

Thirteen

Morality and Doing the Right Thing, Once You've Figured Out What That Is

The two previous chapters examined the thoroughly unique contexts for some human behaviors that are on a continuum with behaviors in other species. Like some other species, we make automatic Us/Them dichotomies and favor the former—though only humans rationalize that tendency with ideology. Like many other species, we are implicitly hierarchical—though only humans view the gap between haves and have-nots as a divine plan.

This chapter considers another domain rife with human uniqueness, namely morality. For us, morality is not only belief in norms of appropriate behavior but also the belief that they should be shared and transmitted culturally.

Work in the field is dominated by a familiar sort of question. When we make a decision regarding morality, is it mostly the outcome of moral reasoning or of moral intuition? Do we mostly think or feel our way to deciding what is right?

This raises a related question. Is human morality as new as the cultural institutions we've hatched in recent millennia, or are its rudiments a far older primate legacy?

This raises more questions. What's more impressive, consistencies and universalities of human moral behavior or variability and its correlation with cultural and ecological factors?

Finally, there will be unapologetically prescriptive questions. When it comes to moral decision making, when is it "better" to rely on intuition, when on reasoning? And when we resist temptation, is it mostly an act of will or of grace?

People have confronted these issues since students attended intro philosophy in togas. Naturally, these questions are informed by science.

THE PRIMACY OF REASONING IN MORAL DECISION MAKING

One single fact perfectly demonstrates moral decision making being based on cognition and reasoning. Have you ever picked up a law textbook? They're humongous.

Every society has rules about moral and ethical behavior that are reasoned and call upon logical operations. Applying the rules requires reconstructing scenarios, understanding proximal and distal causes of events, and assessing magnitudes and probabilities of consequences of actions. Assessing individual behavior requires perspective taking, Theory of Mind, and distinguishing between outcome and intent. Moreover, in many cultures rule implementation is typically entrusted to people (e.g., lawyers, clergy) who have undergone long training.

Harking back to chapter 7, the primacy of reasoning in moral decision making is anchored in child development. The Kohlbergian emergence of increasingly complex stages of moral development is built on the Piagetian emergence of increasingly complex logical operations. They are similar, neurobiologically. Logical and moral reasoning about the correctness of an economic or ethical decision, respectively, both activate the (cognitive) dlPFC. People with obsessive-compulsive disorder get mired in both everyday decision making and moral decision making, and their dlPFCs go wild with activity for both.¹

Similarly, there's activation of the temporoparietal junction (TPJ) during Theory of Mind tasks, whether they are perceptual (e.g., visualizing a complex scene from another viewer's perspective), amoral (e.g., keeping straight who's in love with whom in *A Midsummer Night's Dream*), or moral/social (e.g., inferring the ethical motivation behind a person's act). Moreover, the more the TPJ activation, the more people take intent into account when making moral judgments, particularly when there was intent to harm but no actual harm done. Most important, inhibit the TPJ with transcranial magnetic stimulation, and subjects become less concerned about intent.²

The cognitive processes we bring to moral reasoning aren't perfect, in that there are fault lines of vulnerability, imbalances, and asymmetries.³ For example, doing harm is worse than allowing it—for equivalent outcomes we typically judge commission more harshly than omission and must activate the dlPFC more to judge them as equal. This makes sense—when we do one thing, there are innumerable other things we didn't do; no wonder the former is psychologically weightier. As another cognitive skew, as discussed in chapter 10, we're better at detecting violations of social contracts that have malevolent rather than benevolent consequences (e.g., giving less versus more than promised). We also search harder for causality (and come up with more false attributions) for malevolent than for benevolent events.

This was shown in one study. First scenario: A worker proposes a plan to the boss, saying, "If we do this, there'll be big profits, and we'll harm the environment in the process." The boss answers: "I don't care about the environment. Just do it." Second scenario: Same setup, but this time there'll be big profits and *benefits* to the environment. Boss: "I don't care about the environment. Just do it." In the first scenario 85 percent of subjects stated that the boss harmed the environment *in order to* increase profits; however, in the second scenario only 23 percent said that the boss helped the environment *in order to* increase profits.⁴

Okay, we're not perfect reasoning machines. But that's our goal, and numerous moral philosophers emphasize the preeminence of reasoning, where emotion and intuition, if they happen to show up, just soil the carpet. Such philosophers range from Kant, with his search for a mathematics of morality, to Princeton philosopher Peter Singer, who kvetches that if things like sex and bodily functions are pertinent to philosophizing, time to hang up his spurs: "It would be best to forget all about our particular moral judgments." Morality is anchored in reason.⁵

YEAH, SURE IT IS: SOCIAL INTUITIONISM

Except there's a problem with this conclusion—people often haven't a clue *why* they've made some judgment, yet they fervently believe it's correct.

This is straight out of chapter 11's rapid implicit assessments of Us versus Them and our post-hoc rational justifications for visceral prejudice. Scientists studying moral philosophy increasingly emphasize moral decision making as implicit, intuitive, and anchored in emotion.

The king of this “social intuitionist” school is Jonathan Haidt, whom we've encountered previously.⁶ Haidt views moral decisions as primarily based on intuition and believes reasoning is what we then use to convince everyone, including ourselves, that we're making sense. In an apt phrase of Haidt's, “moral thinking is for social doing,” and sociality always has an emotional component.

The evidence for the social intuitionist school is plentiful:

When contemplating moral decisions, we don't just activate the eggheady dlPFC.⁷ There's also activation of the usual emotional cast—amygdala, vmPFC and the related orbitofrontal cortex, insular cortex, anterior cingulate. Different types of moral transgressions preferentially activate different subsets of these regions. For example, moral quandaries eliciting pity preferentially activate the insula; those eliciting indignation activate the orbitofrontal cortex. Quandaries generating intense conflict preferentially activate the anterior cingulate. Finally, for acts assessed as equally morally wrong, those involving nonsexual transgression (e.g., stealing from a sibling) activate the amygdala, whereas those involving sexual transgressions (e.g., sex with a sibling) also activate the insula.*

Moreover, when such activation is strong enough, we also activate the sympathetic nervous system and feel arousal—and we know how those peripheral effects feedback and influence behavior. When we confront a moral choice, the dlPFC doesn't adjudicate in contemplative silence. The waters roil below.

The pattern of activation in these regions predicts moral decisions better than does the dlPFC's profile. And this matches behavior—people punish to the extent that they *feel* angered by someone acting unethically.⁸

People tend toward instantaneous moral reactions; moreover, when subjects shift from judging nonmoral elements of acts to moral ones, they make assessments faster, the antithesis of moral decision making being about grinding cognition. Most strikingly, when facing a moral quandary, activation in the amygdala, vmPFC, and insula typically *precedes* dlPFC activation.⁹

Damage to these intuitionist brain regions makes moral judgments more pragmatic, even coldhearted. Recall from chapter 10 how people with damage to the (emotional) vmPFC readily advocate sacrificing one relative to save five strangers, something control subjects never do.

Most telling is when we have strong moral opinions but can't tell why, something Haidt calls “moral dumbfounding”—followed by clunky post-hoc rationalizing.¹⁰ Moreover, such moral decisions can differ markedly in different affective or visceral circumstances, generating very different rationalizations. Recall from the last chapter how people become more conservative in their social judgments when they're smelling a foul odor or sitting at a dirty desk. And then there's that doozy of a finding—knowing a judge's opinions about Plato, Nietzsche, Rawls, and any other philosopher whose name I just looked up gives you less predictive power about her judicial decisions than knowing if she's hungry.

The social intuitionist roots of morality are bolstered further by evidence of moral judgment in two classes of individuals with limited capacities for moral reasoning.

AGAIN WITH BABIES AND ANIMALS

Much as infants demonstrate the rudiments of hierarchical and Us/Them thinking, they possess building blocks of moral reasoning as well. For starters, infants have the bias concerning commission versus omission. In one clever study, six-month-olds watched a scene containing two of the same objects, one blue and one red; repeatedly, the scene would show a person picking the blue object. Then, one time, the red one is picked. The kid becomes interested, looks more, breathes faster, showing that this seems discrepant. Now, the scene shows two of the same objects, one blue, one a different color. In each repetition of the scene, a person picks the one that is not blue (its color changes with each repetition). Suddenly, the blue one is picked. The kid isn't particularly interested. "He always picks the blue one" is easier to comprehend than "He never picks the blue one." Commission is weightier.¹¹

Infants and toddlers also have hints of a sense of justice, as shown by Kiley Hamlin of the University of British Columbia, and Paul Bloom and Karen Wynn of Yale. Six- to twelve-month-olds watch a circle moving up a hill. A nice triangle helps to push it. A mean square blocks it. Afterward the infants can reach for a triangle or a square. They choose the triangle.* Do infants prefer nice beings, or shun mean ones? Both. Nice triangles were preferred over neutral shapes, which were preferred over mean squares.

Such infants advocate punishing bad acts. A kid watches puppets, one good, one bad (sharing versus not). The child is then presented with the puppets, each sitting on a pile of sweets. Who should lose a sweet? The bad puppet. Who should gain one? The good puppet.

Remarkably, toddlers even assess secondary punishment. The good and bad puppets then interact with two additional puppets, who can be nice or bad. And whom did kids prefer of those second-layer puppets? Those who were nice to nice puppets and those who punished mean ones.

Other primates also show the beginnings of moral judgments. Things started with a superb 2003 paper by Frans de Waal and Sarah Brosnan.¹² Capuchin monkeys were trained in a task: A human gives them a mildly interesting small object—a pebble. The human then extends her hand palm up, a capuchin

begging gesture. If the monkey puts the pebble in her hand, there's a food reward. In other words, the animals learned how to buy food.

Now there are two capuchins, side by side. Each gets a pebble. Each gives it to the human. Each gets a grape, very rewarding.

Now change things. Both monkeys pay their pebble. Monkey 1 gets a grape. But monkey 2 gets some cucumber, which blows compared with grapes—capuchins prefer grapes to cucumber 90 percent of the time. Monkey 2 was shortchanged.

And monkey 2 would then typically fling the cucumber at the human or bash around in frustration. Most consistently, they wouldn't give the pebble the next time. As the *Nature* paper was entitled, "Monkeys reject unequal pay."

This response has since been demonstrated in various macaque monkey species, crows, ravens, and dogs (where the dog's "work" would be shaking her paw).^{[*13](#)}

Subsequent work by Brosnan, de Waal, and others fleshed out this phenomenon further:^{[14](#)}

- One criticism of the original study was that maybe capuchins refused to work for cucumbers because grapes were visible, regardless of whether the other guy was getting paid in grapes. But no—the phenomenon required unfair payment.
- Both animals are getting grapes, then one gets switched to cucumber. What's key—that the other guy is still getting grapes, or that I no longer am? The former—if doing the study with a single monkey, switching from grapes to cucumbers would not evoke refusal. Nor would it if both monkeys got cucumbers.
- Across the various species, males were more likely than females to reject "lower pay"; dominant animals were more likely than subordinates to reject.
- It's about the work—give one monkey a free grape, the other free cucumber, and the latter doesn't get pissed.
- The closer in proximity the two animals are, the more likely the one getting cucumber is to go on strike.
- Finally, rejection of unfair pay isn't seen in species that are solitary (e.g., orangutans) or have minimal social cooperation (e.g., owl monkeys).

Okay, very impressive—other social species show hints of a sense of justice, reacting negatively to unequal reward. But this is worlds away from juries awarding money to plaintiffs harmed by employers. Instead it's self-interest—"This isn't fair; I'm getting screwed."

How about evidence of a sense of fairness in the treatment of another individual? Two studies have examined this in a chimp version of the Ultimatum Game. Recall the human version—in repeated rounds, player 1 in a pair decides how money is divided between the two of them. Player 2 is powerless in the decision making but, if unhappy with the split, can refuse, and no one gets any money. In other words, player 2 can forgo immediate reward to punish selfish player 1. As we saw in chapter 10, Player 2s tend to accept 60:40 splits.

In the chimp version, chimp 1, the proposer, has two tokens. One indicates that each chimp gets two grapes. The other indicates that the proposer gets three grapes, the partner only one. The proposer chooses a token and passes it to chimp 2, who then decides whether to pass the token to the human grape dispenser. In other words, if chimp 2 thinks chimp 1 is being unfair, no one gets grapes.

In one such study, Michael Tomasello (a frequent critic of de Waal—stay tuned) at the Max Planck Institutes in Germany, found no evidence of chimp fairness—the proposer always chose, and the partner always accepted unfair splits.¹⁵ De Waal and Brosnan did the study in more ethologically valid conditions and reported something different: proposer chimps tended toward equitable splits, but if they could give the token directly to the human (robbing chimp 2 of veto power), they'd favor unfair splits. So chimps will opt for fairer splits—but only when there is a downside to being unfair.

Sometimes other primates are fair when it's at no cost to themselves. Back to capuchin monkeys. Monkey 1 chooses whether both he and the other guy get marshmallows or it's a marshmallow for him and yucky celery for the other guy. Monkeys tended to choose marshmallows for the other guy.* Similar "other-regarding preference" was shown with marmoset monkeys, where the first individual got nothing and merely chose whether the other guy got a cricket to eat (of note, a number of studies have failed to find other-regarding preference in chimps).¹⁶

Really interesting evidence for a nonhuman sense of justice comes in a small side study in a Brosnan/de Waal paper. Back to the two monkeys getting cucumbers for work. Suddenly one guy gets shifted to grapes. As we saw, the

one still getting the cucumber refuses to work. Fascinatingly, the grape mogul often refuses as well.

What is this? Solidarity? “I’m no strike-breaking scab”? Self-interest, but with an atypically long view about the possible consequences of the cucumber victim’s resentment? Scratch an altruistic capuchin and a hypocritical one bleeds? In other words, all the questions raised by human altruism.

Given the relatively limited reasoning capacities of monkeys, these findings support the importance of social intuitionism. De Waal perceives even deeper implications—the roots of human morality are older than our cultural institutions, than our laws and sermons. Rather than human morality being spiritually transcendent (enter deities, stage right), it transcends our species boundaries.^{[17](#)}

MR. SPOCK AND JOSEPH STALIN

Many moral philosophers believe not only that moral judgment is built on reasoning but also that it *should* be. This is obvious to fans of Mr. Spock, since the emotional component of moral intuitionism just introduces sentimentality, self-interest, and parochial biases. But one remarkable finding counters this.

Relatives are special. Chapter 10 attests to that. Any social organism would tell you so. Joseph Stalin thought so concerning Pavlik Morozov ratting out his father. As do most American courts, where there is either de facto or de jure resistance to making someone testify against their own parent or child. Relatives are special. But not to people lacking social intuitionism. As noted, people with vmPFC damage make extraordinarily practical, unemotional moral decisions. And in the process they do something that everyone, from clonal yeast to Uncle Joe to the Texas Rules of Criminal Evidence considers morally suspect: they advocate harming kin as readily as strangers in an “Is it okay to sacrifice one person to save five?” scenario.¹⁸

Emotion and social intuition are not some primordial ooze that gums up that human specialty of moral reasoning. Instead, they anchor some of the few moral judgments that most humans agree upon.

CONTEXT

So social intuitions can have large, useful roles in moral decision making. Should we now debate whether reasoning or intuition is more important? This is silly, not least of all because there is considerable overlap between the two. Consider, for example, protesters shutting down a capital to highlight income inequity. This could be framed as the Kohlbergian reasoning of people in a postconventional stage. But it could also be framed à la Haidt in a social intuitionist way—these are people who resonate more with moral intuitions about fairness than with respect for authority.

More interesting than squabbling about the relative importance of reasoning and intuition are two related questions: What circumstances bias toward emphasizing one over the other? Can the differing emphases produce different decisions?

As we've seen, then—graduate student Josh Greene and colleagues helped jump-start “neuroethics” by exploring these questions using the poster child of “Do the ends justify the means?” philosophizing, namely the runaway trolley problem. A trolley's brake has failed, and it is hurtling down the tracks and will hit and kill five people. Is it okay to do something that saves the five but kills someone else in the process?

People have pondered this since Aristotle took his first trolley ride;* Greene et al. added neuroscience. Subjects were neuroimaged while pondering trolley ethics. Crucially, they considered two scenarios. Scenario 1: Here comes the trolley; five people are goners. Would you pull a lever that diverts the trolley onto a different track, where it will hit and kill someone (the original scenario)? Scenario 2: Same circumstance. Would you *push* the person onto the tracks to stop the trolley?¹⁹

By now I bet readers can predict which brain region(s) activates in each circumstance. Contemplate pulling the lever, and dlPFC activity predominates, the detached, cerebral profile of moral reasoning. Contemplate consigning the person to death by pushing them, and it's vmPFC (and amygdala), the visceral profile of moral intuition.

Would you pull the lever? Consistently, 60 to 70 percent of people, with their dlPFCs churning away, say yes to this utilitarian solution—kill one to save five. Would you push the person with your own hands? Only 30 percent are willing; the more the vmPFC and/or amygdaloid activation, the more likely they are to refuse.* This is hugely important—a relatively minor variable determines whether people emphasize moral reasoning or intuition, and they engage different brain circuits in the process, producing radically different decisions. Greene has explored this further.

Are people resistant to the utilitarian trade-off of killing one to save five in the pushing scenario because of the visceral reality of actually touching the person whom they have consigned to death? Greene's work suggests not—if instead of pushing with your hands, you push with a pole, people are still resistant. There's something about the personal force involved that fuels the resistance.

Are people willing in the lever scenario because the victim is at a distance, rather than right in front of them? Probably not—people are just as willing if the lever is right next to the person who will die.

Greene suggests that intuitions about intentionality are key. In the lever scenario, the five people are saved because the trolley has been diverted to another track; the killing of the individual is a side effect and the five would still have been saved if that person hadn't been standing on the tracks. In contrast, in the pushing scenario the five are saved *because* the person is killed, and the intentionality feels intuitively wrong. As evidence, Greene would give subjects another scenario: Here comes the trolley, and you are rushing to throw a switch that will halt it. Is it okay to do this if you know that in the process of lunging for the switch, you must push a person out of the way, who falls to the ground and dies? About 80 percent of people say yes. Same pushing the person, same proximity, but done unintentionally, as a side effect. The person wasn't killed as a *means* to save the five. Which seems much more okay.

Now a complication. In the "loop" scenario, you pull a lever that diverts the trolley to another track. But—oh no!—it's just a loop; it merges back on to the original track. The trolley will still kill the five people—except that there's a person on the side loop who will be killed, stopping the trolley. This is as intentional a scenario as is pushing with your hands—diverting to another track isn't enough; the person has to be killed. By all logic only about 30 percent of people should sign on, but instead it's in the 60 to 70 percent range.

Greene concludes (from this and additional scenarios resembling the loop) that the intuitionist universe is very local. Killing someone intentionally as a means to save five feels intuitively wrong, but the intuition is strongest when the killing would occur right here, right now; doing it in more complicated sequences of intentionality doesn't feel as bad. This is not because of a cognitive limit—it's not that subjects don't realize the necessity of killing the person in the loop scenario. It just doesn't *feel* the same. In other words, intuitions discount heavily over space and time. Exactly the myopia about cause and effect you'd expect from a brain system that operates rapidly and automatically. This is the same sort of myopia that makes sins of commission feel worse than those of omission.

Thus these studies suggest that when a sacrifice of one requires active, intentional, and local actions, more intuitive brain circuitry is engaged, and ends don't justify means. And in circumstances where either the harm is unintentional or the intentionality plays out at a psychological distance, different neural circuitry predominates, producing an opposite conclusion about the morality of ends and means.

These trolleyology studies raise a larger point, which is that moral decision making can be wildly context dependent.²⁰ Often the key thing that a change in context does is alter the locality of one's intuitionist morals, as summarized by Dan Ariely of Duke University in his wonderful book *Predictably Irrational*. Leave money around a common work area and no one takes it; it's not okay to steal money. Leave some cans of Coke and they're all taken; the one-step distance from the money involved blunts the intuitions about the wrongness of stealing, making it easier to start rationalizing (e.g., someone must have left them out for the taking).

The effects of proximity on moral intuitionism are shown in a thought experiment by Peter Singer.²¹ You're walking by a river in your hometown. You see that a child has fallen in. Most people feel morally obliged to jump in and save the child, even if the water destroys their \$500 suit. Alternatively, a friend in Somalia calls and tells you about a poor child there who will die without \$500 worth of medical care. Can you send money? Typically not. The locality and moral discounting over distance is obvious—the child in danger in your hometown is far more of an Us than is this dying child far away. And this is an intuitive rather than cognitive core—if you were walking along in Somalia and saw a child fall into a river, you'd be more likely to jump in and sacrifice the suit than to send \$500 to that friend making the phone call. Someone being right

there, in the flesh, in front of our eyes is a strong implicit prime that they are an Us.

Moral context dependency can also revolve around language, as noted in chapter 3.²² Recall, for example, people using different rules about the morality of cooperation if you call the same economic game the “Wall Street game” or the “community game.” Framing an experimental drug as having a “5 percent mortality rate” versus a “95 percent survival rate” produces different decisions about the ethics of using it.

Framing also taps into the themes of people having multiple identities, belonging to multiple Us groups and hierarchies. This was shown in a hugely interesting 2014 *Nature* paper by Alain Cohn and colleagues at the University of Zurich.²³ Subjects, who worked for an (unnamed) international bank, played a coin-toss game with financial rewards for guessing outcomes correctly. Crucially, the game’s design made it possible for subjects to cheat at various points (and for the investigators to detect the cheating).

In one version subjects first completed a questionnaire filled with mundane questions about their everyday lives (e.g., “How many hours of television do you watch each week?”). This produced a low, baseline level of cheating.

Then, in the experimental version, the questionnaire was about their bank job. Questions like these primed the subjects to implicitly think more about banking (e.g., they became more likely in a word task to complete “__oker” with “broker” than with “smoker”).

So subjects were thinking about their banking identity. And when they did, rates of cheating rose 20 percent. Priming people in other professions (e.g., manufacturing) to think about their jobs, or about the banking world, didn’t increase cheating. These bankers carried in their heads two different sets of ethical rules concerning cheating (banking and nonbanking), and unconscious cuing brought one or the other to the forefront.* Know thyself. Especially in differing contexts.

“But This Circumstance Is Different”

The context dependency of morality is crucial in an additional realm.

It is a nightmare of a person who, with remorseless sociopathy, believes it is okay to steal, kill, rape, and plunder. But far more of humanity’s worst behaviors are due to a different kind of person, namely most of the rest of us, who will say

that of course it is wrong to do X . . . but here is why these special circumstances make me an exception right now.

We use different brain circuits when contemplating our own moral failings (heavy activation of the vmPFC) versus those of others (more of the insula and dlPFC).²⁴ And we consistently make different judgments, being more likely to exempt ourselves than others from moral condemnation. Why? Part of it is simply self-serving; sometimes a hypocrite bleeds because you've scratched a hypocrite. The difference may also reflect different emotions being involved when we analyze our own actions versus those of others. Considering the moral failings of the latter may evoke anger and indignation, while their moral triumphs prompt emulation and inspiration. In contrast, considering our own moral failings calls forth shame and guilt, while our triumphs elicit pride.

The affective aspects of going easy on ourselves are shown when stress makes us more this way.²⁵ When experimentally stressed, subjects make more egoistic, rationalizing judgments regarding emotional moral dilemmas and are less likely to make utilitarian judgments—but only when the latter involve a personal moral issue. Moreover, the bigger the glucocorticoid response to the stressor, the more this is the case.

Going easy on ourselves also reflects a key cognitive fact: we judge ourselves by our internal motives and everyone else by their external actions.²⁶ And thus, in considering our own misdeeds, we have more access to mitigating, situational information. This is straight out of Us/Them—when Them's do something wrong, it's because they're simply rotten; when Us-es do it, it's because of an extenuating circumstance, and “Me” is the most focal Us there is, coming with the most insight into internal state. Thus, on this cognitive level there is no inconsistency or hypocrisy, and we might readily perceive a wrong to be mitigated by internal motives in the case of anyone's misdeeds. It's just easier to know those motives when we are the perpetrator.

The adverse consequences of this are wide and deep. Moreover, the pull toward judging yourself less harshly than others easily resists the rationality of deterrence. As Ariely writes in his book, “Overall cheating is not limited by risk; it is limited by our ability to rationalize the cheating to ourselves.”

Cultural Context

So people make different moral judgments about the same circumstance depending on whether it's about them or someone else, which of their identities has been primed, the language used, how many steps the intentionality is removed, and even the levels of their stress hormones, the fullness of their stomach, or the smelliness of their environment. After chapter 9 it is no surprise that moral decision making can also vary dramatically by culture. One culture's sacred cow is another's meal, and the discrepancy can be agonizing.

When thinking about cross-cultural differences in morality, key issues are what universals of moral judgment exist and whether the universals or the differences are more interesting and important.

Chapter 9 noted some moral stances that are virtually universal, whether *de facto* or *de jure*. These include condemnation of at least some forms of murder and of theft. Oh, and of some form of sexual practice.

More broadly, there is the near universal of the Golden Rule (with cultures differing as to whether it is framed as "Do only things you'd want done to you" or "Don't do things you wouldn't want done to you"). Amid the power of its simplicity, the Golden Rule does not incorporate people differing as to what they would/wouldn't want done to them; we have entered complicated terrain when we can make sense of an interchange where a masochist says, "Beat me," and the sadist sadistically answers, "No."

This criticism is overcome with the use of a more generalized, common currency of reciprocity, where we are enjoined to give concern and legitimacy to the needs and desires of people in circumstances where we would want the same done for us.

Cross-cultural universals of morality arise from shared categories of rules of moral behavior. The anthropologist Richard Shweder has proposed that all cultures recognize rules of morality pertinent to autonomy, community, and divinity. As we saw in the last chapter, Jonathan Haidt breaks this continuum into his foundations of morality that humans have strong intuitions about. These are issues related to harm, fairness and reciprocity (both of which Shweder would call autonomy), in-group loyalty and respect for authority (both of which Shweder would call community), and issues of purity and sanctity (i.e., Shweder's realm of divinity).^{[*27](#)}

The existence of universals of morality raises the issue of whether that means that they should trump more local, provincial moral rules. Between the moral absolutists on one side and the relativists on the other, people like the historian of science Michael Shermer argue reasonably for provisional morality—if a

moral stance is widespread across cultures, start off by giving the benefit of the doubt to its importance, but watch your wallet.²⁸

It's certainly interesting that, for example, all cultures designate certain things as sacred; but it is far more so to look at the variability in what is considered sacred, how worked up people get when such sanctity is violated,^{*,} and what is done to keep such violations from reoccurring. I'll touch on this huge topic with three subjects—cross-cultural differences concerning the morality of cooperation and competition, affronts to honor, and the reliance on shame versus guilt.

COOPERATION AND COMPETITION

Some of the most dramatic cross-cultural variability in moral judgment concerns cooperation and competition. This was shown to an extraordinary extent in a 2008 *Science* paper by a team of British and Swiss economists.

Subjects played a “public good” economic game where players begin with a certain number of tokens and then decide, in each of a series of rounds, how many to contribute to a shared pool; the pool is then multiplied and shared evenly among all the players. The alternative to contributing is for subjects to keep the tokens for themselves. Thus, the worst payoff for an individual player would be if they contributed all their tokens to the pool, while no other player contributed any; the best would be if the individual contributed no tokens and everyone else contributed everything. As a feature of the design, subjects could “pay” to punish other players for the size of their contribution. Subjects were from around the world.

First finding: Across all cultures, people were more prosocial than sheer economic rationality would predict. If everyone played in the most brutally asocial, realpolitik manner, no one would contribute to the pool. Instead subjects from all cultures consistently contributed. Perhaps as an explanation, subject from all cultures punished people who made lowball contributions, and to roughly equal extents.

Where the startling difference came was with a behavior that I'd never even seen before in the behavioral economics literature, something called “antisocial punishment.” Free-riding punishment is when you punish another player for contributing less than you (i.e., being selfish). Antisocial punishment is when you punish another player for contributing *more* than you (i.e., being generous).

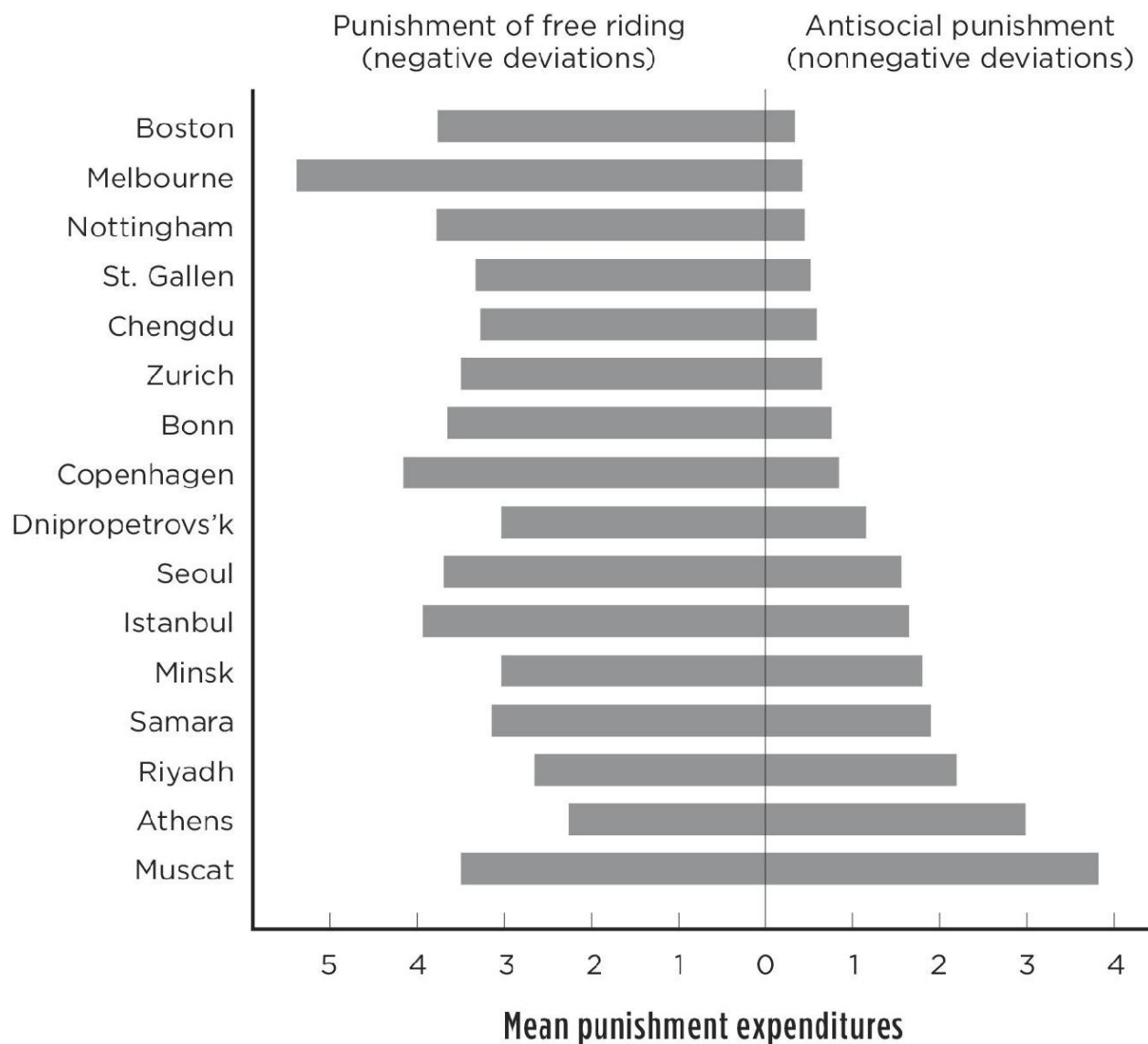
What is that about? Interpretation: This hostility toward someone being overly generous is because they're going to up the ante, and soon everyone (i.e.,

me) will be expected to be generous. Kill 'em, spoiling things for everyone. It's a phenomenon where you punish someone for being nice, because what if that sort of crazy deviance becomes the norm and you feel pressure to be nice back?

At one extreme were subjects from countries (the United States and Australia) where this weird antisocial punishment was nearly nonexistent. And at the mind-boggling other extreme were subjects from Oman and Greece, who were willing to spend *more* to punish generosity than to punish selfishness. And this was not a comparison of, say, theologians in Boston with Omani pirates. Subjects were all urban university students.

So what's different among these cities? The authors found a key correlation—the lower the social capital in a country, the higher the rates of antisocial punishment. In other words, when do people's moral systems include the idea that being generous deserves punishment? When they live in a society where people don't trust one another and feel as if they have no efficacy.

Fascinating work has also been done specifically on people in non-Western cultures, as reported in a pair of studies by Joseph Henrich, of the University of British Columbia, and colleagues.²⁹ Subjects were in the thousands and came from twenty-five different “small-scale” cultures from around the world—they were nomadic pastoralists, hunter-gatherers, sedentary forager/horticulturalists, and subsistence farmers/wage earners. There were two control groups, namely urbanites from Missouri and Accra, Ghana. As a particularly thorough feature of the study, subjects played three economic games: (a) The Dictator Game, where the subject simply decides how money is split between them and another player. This measures a pure sense of fairness, independent of consequence. (b) The Ultimatum Game, where you can pay to punish someone treating you unfairly (i.e., self-interested second-party punishment). (c) A third-party punishment scenario, where you can pay to punish someone treating a third party unfairly (i.e., altruistic punishment).



B. Herrmann et al., "Antisocial Punishment Across Societies," *Sci* 319 (2008): 1362.

Visit bit.ly/2neVZaA for a larger version of this graph.

The authors identified three fascinating variables that predicted patterns of play:

Market integration: How much do people in a culture interact economically, with trade items? The authors operationalized this as the percentage of people's calories that came from purchases in market interactions, and it ranged from 0 percent for the hunter-gathering Hadza of Tanzania to nearly 90 percent for sedentary fishing cultures. And across the cultures a greater degree of market integration strongly predicted people making fairer offers in all three games and being willing to pay for both self-interested second-party and altruistic third-

party punishment of creeps. For example, the Hadza, at one extreme, kept an average of 73 percent of the spoils for themselves in the Dictator Game, while the sedentary fishing Sanquianga of Colombia, along with people in the United States and Accra, approached dictating a 50:50 split. Market integration predicts more willingness to punish selfishness and, no surprise, less selfishness.

Community size: The bigger the community, the more the incidence of second- and third-party punishment of cheapskates. Hadza, for example, in their tiny bands of fifty or fewer, would pretty much accept any offer above zero in the Ultimatum Game—there was no punishment. At the other extreme, in communities of five thousand or more (sedentary agriculturalists and aquaculturalists, plus the Ghanaian and American urbanites), offers that weren't in the ballpark of 50:50 were typically rejected and/or punished.

Religion: What percentage of the population belonged to a worldwide religion (i.e., Christianity or Islam)? This ranged from none of the Hadza to 60 to 100 percent for all the other groups. The greater the incidence of belonging to a Western religion, the more third-party punishment (i.e., willingness to pay to punish person A for being unfair to person B).

What to make of these findings?

First the religion angle. This was a finding not about religiosity generally but about religiosity within a worldwide religion, and not about generosity or fairness but about altruistic third-party punishment. What is it about worldwide religions? As we saw in chapter 9, it is only when groups get large enough that people regularly interact with strangers that cultures invent moralizing gods. These are not gods who sit around the banquet table laughing with detachment at the foibles of humans down below, or gods who punish humans for lousy sacrificial offerings. These are gods who punish humans for being rotten to other humans—in other words, the large religions invent gods who do third-party punishment. No wonder this predicts these religions' adherents being third-party punishers themselves.

Next the twin findings that more market integration and bigger community size were associated with fairer offers (for the former) and more willingness to punish unfair players (for both). I find this to be a particularly challenging pair of findings, especially when framed as the authors thoughtfully did.

The authors ask where the uniquely extreme sense of fairness comes from in humans, particularly in the context of large-scale societies with strangers frequently interacting. And they offer two traditional types of explanations that

are closely related to our dichotomies of intuition versus reasoning and animal roots versus cultural inventions:

- Our moral anchoring in fairness in large-scale societies is a residue and extension of our hunter-gatherer and nonhuman primate past. This was life in small bands, where fairness was mostly driven by kin selection and easy scenarios of reciprocal altruism. As our community size has expanded and we now mostly have one-shot interactions with unrelated strangers, our prosociality just represents an expansion of our small-band mind-set, as we use various green-beard marker shibboleths as proxies for relatedness. I'd gladly lay down my life for two brothers, eight cousins, or a guy who is a fellow Packers fan.
- The moral underpinnings of a sense of fairness lie in cultural institutions and mind-sets that we invented as our groups became larger and more sophisticated (as reflected in the emergence of markets, cash economies, and the like).

This many pages in, it's obvious that I think the former scenario is pretty powerful—look, we see the roots of a sense of fairness and justice in the egalitarian nature of nomadic hunter-gatherers, in other primates, in infants, in the preeminent limbic rather than cortical involvement. But, inconveniently for that viewpoint, that's totally counter to what emerges from these studies—across the twenty-five cultures it's the hunter-gatherers, the ones most like our ancestors, living in the smallest groups, with the highest degrees of relatedness and with the least reliance on market interactions, who show the least tendency toward making fair offers and are least likely to punish unfairness, whether to themselves or to the other guy. None of that prosociality is there, a picture counter to what we saw in chapter 9.

I think an explanation is that these economic games tap into a very specific and artificial type of prosociality. We tend to think of market interactions as being the epitome of complexity—finding a literal common currency for the array of human needs and desires in the form of this abstraction called money. But at their core, market interactions represent an impoverishment of human reciprocity. In its natural form, human reciprocity is a triumph of comfortably and intuitively doing long-term math with apples and oranges—this guy over

here is a superstar hunter; that other guy isn't in his league but has your back if there's a lion around; meanwhile, she's amazing at finding the best mongongo nuts, that older woman knows all about medicinal herbs, and that geeky guy remembers the best stories. We know where one another live, the debit columns even out over time, and if someone is really abusing the system, we'll get around to collectively dealing with them.

In contrast, at its core, a cash-economy market interaction strips it all down to "I give you this now, so you give me that now"—myopic present-tense interactions whose obligations of reciprocity must be balanced immediately. People in small-scale societies are relatively new to functioning this way. It's not the case that small-scale cultures that are growing big and market reliant are newly schooled in how to be fair. Instead they're newly schooled in how to be fair in the artificial circumstances modeled by something like the Ultimatum Game.

HONOR AND REVENGE

Another realm of cross-cultural differences in moral systems concerns what constitutes appropriate response to personal affronts. This harks back to chapter 9's cultures of honor, from Maasai tribesmen to traditional American Southerners. As we saw, such cultures have historical links to monotheism, warrior age groups, and pastoralism.

To recap, such cultures typically see an unanswered challenge to honor as the start of a disastrous slippery slope, rooted in the intrinsic vulnerability of pastoralism—while no one can raid farmers and steal all their crops, someone can rustle a herd overnight—and if this sum'a bitch gets away with insulting my family, he'll be coming for my cattle next. These are cultures that place a high moral emphasis on revenge, and revenge at least in kind—after all, an eye for an eye was probably the invention of Judaic pastoralists. The result is a world of Hatfields and McCoys, with their escalating vendettas. This helps explain why the elevated murder rates in the American South are not due to urban violence or things like robberies but are instead about affronts to honor between people who know each other. And it helps explain why Southern prosecutors and juries are typically more forgiving of such crimes of affronted honor. And it also helps explain the command apparently given by many Southern matriarchs to their sons marching off to join the Confederate fight: come back a winner or come back in a coffin. The shame of surrender is not an option.

SHAMED COLLECTIVISTS AND GUILTY INDIVIDUALISTS

We return to our contrast between collectivist and individualistic cultures (in the studies, as a reminder, “collectivist” has mostly meant East Asian societies, while “individualistic” equals Western Europeans and North Americans). Implicit in the very nature of the contrast are markedly different approaches to the morality of ends and means. By definition, collectivist cultures are more comfortable than individualistic ones with people being used as a means to a utilitarian end. Moreover, moral imperatives in collectivist cultures tend to be about social roles and duties to the group, whereas those in individualistic cultures are typically about individual rights.

Collectivist and individualistic cultures also differ in how moral behavior is enforced. As first emphasized by the anthropologist Ruth Benedict in 1946, collectivist cultures enforce with shame, while individualistic cultures use guilt. This is a doozy of a contrast, as explored in two excellent books, Stanford psychiatrist Herant Katchadourian’s *Guilt: The Bite of Conscience* and NYU environmental scientist Jennifer Jacquet’s *Is Shame Necessary?*³⁰

In the sense used by most in the field, including these authors, shame is external judgment by the group, while guilt is internal judgment of yourself. Shame requires an audience, is about honor. Guilt is for cultures that treasure privacy and is about conscience. Shame is a negative assessment of the entire individual, guilt that of an act, making it possible to hate the sin but love the sinner. Effective shaming requires a conformist, homogeneous population; effective guilt requires respect for law. Feeling shame is about wanting to hide; feeling guilt is about wanting to make amends. Shame is when everyone says, “You can no longer live with us”; guilt is when you say, “How am I going to live with myself?”*

From the time that Benedict first articulated this contrast, there has been a self-congratulatory view in the West that shame is somehow more primitive than guilt, as the West has left behind dunce caps, public flogging, and scarlet letters. Shame is the mob; guilt is internalizing rules, laws, edicts, decrees, and statutes. Yet, Jacquet convincingly argues for the continued usefulness of shaming in the West, calling for its rebirth in a postmodernist form. For her, shaming is particularly useful when the powerful show no evidence of feeling guilt and evade punishment. We have no shortage of examples of such evasion in the American legal system, where one can benefit from the best defense that money or power can buy; shaming can often step into that vacuum. Consider a 1999 scandal at UCLA, when more than a dozen healthy, strapping football players

were discovered to have used connections, made-up disabilities, and forged doctors' signatures to get handicapped parking permits. Their privileged positions resulted in what was generally seen as slaps on the wrist by both the courts and UCLA. However, the element of shaming may well have made up for it—as they left the courthouse in front of the press, they walked past a phalanx of disabled, wheelchair-bound individuals jeering them.³¹

Anthropologists, studying everyone from hunter-gatherers to urbanites, have found that about two thirds of everyday conversation is gossip, with the vast majority of it being negative. As has been said, gossip (with the goal of shaming) is a weapon of the weak against the powerful. It has always been fast and cheap and is infinitely more so now in the era of the Scarlet Internet.

Shaming is also effective when dealing with outrages by corporations.³² Bizarrely, the American legal system considers a corporation to be an individual in many ways, one that is psychopathic in the sense of having no conscience and being solely interested in profits. The people running a corporation are occasionally criminally responsible when the corporation has done something illegal; however, they are not when the corporation does something legal yet immoral—it is outside the realm of guilt. Jacquet emphasizes the potential power of shaming campaigns, such as those that forced Nike to change its policies about the horrific working conditions in its overseas sweatshops, or paper giant Kimberly-Clark to address the cutting of old-growth forests.

Amid the potential good that can come from such shaming, Jacquet also emphasizes the dangers of contemporary shaming, which is the savagery with which people can be attacked online and the distance such venom can travel—in a world where getting to anonymously hate the sinner seems more important than anything about the sin itself.

FOOLS RUSH IN: APPLYING THE FINDINGS OF THE SCIENCE OF MORALITY

How can the insights we already have in hand be used to foster the best of our behaviors and lessen the worst?

Which Dead White Male Was Right?

Let's start with a question that has kept folks busy for millennia, namely, what is the optimal moral philosophy?

People pondering this question have grouped the different approaches into three broad categories. Say there's money sitting there, and it's not yours but no one is looking; why not grab it?

Virtue ethics, with its emphasis on the actor, would answer: because you are a better person than that, because you'll have to live with yourself afterward, etc.

Deontology, with its emphasis on the act: because it's not okay to steal.

Consequentialism, with its emphasis on the outcome: what if everyone started acting that way, think about the impact on the person whose money you've stolen, etc.



Virtue ethics has generally taken a backseat to the other two in recent years, having acquired a quaint veneer of antiquarian fretting over how an improper act tarnishes one's soul. As we'll see, I think that virtue ethics returns through the back door with considerable relevance.

By focusing on deontology versus consequentialism, we are back on the familiar ground of whether ends justify means. For deontologists the answer is "No, people can never be pawns." For the consequentialist the answer is "Yes, for the right outcome." Consequentialism comes in a number of stripes, taken seriously to varying degrees, depending on its features—for example, yes, the end justifies the means if the end is to maximize my pleasure (hedonism), to maximize overall levels of wealth,* to strengthen the powers that be (state consequentialism). For most, though, consequentialism is about classical

utilitarianism—it is okay to use people as a means to the end of maximizing overall levels of happiness.

When deontology and consequentialism contemplate trolleys, the former is about moral intuitions rooted in the vmPFC, amygdala, and insula, while the latter is the domain of the dlPFC and moral reasoning. Why is it that our automatic, intuitive moral judgments tend to be nonutilitarian? Because, as Greene states in his book, “Our moral brains evolved to help us spread our genes, not to maximize our collective happiness.”

The trolley studies show people’s moral heterogeneity. In them approximately 30 percent of subjects were consistently deontologists, unwilling to either pull a lever or push a person, even at the cost of those five lives. Another 30 percent were always utilitarian, willing to pull or push. And for everyone else, moral philosophies were context dependent. The fact that a plurality of people fall into this category prompts Greene’s “dual process” model, stating that we are usually a mixture of valuing means and ends. What’s your moral philosophy? If harm to the person who is the means is unintentional or if the intentionality is really convoluted and indirect, I’m a utilitarian consequentialist, and if the intentionality is right in front of my nose, I’m a deontologist.

The different trolley scenarios reveal what circumstances push us toward intuitive deontology, which toward utilitarian reasoning. Which outcome is better?

For the sort of person reading this book (i.e., who reads and thinks, things to be justifiably self-congratulatory about), when considering this issue at a calm distance, utilitarianism seems like the place to start—maximizing collective happiness. There is the emphasis on equity—not equal treatment but taking everyone’s well-being into equal consideration. And there is the paramount emphasis on impartiality: if someone thinks the situation being proposed is morally equitable, they should be willing to flip a coin to determine which role they play.

Utilitarianism can be critiqued on practical grounds—it’s hard to find a common currency of people’s differing versions of happiness, the emphasis on ends over means requires that you be good at predicting what the actual ends will be, and true impartiality is damn hard with our Us/Them minds. But in theory, at least, there is a solid, logical appeal to utilitarianism.

Except that there’s a problem—unless someone is missing their vmPFC, the appeal of utilitarianism inevitably comes to a screeching halt at some point. For

most people it's pushing the person in front of the trolley. Or smothering a crying baby to save a group of people hiding from Nazis. Or killing a healthy person to harvest his organs and save five lives. As Greene emphasizes, virtually everyone immediately grasps the logic and appeal of utilitarianism yet eventually hits a point where it is clear that it's not a good guide for everyday moral decision making.

Greene and, independently, the neuroscientist John Allman of Caltech and historian of science James Woodward of the University of Pittsburgh have explored the neurobiological underpinnings of a key point—the utilitarianism being considered here is unidimensional and artificial; it hobbles the sophistication of both our moral intuitions and our moral reasoning. A pretty convincing case can be made for utilitarian consequentialism. As long as you consider the immediate consequences. And the longer-term consequences. And the long-long-term consequences. And then go and consider them all over again a few times.

When people hit a wall with utilitarianism, it's because what is on paper a palatable trade-off in the short run (“Intentionally kill one to save five—that obviously increases collective happiness”) turns out not to be so in the long run. “Sure, that healthy person’s involuntary organ donation just saved five lives, but who else is going to get dissected that way? What if they come for me? I kinda like my liver. What else might they start doing?” Slippery slopes, desensitization, unintended consequences, intended consequences. When shortsighted utilitarianism (what Woodward and Allman call “parametric” consequentialism) is replaced with a longer-viewed version (what they call “strategic” consequentialism and what Greene calls “pragmatic utilitarianism”), you get better outcomes.

Our overview of moral intuition versus moral reasoning has generated a dichotomy, something akin to how guys can't have lots of blood flow to their crotch and their brain at the same time; they have to choose. Similarly, you have to choose whether your moral decision making will be about the amygdala or the dlPFC. But this is a false dichotomy, because we reach our best long-term, strategic, consequentialist decisions when we engage both our reasoning and our intuition. “Sure, being willing to do X in order to accomplish Y seems like a good trade-off in the short run. But in the long run, if we do that often enough, doing Z is going to start to seem okay also, and I'd feel awful if Z were done to me, and there's also a good chance that W would happen, and that's going to generate really bad feelings in people, which will result in . . .” And the “feel”

part of that process is not the way Mr. Spock would do it, logically and dispassionately remembering that those humans are irrational, flighty creatures and incorporating that into his rational thinking about them. Instead, this is feeling what the feelings would feel like. This is straight out of chapter 2's overview of Damasio's somatic marker hypothesis: when we are making decisions, we are running not only thought experiments but somatic feeling experiments as well—how is it going to *feel* if this happens?—and this combination is the goal in moral decision making.

Thus, “No way I’d push someone onto the trolley tracks; it’s just wrong” is about the amygdala, insula, and vmPFC. “Sacrifice one life to save five, sure” is the dlPFC. But do long-term strategic consequentialism, and all those regions are engaged. And this yields something more powerful than the cocksureness of knee-jerk intuitionism, the “I can’t tell you why, but this is simply wrong.” When you’ve engaged all those brain systems, when you’ve done the thought experiments and feeling experiments of how things might play out in the long run, and when you’ve prioritized the inputs—gut reactions are taken seriously, but they’re sure not given veto power—you’ll know exactly why something seems right or wrong.

The synergistic advantages of combining reasoning with intuition raise an important point. If you’re a fan of moral intuitions, you’d frame them as being foundational and primordial. If you don’t like them, you’d present them as simplistic, reflexive, and primitive. But as emphasized by Woodward and Allman, our moral intuitions are neither primordial nor reflexively primitive. They are the end products of learning; they are cognitive conclusions to which we have been exposed so often that they have become automatic, as implicit as riding a bicycle or reciting the days of the week forward rather than backward. In the West we nearly all have strong moral intuitions about the wrongness of slavery, child labor, or animal cruelty. But that sure didn’t used to be the case. Their wrongness has become an implicit moral intuition, a gut instinct concerning moral truth, only because of the fierce moral reasoning (and activism) of those who came before us, when the average person’s moral intuitions were unrecognizably different. Our guts learn their intuitions.

Slow and Fast: The Separate Problems of “Me Versus Us” and “Us Versus Them”

The contrast between rapid, automatic moral intuitionism and conscious, deliberative moral reasoning plays out in another crucial realm and is the subject of Greene's superb 2014 book *Moral Tribes: Emotion, Reason, and the Gap Between Us and Them*.³³

Greene starts with the classic tragedy of the commons. Shepherds bring their flocks to a common grazing field. There are so many sheep that there is the danger of destroying the commons, unless people decrease the size of their herds. And the tragedy is that if it is truly a commons, there is no incentive to ever cooperate—you'd range from being a fool if no one else was cooperating to being a successful free rider if everyone else was.

This issue, namely how to jump-start and then maintain cooperation in a sea of noncooperators, ran through all of chapter 10 and, as shown in the widespread existence of social species that cooperate, this is solvable (stay tuned for more in the final chapter). When framed in the context of morality, averting the tragedy of the commons requires getting people in groups to not be selfish; it is an issue of Me versus Us.

But Greene outlines a second type of tragedy. Now there are two different *groups* of shepherds, and the challenge is that each group has a different approach to grazing. One, for example, treats the pasture as a classic commons, while the other believes that the pasture should be divided up into parcels of land belonging to individual shepherds, with high, strong fences in between. In other words, mutually contradictory views about using the pasture.

The thing that fuels the danger and tragedy of this situation is that each group has such a tightly reasoned structure in their heads as to why their way is correct that it can acquire moral weight, be seen as a "right." Greene dissects that word brilliantly. For each side, perceiving themselves as having a "right" to do things their way mostly means that they have slathered enough post-hoc, Haidtian rationalizations on a shapeless, self-serving, parochial moral intuition; have lined up enough of their gray-bearded philosopher-king shepherds to proclaim the moral force of their stance; feel in the most sincere, pained way that the very essence of what they value and who they are is at stake, that the very moral rightness of the universe is wobbling; all of that so strongly that they can't recognize the "right" for what it is, namely "I can't tell you why, but this is how things should be done." To cite a quote attributed to Oscar Wilde, "Morality is simply the attitude we adopt towards people whom we personally dislike."

It's Us versus Them framed morally, and the importance of what Greene calls "the Tragedy of Commonsense Morality" is shown by the fact that most

intergroup conflicts on our planet ultimately are cultural disagreements about whose “right” is righter.

This is an intellectualized, bloodless way of framing the issue. Here’s a different way.

Say I decide that it would be a good thing to have pictures here demonstrating cultural relativism, displaying an act that is commonsensical in one culture but deeply distressing in another. “I know,” I think, “I’ll get some pictures of a Southeast Asian dog-meat market; like me, most readers will likely resonate with dogs.” Good plan. On to Google Images, and the result is that I spend hours transfixed, unable to stop, torturing myself with picture after picture of dogs being carted off to market, dogs being butchered, cooked, and sold, pictures of humans going about their day’s work in a market, indifferent to a crate stuffed to the top with suffering dogs.



I imagine the fear those dogs feel, how they are hot, thirsty, in pain. I think, “What if these dogs had come to trust humans?” I think of their fear and confusion. I think, “What if one of the dogs whom I’ve loved had to experience that? What if this happened to a dog my children loved?” And with my heart racing, I realize that I hate these people, *hate* every last one of them and despise their culture.

And it takes a locomotive’s worth of effort for me to admit that I can’t justify that hatred and contempt, that mine is a mere moral intuition, that there are things that I do that would evoke the same response in some distant person whose humanity and morality are certainly no less than mine, and that but for the randomness of where I happen to have been born, I could have readily had their views instead.

The thing that makes the tragedy of commonsense morality so tragic is the intensity with which you just know that They are deeply wrong.

In general, our morally tinged cultural institutions—religion, nationalism, ethnic pride, team spirit—bias us toward our best behaviors when we are single shepherds facing a potential tragedy of the commons. They make us less selfish in Me versus Us situations. But they send us hurtling toward our worst behaviors when confronting Them and their different moralities.

The dual process nature of moral decision making gives some insights into how to avert these two very different types of tragedies.

In the context of Me versus Us, our moral intuitions are shared, and emphasizing them hums with the prosociality of our Us-ness. This was shown in a study by Greene, David Rand of Yale, and colleagues, where subjects played a one-shot public-goods game that modeled the tragedy of the commons.³⁴ Subjects were given differing lengths of time to decide how much money they would contribute to a common pot (versus keeping it for themselves, to everyone else's detriment). And the faster the decision required, the more cooperative people were. Ditto if you had primed subjects to value intuition (by having them relate a time when intuition led them to a good decision or where careful reasoning did the opposite)—more cooperation. Conversely, instruct subjects to "carefully consider" their decision, or prime them to value reflection over intuition, and they'd be more selfish. The more time to think, the more time to do a version of "Yes, we all agree that cooperation is a good thing . . . but here is why I should be exempt this time"—what the authors called "calculated greed."

What would happen if subjects played the game with someone screamingly different, as different a human as you could find, by whatever the subject's standards of comfort and familiarity? While the study hasn't been done (and would obviously be hard to do), you'd predict that fast, intuitive decisions would overwhelmingly be in the direction of easy, unconflicted selfishness, with "Them! Them!" xenophobia alarms ringing and automatic beliefs of "Don't trust Them!" instantly triggered.

When facing Me-versus-Us moral dilemmas of resisting selfishness, our rapid intuitions are good, honed by evolutionary selection for cooperation in a sea of green-beard markers.³⁵ And in such settings, regulating and formalizing the prosociality (i.e., moving it from the realm of intuition to that of cogitation) can even be counterproductive, a point emphasized by Samuel Bowles.*

In contrast, when doing moral decision making during Us-versus-Them scenarios, keep intuitions as far away as possible. Instead, think, reason, and question; be deeply pragmatic and strategically utilitarian; take their perspective,

try to think what they think, try to feel what they feel. Take a deep breath, and then do it all again.*

Veracity and Mendacity

The question rang out, clear and insistent, a question that could not be ignored or evaded. Chris swallowed once, tried for a voice that was calm and steady, and answered, “No, absolutely not.” It was a bald-faced lie.

Is this a good thing or bad thing? Well, it depends on what the question was: (a) “When the CEO gave you the summary, were you aware that the numbers had been manipulated to hide the third-quarter losses?” asked the prosecutor. (b) “Is this a toy you already have?” asked Grandma tentatively. (c) “What did the doctor say? Is it fatal?” (d) “Does this outfit make me look ____?” (e) “Did you eat the brownies that were for tonight?” (f) “Harrison, are you harboring the runaway slave named Jack?” (g) “Something’s not adding up. Are you lying about being at work late last night?” (h) “OMG, did you just cut one?”

Nothing better typifies the extent to which the meanings of our behaviors are context dependent. Same untruth, same concentration on controlling your facial expression, same attempt to make just the right amount of eye contact. And depending on the circumstance, this could be us at our best or worst. On the converse side of context dependency, sometimes being honest is the harder thing—telling an unpleasant truth about another person activates the medial PFC (along with the insula).^{*36}

Given these complexities, it is no surprise that the biology of honesty and duplicity is very muddy.

As we saw in chapter 10, the very nature of competitive evolutionary games selects for both deception and vigilance against it. We even saw protoversions of both in social yeast. Dogs attempt to deceive one another, with marginal success—when a dog is terrified, fear pheromones emanate from his anal scent glands, and it’s not great if the guy you’re facing off against knows you’re scared. A dog can’t consciously choose to be deceptive by not synthesizing and secreting those pheromones. But he can try to squelch their dissemination by putting a lid on those glands, by putting his tail between his legs—“I’m not scared, no siree,” squeaked Sparky.

No surprise, nonhuman primate duplicity takes things to a whole other level.³⁷ If there is a good piece of food and a higher-ranking animal nearby, capuchins will give predator alarm calls to distract the other individual; if it is a lower-ranking animal, no need; just take the food. Similarly, if a low-ranking capuchin knows where food has been hidden and there is a dominant animal around, he will move away from the hiding place; if it's a subordinate animal, no problem. The same is seen in spider monkeys and macaques. And other primates don't just carry out "tactical concealment" about food. When a male gelada baboon mates with a female, he typically gives a "copulation call." Unless he is with a female who has snuck away from her nearby consortship male. In which case he doesn't make a sound. And, of course, all of these examples pale in comparison with what politico chimps can be up to. Reflecting deception as a task requiring lots of social expertise, across primate species, a larger neocortex predicts higher rates of deception, independent of group size.*

That's impressive. But it is highly unlikely that there is conscious strategizing on the part of these primates. Or that they feel bad or even morally soiled about being deceptive. Or that they actually believe their lies. For those things we need humans.

The human capacity for deception is enormous. We have the most complex innervation of facial muscles and use massive numbers of motor neurons to control them—no other species can be poker-faced. And we have language, that extraordinary means of manipulating the distance between a message and its meaning.

Humans also excel at lying because our cognitive skills allow us to do something beyond the means of any perfidious gelada baboon—we can finesse the truth.

A cool study shows our propensity for this. To simplify: A subject would roll a die, with different results yielding different monetary rewards. The rolls were made in private, with the subject reporting the outcome—an opportunity to cheat.

Given chance and enough rolls, if everyone was honest, each number would be reported about one sixth of the time. If everyone always lied for maximal gain, all rolls would supposedly have produced the highest-paying number.

There was lots of lying. Subjects were over 2,500 college students from twenty-three countries, and higher rates of corruption, tax evasion, and political fraud in a subject's country predicted higher rates of lying. This is no surprise,

after chapter 9's demonstration that high rates of rule violations in a community decrease social capital, which then fuels individual antisocial behavior.

What was most interesting was that across all the cultures, lying was of a particular type. Subjects actually rolled a die twice, and only the first roll counted (the second, they were told, tested whether the die was "working properly"). The lying showed a pattern that, based on prior work, could be explained by only one thing—people rarely made up a high-paying number. Instead they simply reported the higher roll of the two.

You can practically hear the rationalizing. "Darn, my first roll was a 1 [a bad outcome], my second a 4 [better]. Hey, rolls are random; it could just as readily have been 4 as a 1, so . . . let's just say I rolled a 4. That's not really cheating."

In other words, lying most often included rationalizing that made it feel less dishonest—not going whole hog for that filthy lucre, so that your actions feel like only slightly malodorous untruthiness.

When we are lying, naturally, regions involved in Theory of Mind are involved, particularly with circumstances of strategic social deception. Moreover, the dlPFC and related frontal regions are central to a neural circuit of deception. And then insight grinds to a halt.³⁸

Back to the theme introduced in chapter 2 of the frontal cortex, and the dlPFC in particular, getting you to do the harder thing when it's the right thing to do. And in our value-free sense of "right," you'd expect the dlPFC to activate when you're struggling to do (a) the morally right thing, which is to avoid the temptation to lie, as well as (b) the strategically right thing, namely, once having decided to lie, doing it effectively. It can be *hard* to deceive effectively, having to think strategically, carefully remember what lie you're actually saying, and create a false affect ("Your Majesty, I bring terrible, sad news about your son, the heir to the throne [yeah, we ambushed him—high fives!]).^{*} Thus activation of the dlPFC will reflect both the struggle to resist temptation and the executive effort to wallow effectively in the temptation, once you've lost that struggle. "Don't do it" + "if you're going to do it, do it right."

This confusion arises in neuroimaging studies of compulsive liars.^{*39} What might one expect? These are people who habitually fail to resist the temptation of lying; I bet they have atrophy of something frontocortical. These are people who habitually lie and are good at it (and typically have high verbal IQs); I bet they have expansion of something frontocortical. And the studies bear out both predictions—compulsive liars have increased amounts of white matter (i.e., the axonal cables connecting neurons) in the frontal cortex, but lesser amounts of

gray matter (i.e., the cell bodies of the neurons). It's not possible to know if there's causality in these neuroimaging/behavior correlates. All one can conclude is that frontocortical regions like the dlPFC show multiple and varied versions of "doing the harder thing."

You can dissociate the frontal task of resisting temptation from the frontal task of lying effectively by taking morality out of the equation.⁴⁰ This is done in studies where people are *told* to lie. (For example, subjects are given a series of pictures; later they are shown an array of pictures, some of which are identical to ones in their possession, and asked, "Is this a picture you have?" A signal from the computer indicates whether the subject should answer honestly or lie.) In this sort of scenario, lying is most consistently associated with activation of the dlPFC (along with the nearby and related ventrolateral PFC). This is a picture of the dlPFC going about the difficult task of lying effectively, minus worrying about the fate of its neuronal soul.

The studies tend to show activation of the anterior cingulate cortex (ACC) as well. As introduced in chapter 2, the ACC responds to circumstances of conflicting choices. This occurs for conflict in an emotional sense, as well as in a cognitive sense (e.g., having to choose between two answers when both seem to work). In the lying studies the ACC isn't activating because of moral conflict about lying, since subjects were instructed to lie. Instead, it's monitoring the conflict between reality and what you've been instructed to report, and this gums up the works slightly; people show minutely longer response times during lying trials than during honest ones.

This delay is useful in polygraph tests (i.e., lie detectors). In the classic form, the test detected arousal of the sympathetic nervous system, indicating that someone was lying and anxious about not getting caught. The trouble is that you'd get the same anxious arousal if you're telling the truth but your life's over if that fallible machine says otherwise. Moreover, sociopaths are undetectable, since they don't get anxiously aroused when lying. Plus subjects can take countermeasures to manipulate their sympathetic nervous system. As a result, this use of polygraphs is no longer admissible in courts. Contemporary polygraph techniques instead home in on that slight delay, on the physiological indices of that ACC conflict—not the moral one, since some miscreant may have no moral misgivings, but the cognitive conflict—"Yeah, I robbed the store, but no, wait, I have to say that I didn't." Unless you thoroughly believe your lie, there's likely to be that slight delay, reflecting the ACC-ish cognitive conflict between reality and your claim.

Thus, activation of the ACC, dlPFC, and nearby frontal regions is associated with lying on command.⁴¹ At this point we have our usual issue of causality—is activation of, say, the dlPFC a cause, a consequence, or a mere correlate of lying? To answer this, transcranial direct-current stimulation has been used to inactivate the dlPFC in people during instructed-lying tasks. Result? Subjects were slower and less successful in lying—implying a causal role for the dlPFC. And to remind us of how complicated this issue is, people with damage to the dlPFC are less likely to take honesty into account when honesty and self-interest are pitted against each other in an economic game. So this most eggheady, cognitive part of the PFC is central to both resisting lying and, once having decided to lie, doing it well.

This book's focus is not really how good a liar someone is. It's whether we lie, whether we do the harder thing and resist the temptation to deceive. For more understanding of that, we turn to a pair of thoroughly cool neuroimaging studies where subjects who lied did so not because they were instructed to but because they were dirty rotten cheaters.

The first was carried out by the Swiss scientists Thomas Baumgartner, Ernst Fehr (whose work has been noted previously), and colleagues.⁴² Subjects played an economic trust game where, in each round, you could be cooperative or selfish. Beforehand a subject would tell the other player what their strategy would be (always/sometimes/never cooperate). In other words, they made a promise.

Some subjects who promised to always cooperate broke their promise at least once. At such times there was activation of the dlPFC, the ACC, and, of course, the amygdala.*⁴³

A pattern of brain activation before each round's decision *predicted* breaking of a promise. Fascinatingly, along with predictable activation of the ACC, there'd be activation of the insula. Does the scoundrel think, "I'm disgusted with myself, but I'm going to break my promise"? Or is it "I don't like this guy because of X; in fact, he's kind of disgusting; I owe him nothing; I'm breaking my promise"? While it's impossible to tell, given our tendency to rationalize our own transgressions, I'd bet it's the latter.

The second study comes from Greene and colleague Joseph Paxton.⁴⁴ Subjects in a scanner would predict the outcome of coin tosses, earning money for correct guesses. The study's design contained an extra layer of distracting nonsense. Subjects were told the study was about paranormal mental abilities, and for some of the coin tosses, for this concocted reason, rather than state their

prediction beforehand, subjects would just think about their choice and then tell afterward if they were right. In other words, amid a financial incentive to guess correctly, there were intermittent opportunities to cheat. Crucially, this was detectable—during the periods of forced honesty, subjects averaged a 50 percent success rate. And if accuracy jumped a lot higher during opportunities to cheat, subjects were probably cheating.

The results were pretty depressing. Using this form of statistical detection, about a third of the subjects appeared to be big-time cheaters, with another sixth on the statistical border. When cheaters cheated, there was activation of the dlPFC, as we'd expect. Were they struggling with the combination of moral and cognitive conflict? Not particularly—there wasn't activation of the ACC, nor was there the slight lag time in response. Cheaters typically didn't cheat at every opportunity; what did things look like when they resisted? Here's where you saw the struggling—even greater activation of the dlPFC (along with the vlPFC), the ACC roaring into action, and a significant delay in response time. In other words, for people capable of cheating, the occasional resistance seems to be the outcome of major neurobiological Sturm und Drang.

And now for probably the most important finding in this chapter. What about subjects who never cheated? There are two very different scenarios, as framed by Greene and Paxton: Is resisting temptation at every turn an outcome of “will,” of having a stoked dlPFC putting Satan into a hammerlock of submission? Or is it an act of “grace,” where there's no struggle, because it's simple; you don't cheat?

It was grace. In those who were always honest, the dlPFC, vlPFC, and ACC were in veritable comas when the chance to cheat arose. There's no conflict. There's no working hard to do the right thing. You simply don't cheat.

Resisting temptation is as implicit as walking up stairs, or thinking “Wednesday” after hearing “Monday, Tuesday,” or as that first piece of regulation we mastered way back when, being potty trained. As we saw in chapter 7, it's not a function of what Kohlbergian stage you're at; it's what moral imperatives have been hammered into you with such urgency and consistency that doing the right thing has virtually become a spinal reflex.

This is not to suggest that honesty, even impeccable honesty that resists all temptation, can only be the outcome of implicit automaticity.⁴⁵ We can think and struggle and employ cognitive control to produce similar stainless records, as

shown in some subsequent work. But in circumstances like the Greene and Paxton study, with repeated opportunities to cheat in rapid succession, it's not going to be a case of successfully arm wrestling the devil over and over. Instead, automaticity is required.

We've seen something equivalent with the brave act, the person who, amid the paralyzed crowd, runs into the burning building to save the child. "What were you thinking when you decided to go into the house"? (Were you thinking about the evolution of cooperation, of reciprocal altruism, of game theory and reputation?) And the answer is always "I wasn't thinking anything. Before I knew it, I had run in." Interviews of Carnegie Medal recipients about that moment shows precisely that—a first, intuitive thought of needing to help, resulting in the risking of life without a second thought. "Heroism feels and never reasons," to quote Emerson.⁴⁶

It's the same thing here: "Why did you never cheat? Is it because of your ability to see the long-term consequences of cheating becoming normalized, or your respect for the Golden Rule, or . . . ?" The answer is "I don't know [shrug]. I just don't cheat." This isn't a deontological or a consequentialist moment. It's virtue ethics sneaking in the back door in that moment—"I don't cheat; that's not who I am." Doing the right thing *is* the easier thing.