequate answers to sentimental diversity. After all, in the political case, one side wins and the legislates for a diverse populace. An alternative solution would go radically local, scaling legal jurisdictions down to a size the was more commensurate with moral communities. But this strategy brings us into the libertarian situation, in which local groups take care of their own, and abuse outliers or neglect distant others in need. Some evidence suggests that the increase in group size and global economic ties has decreased violence dramatically (Pinker). Small scale societies killed at rates the exceed Russian and German mortality during the second world war (Wrangham).

So constructive sentimentalism leaves us with a practical dilemma. Once we admit that there is moral variation that cannot be feasible settled by universal reason, we must decide between living in a world where nations and global alliances impose values on populations that are radically heterogeneous, or return to a kind of moral feudalism. Neither is especially attractive, and the choice looks suspiciously like the irresolvable and value-laden disagreement between liberals and conservatives about centralism and state's rights.

Here there is room for an interesting project. I think we need to think more seriously about what a relativist normative ethics would look like. And we need to think about relativist political theories and legal institutions. To some extent, that's what we have already, but I don't think the founders or their followers were convinced of relativism, so we are operating within social structures that are shaped under the weight of diversity, but not necessarily ideally suited for that end. We might think we've hit on something optimal (or nearly so), but we must think carefully about who the "we" is.

Let me end with two rosier observations. If values are inculcated and unfettered from cool reason, then we may not find a universal basis for morality. On the other hand, we can think of morality as self-expressive in the same way we value other cultural institutions, and we can embrace our morals happily, while recognizing that others' values are not based on confusion or iniquity. Alternatively, we can recognize that our values are historical accidents, and pursue programs of moral reform. Reason may not serve as a rudder here, but other non-moral goals can guide us. We can adopt values that help us live more successful, healthy, and comfortable lives. Politically, this may suggest an agenda of moral reform rather than poll matching. In terms of legal institutions, we might try to bracket intuitions of justice and focus instead on what non-moral ends we want those institutions to serve. Of course, reform carries risk. Liberation movements are counterbalanced by catastrophic social experiments. So, optimism must always be tempered by caution.