CHAPTER 18
Morality Is for Choosing Sides

Peter DeScioli
Robert Kurzban

Why did moral judgment evolve?
To help people choose sides when conflicts erupt within groups with complex coalitions and power hierarchies.

Theories of morality have largely tried to explain the brighter side of behavior, answering questions about why people behave in ways that are kind, generous, and good. Our proposal focuses not on explaining moral behavior but, rather, on explaining moral judgment. Consider someone reading a news story about a man who pays a woman to have sex with him. Many people would judge—intuitively (Haidt, 2012)—that both the man’s and woman’s actions are morally wrong. Our interest lies in the explanation for these and similar judgments.

Theories that attempt to explain moral behavior often point to altruism or benefits (de Waal, 1996; Krebs, 2005; Ridley, 1996; Wright, 1994). The theory of reciprocal altruism (Trivers, 1971), for instance, explains why people enter into voluntary exchanges with one another: to reap the benefits of trade. Such theoretical moves are considerably less straightforward for explaining moral judgment. For example, condemning the exchange of sex for money does not transparently confer benefits to the condemner.

What, then, might be the benefits gained through moral judgments? Consider a situation in which a person accuses someone of witchcraft, such as in Arthur Miller’s The Crucible. Specifically, suppose that a young, low-status woman accuses an older, more prominent woman of witchcraft. Other members of the community can respond in a few different ways.

One obvious move for a self-interested observer is to curry favor with the higher-status woman. Choosing sides based on status often occurs in very hierarchical groups such as the military (Fiske, 1992). It is also observed in nonhuman animals: For instance, hyenas join fights and support the higher-status and more formidable fighter (Holekamp, Sakai, & Lundrigan, 2007). This strategy has a downside: It empowers high-status individuals to win all of their conflicts and hence gives them an incentive to exploit other people; they essentially become dictators (Boehm, 1999). Even so, individuals can benefit by siding with dictators because they avoid making powerful enemies. However, humans often do the oppo-
They oppose the disputant who has taken sides based on status or relationships. Moral side-takers choose sides based on actions, siding with the lower-status accuser, as in the hypothetical (and actual) witchcraft case. When might siding with the lower-status accuser be an advantage?

A second strategy for choosing sides is based on relationships: Support closer family or friends, even if they are lower status. Individuals can gain by supporting allies if those allies in turn support them in the future. But alliances have a downside too: If each disputant has a cohort of close friends, then the dispute will expand to include more people on each side and could be even more costly to both the original disputants and their supporters. Research shows that alliances can be extremely damaging at every scale of conflict, from personal to international disputes (Snyder, 1984, 1997; Cooney, 1998, 2003).

Although humans often support their friends and family in conflicts, they do not always do so. This was the case in The Crucible and in numerous real-world witchcraft accusations. Many societies judge black magic to be so morally wrong that it gives cause to abandon one's closest friends and even to seek their death. Executions for witchcraft continue in modern times. In India, for instance, the National Crime Records Bureau documented 2,097 murders of accused witches between 2000 and 2012, despite new laws prohibiting witch hunts (Kapoor, 2015). Around the world, similarly fatal judgments regularly occur for premarital sex, homosexuality, blasphemy, and other harmless offenses that are punishable by death in some societies (Appiah, 2010; Levy, 1993; Sarhan & Burke, 2009; United Nations Commission on Human Rights, 2000). How could it be advantageous to turn against someone, even family and friends, merely because they have (allegedly) done something deemed morally wrong by the community?

We have proposed (DeScioli & Kurzban, 2013) that the benefit of siding with moralistic accusers occurs when other third parties to the conflict do so as well. Moral judgment functions as a side-taking strategy and provides an alternative strategy to choosing sides based on status or relationships. Moral side-takers choose sides based on actions. They oppose the disputant who has taken the more morally wrong action—whether prostitution, witchcraft, homicide, or blasphemy—as established by previous moral debates in the community.

The moral side-taking strategy avoids two key problems with choosing sides based on status and alliances. First, observers do not empower dictators because they do not always side with the same people. Second, they do not create escalating and expanding alliances because observers all choose the same side, provided they use the same moral rules. Moral judgment allows observers to dynamically coordinate their side-taking choices in the sense that they all take the same side, but they can also dynamically change whom they support based on the actions each party has taken. Notice that moral side-taking is effective at coordination only when everyone agrees, or at least acknowledges, what counts as a morally wrong action.

Hence, moral judgment adds to the human repertoire of strategies for managing other people’s conflicts. It does not entirely displace bandwagon or alliance strategies because choosing sides is a coordination game, and coordination games have multiple equilibria (Schelling, 1960). But morality does explain why people sometimes oppose powerful people and close friends—because morality is designed for exactly this purpose, so as to avoid the costs of those strategies.

The side-taking theory explains why moral condemnation can be so destructive. Moral condemnation causes great harm to alleged wrongdoers for harmless or beneficial behaviors, including witchcraft, premarital sex, homosexuality, interest-bearing loans, and scientific research. Popular theories of morality based on cooperation (de Waal, 1996; Krebs, 2005; Ridley, 1996; Wright, 1994) predict that moral judgment will generally maximize welfare, but instead many humans seek prison or death for harmless offenses. In contrast, the side-taking theory is consistent with this destructive behavior because moral judgment functions not to promote welfare but to synchronize side-taking, even if doing so harms many others.

This view of moral judgment explains another important moral phenomenon: people’s decisions to comply with moral rules even when breaking the rules benefits them. In a social world in which the community gangs up against wrongdoers, it is costly
to engage in prohibited actions. The side-taking hypothesis therefore simultaneously accounts for condemnation as well as conscience, psychological mechanisms designed to inhibit actions deemed wrong by the local community.

The side-taking theory explains why people's moralistic punishments are aimed at retribution rather than deterrence, as documented by moral psychology (Carlsmith, Darley, & Robinson, 2002). Theories based on cooperation straightforwardly predict that moralistic punishment will aim at deterring harmful behavior. Instead, people seek retribution for wrongdoers independent of the potential for punishment to deter future violations. The side-taking account holds that an observer's retributive motives are designed to direct their aggression toward the weaker side of a dispute in order to convincingly join the stronger side, where the stronger side in this case means the side with the moral high ground and hence the majority of supporters. Moralistic punishment is retributive because it is designed for side-taking rather than deterring harm.

The side-taking theory also explains why moral judgment includes an ideal of impartiality. Although people's judgments are, in fact, often biased and partial, people at the same time advocate an ideal of impartiality, especially for their opponent's judgments. The side-taking hypothesis holds that the ideal of impartiality functions to decouple moral side-taking from alliances, ultimately to avoid the costs of escalating alliances in disputes.

Finally, this theory illuminates variation in moral rules across individuals and groups. If the dynamic coordination view is correct, then many different moral rules could serve the function of synchronizing side-taking, as long as the local community agrees on the rules. Different societies have different types of conflicts, and people mint new moral rules to cover them. Further, individuals can differ in how they are personally affected by particular rules. For instance, people who pursue short-term mating are worse off when promiscuity is moralized and punished (Kurzban, Dukes, & Weeden, 2010; Weeden, 2003). Other people who pursue long-term mating might benefit from moralizing promiscuity in the interest of guarding their mates. These differences in incentives explain why people differ and disagree about moral rules (DeScioli, Massenkov, Shaw, Petersen, & Kurzban, 2014; Kurzban et al., 2010; Robinson & Kurzban, 2007).

**Historical Context**

The historical context for the side-taking theory includes two parallel but mostly separate research strands: moral psychology and evolutionary theories of morality.

Research in moral psychology has focused on proximate psychological questions. Studies typically examine people's moral judgments about someone's actions (or inactions) in controlled vignettes. Researchers tend to examine issues such as selfish motives versus the greater good, compliance with moral rules when anonymous, intentional versus accidental violations, taboo trade-offs between wrongful actions and overall welfare, and the desire to punish wrongdoers (reviewed in Haidt, 2012).

This research tradition in moral psychology has been largely silent about the evolutionary functions that explain why humans make moral judgments at all. Many researchers either do not address the evolved functions of moral judgment or refer to generic and vague functions, such as the folk wisdom that morality holds society together.

In parallel, evolutionary scholars have viewed morality through the lens of altruism. Starting with Darwin (1871), this was primarily a theoretical problem, asking how natural selection could favor altruistic behavior. Researchers developed models to show how cooperation can evolve, including kin selection, reciprocity, partner choice, and costly signaling. Evolutionary researchers tested these models in thousands of empirical studies, often on nonhuman animals, and many models have extremely impressive empirical support.

However, very little work in the evolutionary tradition measures, or even engages with, moral judgment. Researchers with an evolutionary perspective have largely assumed that cooperation and morality are the same thing (e.g., de Waal, 1996; Krebs,
Hence there is a stark divide between moral psychology, which has proceeded with relatively little theory, and evolutionary accounts of morality, largely uninformed by empirical findings from moral psychology.

Haidt (2007, 2012) began to fuse these two research traditions. He combined moral psychology, cross-cultural research, and evolutionary theories to create a set of fundamental moral foundations, each grounded by different evolutionary models—kin selection, reciprocity, group cooperation, dominance hierarchies, and pathogen avoidance. The result was an impressive overarching theory that had strong appeal both for moral psychologists and for evolutionary researchers.

Haidt’s moral foundations theory is an impressive attempt to reconcile moral psychology with evolution. However, the account misses distinctive elements of human morality. The evolutionary ideas that animate moral foundations theory apply to many different animal species, but moral judgment is an extreme and unusual—possibly unique—human trait, analogous to an elephant’s trunk. If researchers applied only broad theories about animal noses to understand an elephant’s trunk, they would be missing the trunk’s unique grasping and communication abilities.

The theories underlying the moral foundations explain why people show behaviors such as parental care, trade, and dominance. But they do not explain why people make moral judgments about these behaviors. To return to the opening example, reciprocity theory does not explain why people morally judge the act of exchanging sex for money, especially because prostitution is an exchange. Similarly, none of the foundations explain why moral judgment focuses particularly on actions and differs in this respect from people’s decisions about welfare, precautions, economics, and conventions. Traditional evolutionary models predict consequentialist rather than deontological mechanisms (DeScioli & Kurzban, 2009a). Last, traditional evolutionary models do not explain why people disagree about morality and why they debate the moral rules in their community.

The side-taking theory develops additional game-theoretic tools to understand what is distinctive about human moral judgment. Rather than using previous evolutionary models, it develops a new model based on side-taking games to explain the unique human behaviors revealed by moral psychology. It addresses why humans assign moral values to actions, announce moral judgments to other individuals, debate moral rules, and show aggression toward wrongdoers. It provides an explanation for why moral judgment is deontological rather than consequentialist, why punishment is aimed at retribution rather than deterrence, why judgments are held to an ideal of impartiality, and why moral rules vary over time and across cultures. We propose that these moral phenomena result from an evolved strategy for choosing sides in disputes. As such, moral judgment is part of a larger repertoire of adaptations for managing one’s own and others’ conflicts, including the cognitive abilities to assess an opponent’s fighting power, recognize property conventions, and form alliances (DeScioli & Karpoff, 2015; DeScioli, Karpoff, & De Freitas, 2017; DeScioli & Kurzban, 2009b; DeScioli, Rosa, & Gutchess, 2015; DeScioli & Wilson, 2011).

Theoretical Stance

The side-taking theory differs from other theories in how it treats some debates in the literature and, more important, in the functions it proposes for moral mechanisms.

First, on the perennial issue of whether morality is universal or culturally relative, some scholars assume that an evolutionary basis for morality implies that humans will have a small set of universal moral rules and, further, that cultural variation undermines evolutionary accounts (e.g., Prinz, 2007). The side-taking theory, in contrast, holds that humans possess the evolved ability to create and learn new moral rules so that they can be tailored to new types of conflict. As a result, moral cognition is, in itself, universal, while at the same time moral rules differ across groups and within groups over time. Some rules are more stable than others.
because they tend to be supported by majorities, such as rules against lying, stealing, and killing, whereas other rules are more variable because they receive mixed support, such as rules about promiscuity and drug use.

Second, there is a related issue about whether morality is innate or learned. We take the position that learning cannot occur without innate mechanisms specialized for learning in that domain (Pinker, 1989), so the usual dichotomy is misleading. The side-taking theory holds that moral cognition includes mechanisms for learning the active moral rules in the social environment, including different rules for different subgroups and types of interactions. It further holds that people do not only passively internalize the group’s rules but rather, they actively advocate for self-serving rules and readily violate rules when they can get away with it.

Third, there is a debate about whether moral judgment is intuitive or deliberative. The side-taking theory holds that moral judgment is largely unconscious, like many complex cognitive processes. However, a critical part of its function is to persuade other people to take the same side. For this purpose, people have the ability to formulate their moral judgments into language so they can be announced to others. Moreover, people can simulate moral debates in their private thoughts in order to build more convincing moral arguments. These communicative elements explain why moral judgment has a deliberative component.

Fourth, there is a question about whether moral judgment is a single process or multiple processes (DeScioli & Kurzban, 2009a; DeScioli, Asao, & Kurzban, 2012). We first note that every major cognitive ability includes a large number of processes, just as any software application does. The real question is whether moral judgment’s many processes are unified by an overarching function, just as word processing or e-mail applications have overarching functions. The side-taking theory holds that moral judgment is indeed unified in this sense because it is structured by the primary function of choosing sides in disputes. In order to perform this function, moral cognition interacts with a wide array of mechanisms specialized for different areas of social life, including mechanisms for processing kin relations, reciprocity, property, hierarchy, and coalitions. This is, again, analogous to the complex interactions of apps for e-mail, photos, and social networks on modern phones. Nevertheless, moral cognition is a distinct and unified program organized around the problem of choosing sides.

Last, the side-taking theory differs in the functions it proposes for moral cognition. Broadly, the primary function of moral judgment is not to guide one’s own behavior (conscience) but to judge other people’s behavior (condemnation). The conscience component of moral judgment is essentially defensive. People morally evaluate their own potential actions in order to avoid other people’s condemnation. Because conscience functions to simulate and avoid condemnation, the structure of moral judgment is best understood from the condemner’s perspective. Condemners face the problem of choosing sides and doing so in a landscape of prior loyalties and status hierarchies. The side-taking theory uses this adaptive problem to understand how moral judgment works, differing from cooperation theories that view morality as designed to motivate good behavior.

Evidence

There is, of course, a tremendous amount of evidence about how moral judgments operate. We focus on a few patterns of evidence that we think are crucial for inferring the functions of moral judgment. First and foremost, we think any theory of morality must explain why moral judgment focuses on the actions people choose rather than only on the consequences they intend.

In moral philosophy, consequentialism is the idea that the morality of an act depends only on the consequences of the act (Sinnott-Armstrong, 2006). In contrast, deontological theories are nonconsequentialist because they also consider the category of the action, such as lying or stealing, independent of the intended consequences. This allows deontological philosophers such as
Kant to conclude, for example, that lying is morally wrong even if it can save lives (Kant, 1785/1993).

Many experiments in moral psychology show deontological patterns in people’s moral judgments. A well-known example is the finding that most people think it is immoral to push one person off of a footbridge in order to save five people from being killed by a runaway trolley. In this case, people judge the action of killing to be morally wrong, even if it leads to better consequences (fewer deaths). In an interesting contrast, however, most people think it is permissible to switch the trolley to a side track, where it will kill one person, to save the five people on the main track. These results, and many others like them (Baron, 1994; Baron & Spranca, 1997; De Freitas, DeScioli, Nemirov, Massenkoff, & Pinker, in press; DeScioli, Bruening, & Kurzban, 2011; DeScioli, Christner, & Kurzban, 2011; Kurzban et al., 2012; Mikhail, 2007; Tetlock, 2003; Waldmann & Dieterich, 2007), illustrate that people’s evaluations of the wrongness of actions depends on the details of the actions themselves, as opposed to the intended outcomes (which are the same in both the footbridge and switch cases).

This basic observation is not predicted by many prominent theories of morality. We take a moment to come at this problem obliquely because it is easy to overlook. Consider a different context—parental care. Kin selection theory explains why some organisms are designed to provide resources to offspring and also explains how parents make trade-offs when allocating resources across multiple offspring (Hamilton, 1964; Trivers, 1974). Now imagine that researchers observed a species of bird in which mothers sometimes eject eggs from the nest. The researchers propose that this ejection behavior maximizes inclusive fitness by optimally allocating resources to higher quality eggs while ejecting lower quality eggs.

Notice first that this is a specific consequentialist function: Mothers maximize a consequence: inclusive fitness. This makes sense because natural selection is a process driven by consequences, and organisms are usually consequentialist—most animals do not shy from killing, lying, stealing, infanticide, siblicide, or cannibalism when they can maximize fitness by doing so (e.g., Mock, 2004).

The allocation hypothesis for ejection makes specific predictions tied to its proposed function. A mother bird’s ejection behavior should be sensitive to factors that affect costs and benefits, such as the number of other eggs, the scarcity of food in the environment, the risk of predation, or the age and reproductive potential of the mother.

What if, instead, researchers observed that a mother’s ejections depended primarily on the egg’s color, independent of its quality or the number of other eggs? This observation would constitute an anomaly left unexplained by the theory. If it was found repeatedly, over and over, that the color of the eggs overrides the cost–benefit calculus of kin selection, then the parental allocation hypothesis would be called into question.

Recognizing the theory’s failure, researchers might look further and find that this bird species is parasitized by cuckoo eggs that tend to differ in color (Brooke & Davies, 1988). Suppose it turns out that the mother’s ejection behavior is not designed to optimally distribute resources among her offspring but rather to remove cuckoo parasites. In this case, the theory’s empirical failure would allow researchers to discover an altogether different type of explanation.

Now consider a theory of morality that proposes that moral judgments are designed to improve the overall welfare of families, friends, or groups (de Waal, 1996; Krebs, 2005; Ridley, 1996; Wright, 1994). Such theories predict that people should condemn and desire to punish acts depending on the welfare consequences. In particular, people should condemn acts that lead to aggregate welfare losses and not condemn acts that lead to fitness gains. The way in which these gains and losses are realized—alogous to the color of the eggs—should be irrelevant. Evidence that people’s moral judgments closely track the way gains are produced, the particular actions taken, the means by which goals are sought, is evidence against the welfare-improvement theory. Even worse for the theory is the condemnation of actions that produce obvious and large welfare gains. If moral judgment were for improving welfare, pushing the man off of the footbridge should be praiseworthy, not blameworthy.
Arguably still worse for altruism theories are moral rules that guarantee welfare losses. Across cultures, moral rules prohibit any number of victimless, mutually profitable transactions. Historically, an obvious example is the prohibition against charging interest, which prevents mutually profitable loans. In India, the prohibition against killing cows has long caused substantial harms (Suri, 2015). Any number of similar rules continue to undermine potential welfare gains.

We suggest that the tremendous array of data showing that people’s judgments are deontological, along with the ubiquity of welfare-destroying moral rules, all constitute serious evidence against welfare-based theories of morality.

The side-taking hypothesis does not run afoul of these problems. This theory requires that a rule is known and that its violation can be recognized by observers; because rules are for coordinated side-taking rather than welfare-enhancement, they can include a wide range of contents, including welfare-destroying contents. In short, deontological judgment is a set of observations that is, we think, fatal for welfare theories but consistent with the side-taking theory.

There are several other areas of active research that provide evidence relevant to the side-taking hypothesis. First, research has found that people’s tendency to moralize an issue depends on their power and alliances (Jensen & Petersen, 2011; Petersen, 2013). This evidence supports the idea that moral judgment is a strategy that people selectively deploy depending on whether they are most advantaged when others choose sides according to moral judgment, power, or alliances. Second, the side-taking theory points to impartiality as a core feature of moral judgment because it is designed as an alternative to partial alliances. Recent work on fairness judgments points to a similar role for impartiality in suppressing alliances in the context of allocating resources (Shaw, 2013). Third, the side-taking hypothesis emphasizes variability in moral rules and also people’s debates and arguments about which moral rules will structure side-taking in their community. Consistent with this idea, research shows that people actively advocate for the moral rules that most advantage them over other people (Aarøe & Petersen, 2013; DeScioli et al., 2014; Kurzban et al., 2010; Petersen, Aarøe, Jensen, & Curry, 2014; Tybur, Liberman, Kurzban, & DeScioli, 2013).

**Extension and Expansion**

One area for expansion is investigating how people decide whether to enter conflicts and which side-taking strategy to use if they do so. The dynamic coordination hypothesis proposes that morality is designed around the problem of taking sides in disputes, but it does not require that people always use moral judgment to choose sides. In some situations, one might choose to side with, for example, one’s close relative or ally or with the higher-status individual. The best strategy depends on the details of the situation. We predict that people will use moral judgment to choose sides as a function of features of the situation, such as the magnitude of the moral violation, the relative status of the individuals involved, the number of observers to the actions, and other elements that affect an individual’s costs and benefits in the side-taking game.

The side-taking proposal also raises the question of why observers do not always sit out of disputes to avoid any fighting costs to themselves. Indeed, if there were no social costs to sitting out, then the dynamic coordination hypothesis would be contradicted, because players would not have an incentive to choose sides in the first place. However, we suspect that sitting out is often costly and damages preexisting relationships, especially when conflicts involve one’s friends and allies. One goal for future research is to measure the damage to relationships caused by sitting out of conflicts when one’s friends and allies are involved. Insofar as one of the functions of friendship is to cultivate allies when disputes arise (DeScioli, Kurzban, Koch, & Liben-Nowell, 2011; DeScioli & Kurzban, 2009b; DeScioli & Kurzban, 2011), we suspect that failing to come to a friend’s aid in conflicts will indeed damage these relationships, possibly to the same degree as siding against one’s friend. If so, then when an observer is confronted by a dispute between two of their friends, sitting out
might, in some cases, be the worst option because it damages both relationships. Further, siding against a friend who is morally wrong (e.g., someone who lied or cheated) might not damage that relationship as much as when the friend is in the right, because at least the friend could still count on the observer’s support when they are not in the wrong in the future. Additional work can examine how observers manage trade-offs between coordinating with other observers and minimizing damage to their own relationships with each side of the dispute.

REFERENCES


Morality Is for Choosing Sides


