

The Tripartite Theory of Machiavellian Morality: Judgment, Influence, and Conscience as Distinct Moral Adaptations

Kelly Asao and David M. Buss

Introduction

Debates surrounding morality dominate much of people's everyday conversations (Dunbar, 2004). Humans are fascinated by uncovering, discussing, and analyzing the moral decisions of themselves and others. People adamantly believe that their opinions on the subject of morality are objective and impartial. Can they be correct in these beliefs?

Consider a recent media explosion, the fallout after the U.S. government ordered drone strikes that killed Anwar al-Awlaki, a U.S. citizen and al-Qaida propagandist in Yemen. Many Americans were outraged that the target of the strike was a fellow citizen despite his affiliation with al-Qaida. They felt numerous morally charged emotions, such as disgust, fear, anger, and confusion. Additionally, people experienced a multitude of diverse cognitions. Some people believed that the strike was justified because of al-Awlaki's terrorist affiliations. Others were concerned that an American citizen, a member of their in-group, was killed without due process of law. This event brought attention to the legality of the targeted killings of the U.S. drone program. Despite the fact that drone bombings have killed several innocent Yemeni citizens, moral outrage was only sparked after the strike was ordered on an al-Qaida member who also happened to be an American citizen.

The authors thank Dan Conroy-Beam, Frank Mann, Todd Shackelford, and Joy Wyckoff for valuable suggestions on an earlier version of this chapter.

K. Asao (✉)

Department of Psychology, The University of Texas at Austin,

1 University Station, Austin, TX 78712-0187, USA

e-mail: kellyasao@gmail.com

D.M. Buss

The University of Texas at Austin, Austin, USA

Group membership appears to dramatically alter people's feelings, beliefs, and behaviors when it comes to moral outrage. This demonstrates, to the surprise of few scientists, that moral intuitions are rarely dispassionate and objective. However, the key point from this example is that real-life situations perceived as moral violations evoke a wide array of emotional, cognitive, and behavioral phenomena. A deeper understanding of the mechanisms responsible for moral intuitions and actions could help elucidate such complicated social issues.

Some moral theorists have treated morality as a unitary phenomenon, treating distinct components of moral reasoning and behavior as singular in nature, and hence amenable to a unitary explanatory framework. Perhaps most frequently, scientists treat morality as more or less synonymous with altruism or cooperation (Gintis, Bowles, Boyd, & Fehr, 2003; Wilson, 2012). Studies demonstrating helping behavior in nonhuman primates and which provide the foundation for the argument that cooperation is the evolutionary root of morality focus solely on one aspect of moral behavior, namely, conscience (Tomasello & Vaish, 2013). Similarly, research has revealed that nonhuman animals and young children avoid unfair distributions of goods (Bloom, 2013; Brosnan & de Waal, 2003; Range, Horn, Viranyi, & Huber, 2009). However, this does not necessarily indicate that maintaining fairness or cooperation is the evolutionary root of morality.

In contrast, other researchers have taken the opposite approach, viewing morality as a multitude of distinct categories of adaptations or decision rules connected only loosely or terminologically (Graham et al., 2013; Haidt & Joseph, 2004; Kohlberg & Hersh, 1977; Shweder, Much, Mahapatra, & Park, 1997; Stich, 2006). Haidt and Joseph, for example, argue that the variety of moral rules can be traced to a few distinct moral foundations: care/harm, fairness/cheating, loyalty/betrayal, authority/subversion, and sanctity/degradation. For example, compassion and kindness evolved as a specific response to the suffering of one's offspring. In the modern environment, compassion is extended to many other contexts including strangers and baby seals.

Within the past decade, however, researchers in evolutionary psychology have made explicit an important implicit distinction between two major classes of morality: condemnation and conscience (DeScioli & Kurzban, 2009, 2013, Sperber & Baumard, 2012). Broadly, a distinction is made between moral decisions applied to others' behavior and those applied to one's own behavior. Furthermore, DeScioli and Kurzban were the first to highlight the importance of distinguishing between moral judgment and moralistic punishment.

In contrast to earlier approaches, we argue that the concept of morality comprises three distinct adaptations: moral judgment, moral influence, and moral conscience, each consisting of mechanisms designed to solve distinct adaptive problems. *Moral judgment*, we propose, is an adaptation subsuming a suite of evolved psychological mechanisms designed to determine whether a conspecific is *exploitative* or *prosocial*, that is, intentionally imposes either a net cost or a net benefit on one's inclusive fitness. The moral judgment mechanisms then store that information in the memory systems by identifying individuals as costly or beneficial relationship partners. This information is then used to avoid costly people, switch relationship partners, and track social relationships across time. For example, if a new acquaintance has a history of betraying friends, he probably will not make a good ally in times of crisis.

Moral influence, we propose, is a closely related, yet functionally distinct, adaptation. It consists of mechanisms that evolved to identify the most efficient and cost-effective way to alter another's future behavior to be less fitness cost-inflicting and more benefit-bestowing. Examples of moral influence include praise, bestowing rewards, rehabilitation, inflicting reputational damage, enlisting others for coordinated ostracism, and inflicting physical punishment.

The third moral component we propose, *moral conscience*, is a set of psychological mechanisms designed to *guide one's own behavior* toward others to avoid negative fitness consequences as a result of judgment and influence mechanisms in others. The emotion of anticipated guilt, for example, may function to deter the temptation to betray a friend to reap a short-term gain because of the long-term cost of a lost friendship. Taken together, these three psychological adaptations make up a tripartite theory of Machiavellian morality.

A key explanatory task is to identify classes of behaviors that humans moralize. The most obvious of these is intentionally exploitative behavior, such as stealing, which involves a gain for the perpetrator and a clear loss to the victim (Buss & Duntley, 2008). Behaviors intended to bestow a benefit on others, in contrast, would be judged as morally good. Examples include returning a lost wallet containing cash to its rightful owner or maintaining loyalty to a friend or mate when being disloyal might produce a temporary gain for the performer, but at a net long-term cost to the other.

The next explanatory step is to posit design features of the hypothesized adaptations, focusing on the inputs to the psychological mechanisms, the decision rules on which these mechanisms operate, and the behavioral or psychological output of the mechanisms. Certain inputs, decision rules, and outputs will be constant across the range of content domains and create the unity and consistency that researchers find in research on morality (DeScioli, Christner, & Kurzban, 2011). For example, the *intentions of the perpetrator* should influence moral judgments whether the violation was in the domain of physical harm, lying, or property damage (Keltikangas-Järvinen & Lindeman, 1997; Knobe, 2003; Nelson, 1980). However, other design features will be unique to solving a specific adaptive problem and add to the complexity and diversity of the moral content and behavior that we find across generations and cultures. Moral judgments about promiscuity should take certain inputs that are specific to sexuality, such as number of previous sexual partners, incidence of mate poaching, and incidence of infidelity. In contrast, moral judgments about disrespect would involve inputs such as existing hierarchical relationship and the size of the audience witnessing the disrespect.

The current chapter extends previous theorizing about morality by starting with an adaptive problem that the moral mechanisms evolved to solve and hypothesizing the design features necessary to solve it. The previous literature on morality lacks clear definitions of what constitutes morality due to the diversity of content areas (Krebs, 2011). Some have hypothesized distinct evolutionary origins of categories of moral content (e.g., purity, authority, harm), but doing so ignores the overwhelming similarities in decision rules across different content domains (DeScioli et al., 2011). Viewing the moral mechanisms as adaptations that function across content

areas by using content-specific mechanisms should help to identify the design features common across content areas and those that are unique to a specific adaptive problem. In turn, this form of analysis aids our understanding of both the universality and diversity of moral rules within and across cultures. Additionally, this approach may lead to more nuanced, novel, and testable predictions. By working through two examples, sexual infidelity and property theft, we will illustrate our tripartite framework of Machiavellian morality.

This approach refocuses research on morality to important content areas that have been neglected in the literature, such as moral rules governing the sexual domain. Additionally, this framework highlights the brighter side of moral influence in the form of gratitude, respect, praise, and rewards. Approaching the topic of morality from this evolutionary perspective may encourage new questions and new lines of research previously unexplored.

The Moral Mechanisms

The moral mechanisms comprise three distinct adaptations, each designed to solve broad classes of adaptive problems. Moral judgment, moral influence, and moral conscience are functionally distinct adaptations that operate across content areas. They encompass many subordinate mechanisms to solve the higher-order adaptive problems. These include categorizing a conspecific's behavior as intentionally cost-inflicting or benefit-bestowing, controlling or changing future behavior of conspecifics to minimize fitness costs and maximize benefits, and monitoring one's own behavior toward conspecifics to avoid condemnation, respectively.

Moral judgment comprises a set of evolved information-processing mechanisms that determines whether a moralized behavior has occurred, calculates a cost/benefit ratio of the behavior to self and to others, and stores that information for use in future social interactions. Specifically, moral judgment is designed for condemnation of cost-inflicting behavior and approbation of prosocial or benefit-bestowing behavior. Moral influence is activated after a moral judgment has been made, and avoidance of the cost-inflicting individual is unlikely. Moral influence mechanisms work to identify the most economical and efficient way of changing or controlling a conspecific's behavior for future interactions. That is, the mechanisms determine the best course of action given the specific circumstances of a moral violation to avoid future exploitative behavior from others and encourage prosocial behavior.

Moral conscience is a separate but related psychological adaptation designed to guide one's own benefit-bestowing or exploitative behavior. Importantly, moral judgment and moral influence are adaptations that focus on the decisions of conspecifics. Moral conscience, on the other hand, is concerned with behaviors of the self. Moral conscience therefore takes as input several internal regulatory variables (IRVs) and uses them to calculate the cost-benefit ratio of engaging in various forms of cost-inflicting or benefit-bestowing behavior (see Tooby, Cosmides, Sell, Lieberman, & Sznycer, 2008, for a fuller discussion of IRVs). Due to the crucial

distinctions between the moral mechanisms, the hypothesized inputs, decision rules, and outputs should diverge and converge in predictable ways. By analyzing each adaptation into its hypothesized design features, researchers can better predict and understand multidimensional phenomena such as morality.

Moral Judgment: Evolution and Design Features

In the ancestral environment of small-scale group living, individuals would have varied in the extent to which they engaged in exploitative strategies that intentionally imposed costs on others (Daly & Wilson, 1988). For example, some individuals would have been, on average, more selfish or more prone to exploitative aggression than others (Duntley & Buss, 2004). Risk of victimization by conspecifics was likely a frequent and recurrent problem in the small-scale societies (Chagnon, 1988). The ability to correctly identify and subsequently avoid these individuals as relationship partners would have afforded an evolutionary advantage. Importantly, moral judgment operates on a continuum ranging from morally evil to morally good. An underemphasized, but equally critical, function of moral judgment is to identify prosocial (i.e., intentionally benefit-bestowing) individuals and seek them out as future relationship partners. If certain individuals were consistently more likely to intentionally bestow benefits or suppress exploitative behaviors, the moral judgment mechanisms could motivate pursuit of these prosocial others as mates, friends, and cooperative partners.

If moral judgment is designed to evaluate a conspecific as either intentionally exploitative or prosocial, the inputs to the mechanisms should be factors of the individual and moralized behavior that shift the likelihood of costs and benefits to oneself, one's kin, and one's close social partners. The mechanism should first determine the likelihood that a given individual engaged in moralized behavior. The mechanisms should gather and review information through direct observation of the behavior, indirect information (e.g., gossip), and probabilistic cues to the moral violation. For example, in the absence of direct evidence, gossip about past moral behavior (e.g., moral reputation) can be used as an indicator of an exploitative or prosocial disposition (Dunbar, 1998).

Another input to the mechanism is whether the behaviors were intentional. If the morally good or bad behaviors were the result of incompetence, accident, or coercion, then these behaviors would not be indicative of the individual's intentions. Furthermore, if the individual's behavior was not intentional, then one could not reliably predict future exploitative or prosocial inclinations. Research has suggested that individuals who commit moral violations resulting from incompetence, chance, and strategy are categorized distinctly in the mind (Delton et al., 2012).

Finally, the key inputs include the overall probabilistic fitness costs and benefits to the self and to others with whom one's fitness is linked. The decision rules of the mechanisms should take the magnitude of the fitness costs (and benefits) imposed and the likelihood of receiving these fitness consequences by interacting with this individual.

The mechanisms then weigh these various inputs to determine how morally good or bad an individual is. The consequences of these weightings lead to a range of emotional, behavioral, and psychological outputs. Specifically, how morally good or evil a person is judged to be depends on the inputs: evidence of the behavior, intentions, and the likelihood of receiving fitness consequences as a result of the behavior. Direct evidence will be weighed more than indirect evidence; for example, catching someone inflicting harm will elicit a stronger reaction than hearing about it later from another person. The more likely a behavior was intentional, the more weight it will be given; for example, someone caught lying would be judged more harshly than someone omitting information because creating a lie is intentional, whereas omitting could be an artifact of forgetfulness. Finally, the magnitude of fitness consequences should shift moral judgment; for example, the act of murder imposes extreme fitness costs on an individual and should be more severely judged than less costly behaviors (Buss, 2006).

The outputs of the moral judgment mechanisms include activation of emotions (e.g., moral outrage, disgust, respect), thoughts (e.g., labeling the person who committed the behavior as morally good or evil), memory systems (e.g., storing the information about this person for future use), and behaviors (e.g., avoiding morally bad and pursuing morally good relationship partners). Although partner choice is a key output of moral judgment mechanisms, other behavioral and cognitive responses are notable. These responses include tracking social relationships based on history of exploitative or prosocial behaviors between third parties. Knowledge of who is an enemy or ally of whom would have been useful in close group living. Such knowledge would enable people to forge beneficial alliances, avoid costly ones, and predict side-taking in future conflicts. Thus, attention to the exploitative or prosocial interactions between third parties would have helped one navigate the complex and ever-changing network of alliances within one's social group. This could partially explain why people are fascinated by moral gossip surrounding unrelated third parties. Research has found that even infants as young as 3 months are able to use evidence of past moral behavior to avoid antisocial others (Hamlin, Wynn, & Bloom, 2010). Furthermore, infants are surprised when individuals choose to affiliate with a previously harmful character over a previously helpful one (Bloom, 2013).

The specific outputs of the moral judgment mechanisms depend on the circumstances surrounding the moralized behavior. A prosocial act, such as food sharing, will activate the emotions of respect and gratitude, will cause one to identify and remember this individual as a morally good person, and will motivate future interactions with this individual. These responses will interact in complex ways. The activation of emotions, in particular, may function to coordinate among the various cognitive, physiological, and behavioral responses of moral judgment mechanisms (Cosmides & Tooby, 2000). This could help explain why emotions are activated first, with other behavioral and psychological responses being activated later. Conscious moral deliberation is slow and unnecessary for the underlying than emotional reactions mechanisms to operate, giving the impression that rational rules matter much less (Haidt, 2001) (Fig. 1).

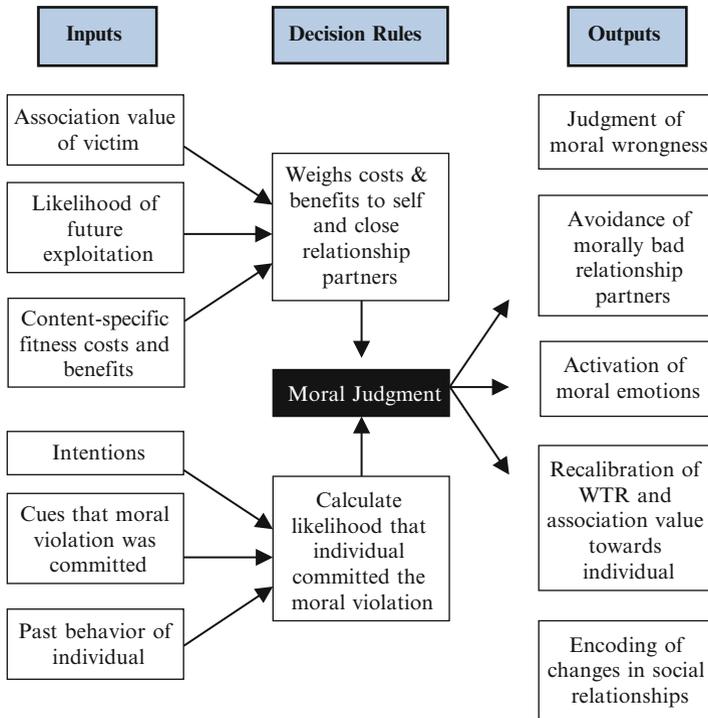


Fig. 1 Moral judgment

Moral Influence: Evolution and Design Features

Whereas moral judgment is concerned with tracking the likelihood of exploitation or prosociality by specific individuals and subsequent partner choice depending on those calculations, moral influence mechanisms function to *control and alter the behavior* of conspecifics to be less exploitative and more prosocial toward oneself and one’s close relationship partners (Frank, 1988). Given the ability to freely choose with whom one interacts, avoidance is a relatively low cost means of solving the adaptive problem of exploitation by others (Baumard, André, & Sperber, 2013). However, in the ancestral environment of small group living, complete avoidance may have been unlikely or prohibitively costly. For example, the exploitative individual would have an extended kin and social alliance network with whom one would likely interact, occasionally putting one in social contact with the exploitative individual. Furthermore, the likelihood that a morally bad group member could impose costs on one’s kin, mate, or friends would have increased the negative fitness consequences of allowing an exploitative individual’s behavior to remain unchecked. When repeated interaction with a cost-inflicting individual was likely, then the

moral influence mechanisms for deciding how best to control and alter future behavior should become activated. Additionally, even if effective future avoidance of exploitative individuals were common in the ancestral environment, allowing oneself to be exploited without retribution could create a reputation as exploitable. This would encourage others within one's social group to inflict fitness costs without fear of retaliation.

Regarding third-party moral punishment, if an individual within one's social group were exploitative toward others within the group, the likelihood of that individual eventually inflicting costs on oneself or one's kin, mates, and friends would often be nonzero. Frequent exploitative behavior toward other group members could be indicative of a disposition that systematically undervalues other people's welfare relative to the self's welfare. This baseline low welfare trade-off ratio (WTR) could be correlated with likelihood of aggressive or exploitative acts that harm one's inclusive fitness. Thus, when the costs of exploitation by an individual outweigh the costs of punishment, then third-party punishment could theoretically evolve. Engaging in third-party punishment could act as an honest signal of one's ability and willingness to retaliate against moral violations without incurring the costs of being a direct victim of exploitation. Additionally, garnering a reputation as someone who punishes wrongdoers could lead to indirect fitness benefits including those associated with reciprocal altruism (Trivers, 1971; Wright, 1995) and strengthening of alliances with victims and their extended social networks. Thus, although punishment should be less costly and less frequent for the exploitation of unrelated others, people should still be willing to engage in third-party punishment under specific conditions: when the likelihood of future exploitation by a wrongdoer is high and reputational benefits are probable. Although evidence of third-party punishment is abundant, recent work attempting to remove audience effects and experimenter demand has shown little or no third-party punishment (Pedersen, Kurzban, & McCullough, 2013). The evolution of mechanisms for moral influence would have presented an effective solution to the adaptive problem of exploitation in situations of repeat interaction and reputation concerns, which likely characterized the close group living of the ancestral environment during which moral influence mechanisms evolved.

Moral influence decision rules should take as input factors specific to the individual. These person variables will moderate the likelihood and magnitude of punishment or rewards. Examples include how valuable a relationship partner the individual is (i.e., the association value of the individual; see Petersen, Sell, Tooby, & Cosmides, 2012) and how physically or socially formidable the individual is (Sell, Tooby, & Cosmides, 2009). Generally, the higher the association value, the less likely one is to seek punishment for a moral offense and the less severe that punishment will be. Family, friends, mates, and irreplaceable association partners all have high association value (Cosmides & Tooby, 1992). Heavy punishments inflicted upon these individuals will have negative fitness consequences for oneself. One will also be less likely to severely punish individuals who are physically or socially formidable for fear of retaliation.

Another key individual variable will be the individual welfare trade-off ratio (WTR), that is, the ratio between how much the individual values another person's welfare relative to his own welfare (Sell et al., 2009). Individuals with a low WTR

value another's welfare much less than their own, indicating the likelihood of their employing an exploitative strategy. The same individual could have a low WTR toward one person and a high WTR toward another. Perhaps even more important than a single measure of WTR would be a measure of an individual's WTR to oneself relative to other people. Since reproductive success is defined relative to conspecifics, if someone uniquely values one's welfare, especially if he does not particularly care about others in the social group, he could be a useful relationship partner. An individual who is not prosocial toward everyone in the group equally, but who instead is exclusively prosocial to oneself, would be preferred to an individual who is indiscriminately prosocial (Lukaszewski & Roney, 2010). Such a finding would provide evidence against group selection since people who have a low WTR toward other group members and only value oneself will be chosen as mates, friends, and allies over group-benefiting others. It also provides evidence against people's everyday application of utilitarian ethics since nearly everyone in the group would be better off if all individuals were indiscriminately prosocial. However, this is not the ethical code most people prefer in others.

Circumstantial factors should also be taken into account, such as the likelihood of future interaction with this person, likelihood of repeat moral violations, and the likelihood of deterring others from committing a similar violation against oneself in the future. If the likelihood of future interaction with the exploitative individual is very low, then there is no point in wasting time, energy, or resources to alter his future behavior (see Krasnow, Cosmides, Pedersen, and Tooby 2012 for a full discussion). However, if the likelihood of future interaction is high, then the magnitude of both rewards for good behavior and punishment for bad behavior should be large.

Other key factors include victim characteristics, such as association value, vulnerability, age, and sex. For example, harsher punishments will be imposed on those who harm or exploit more vulnerable victims since this behavior is indicative of a marked lack of empathy or a heightened tendency toward exploitative strategies. Content-specific variables relevant for solving a particular adaptive problem will also be taken as input.

In support of these hypothesized inputs to moral influence mechanisms, recent research suggests that people selectively impose costly punishment on those with whom they expect to cooperate in the future (Krasnow et al., 2012). Some research suggests that association value of the perpetrator influences whether punishment or rehabilitation is endorsed by third parties judging moral dilemmas (Lieberman & Linke, 2007; Petersen et al., 2012).

The output of moral influence decision rules should show efficiency in controlling moralized behavior with the least amount of effort and cost to self. Potential outputs of the moral influence mechanism are direct bestowal of rewards, public praise, rehabilitation, coordinated ostracism, direct punishment, indirect punishment (e.g., recruiting other individuals to inflict the punishment) or some combination of these strategies. Additional information-processing mechanisms could be responsible for determining the magnitude of consequences depending on the endogenous factors (e.g., WTR of individual) and the exogenous factors (e.g., relationship between self and other) mentioned above. The probability and magnitude of moral consequences will be proportional to the probability and magnitude of fitness costs

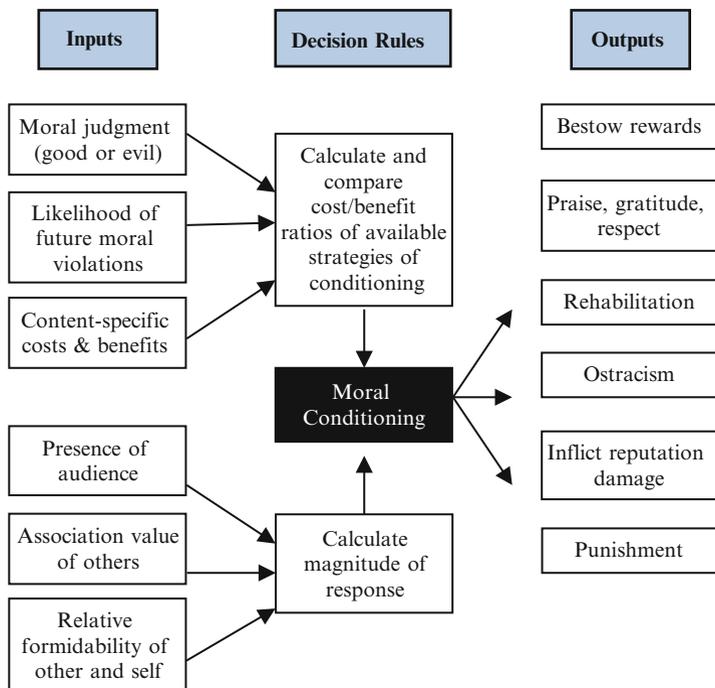


Fig. 2 Moral influence

and benefits to oneself or one's kin and close relationship partners. Behaviors that impose smaller fitness costs on oneself, such as intentional lying, will receive less punishment than relatively more costly violations, such as physical assault. Similarly, behaviors that provide larger benefits to oneself, such as saving one's life, will receive more reward than less beneficial behaviors, such as sharing food (Fig. 2).

Moral Conscience: Evolution and Design Features

Once the mechanisms for moral judgment and moral influence evolved, there would be selection pressure to regulate one's own behavior so as to strategically avoid reputation damage and other forms of punishment from others (DeScioli et al., 2011; Krebs, 2011) and to reap the rewards of building a good moral reputation (Sperber & Baumard, 2012). If moral conscience is a self-regulatory psychological adaptation, the features of the mechanism should be designed to economically and efficiently solve the adaptive problem of regulating one's behavior toward others to avoid negative moral judgment and influence. Specifically, moral conscience has two key functions: (1) to genuinely motivate one to resist costly exploitative

strategies in favor of benefit-bestowing behaviors that would have led to greater long-term fitness benefits and (2) to strategically engage in self-beneficial behaviors while publicly garnering a good moral reputation. Having a reputation as a morally good individual would lead to a variety of long-term fitness benefits, including a large alliance network, increased mate value, direct rewards, and indirect reciprocity (Sperber & Baumard, 2012).

Proactively, moral conscience serves to anticipate the costs and benefits associated with exploitative actions and those associated with prosocial actions to guide behavior toward the least costly option. To do so, the information-processing mechanisms would take as inputs memory of past consequences of morally stigmatized or sanctioned behavior, the likelihood of punishment or reward (e.g., formidability of victim and kin), and the likelihood of reputation damage or enhancement (e.g., presence of an audience). Recent research examining the effects of reputation on morality has found evidence of audience effects (Haley & Fessler, 2005; Kurzban, Descioli, & O'Brien, 2007; Powell, Roberts, & Nettle, 2012; Rigdon, Ishii, Watabe, & Kitayama, 2009).

The inputs are then sent to the decision rules that weigh the costs and benefits of engaging in a given behavior compared to alternatives. The output of the system should take the form of a net cost-benefit ratio that should motivate subsequent behavior. Specifically, if the net cost-benefit ratio is higher than other available strategies, the individual will be motivated to avoid engaging in the behavior. If the net cost-benefit ratio is lower than alternative strategies, the individual will be motivated to engage in the behavior. Even a costly exploitative behavior (e.g., theft) could be the best available strategy if the alternatives are relatively more costly (e.g., starvation). Importantly, moral conscience is designed to forego short-term gains via exploitation in favor of the long-term benefits of prosociality. Therefore, if one does not expect to survive long enough to reap the long-term benefits, then the best strategy would be to engage in exploitation and receive immediate benefits. This could help explain the correlation between future discounting and various indices of criminal behavior. Researchers are increasingly recognizing the importance of considering criminal behavior as evolved strategies for acquiring resources, status, and mates under certain circumstances (Buss, 2012; Duntley & Buss, 2010; Durrant & Ward, 2012).

Retroactively, moral conscience performs two key functions. One mechanism is designed to do strategic damage control after a moral violation has been performed. A second is designed to encode the negative or positive consequences of moralized behavior to prevent oneself from engaging in costly moral violations in the future. The strategic damage control mechanisms should take as input cues to the likelihood of punishment or reputation damage (e.g., presence or absence of an audience, formidability), the magnitude of fitness costs inflicted on the victim, and characteristics of the victim that could influence others' moral judgments (e.g., vulnerability, age, and sex of victim).

These mechanisms should then weigh the different inputs to determine the best course of action to minimize the costs of having already engaged in a morally impermissible behavior. A key function of moral conscience is to engage in strategic

multi-person games, in which the best course of action is dependent on the behavior of others. For example, if many other people know about a moral transgression, the system should motivate expressions of guilt and remorse, apologizing to the victim and the victim's kin and making reparations depending on the self-assessed severity of the violation. If few people are aware of the transgression, the best strategy may be to deny the transgression altogether, make plausible excuses for the immoral behavior (e.g., blame the victim or circumstances surrounding the transgression), hide evidence of wrongdoing, and recruit kin, mates, and allies to defend oneself against retaliation. Individual difference variables such as formidability, and contexts such as need, should lead to different strategies of strategic damage control.

The encoding mechanisms are responsible for gathering information about the fitness consequences of engaging in a moralized behavior. Benefit-bestowing behavior could lead to a variety of positive consequences, including increased mate value, reputation enhancement, strengthening of existing alliances, and formation of new alliances. Conversely, morally impermissible or fitness cost-inflicting behavior could lead to numerous negative fitness consequences, including physical, economical, or emotional retaliation by the victim or the victim's kin, reputation damage, ostracism, loss of close relationships, and creation of enemies. The retroactive mechanisms responsible for encoding moral consequences should search for information about the positive and negative consequences of engaging in moralized behavior and store that information in memory systems.

Since moral conscience is a self-control mechanism, there is no reason why the mechanism should make the Machiavellian decision rules available to consciousness. All that is needed to motivate individuals to suppress their immediate immoral urges to avoid condemnation and seek rewards is positive feedback when committing morally good acts and negative feedback when committing morally bad acts. In much the same way that humans derive pleasure from food and sex without an explicit understanding of the link between food and differential survival or sex and differential reproductive success, moral behavior could lead to positive emotions and cognitions in the absence of conscious awareness of the decision rules underlying moral conscience. The proximate phenomenology of engaging in morally good behavior would include positive emotions and cognitions to motivate such prosocial acts, although the function is selfish in the ultimate sense (Krebs, 2005) (Fig. 3).

Sexual Infidelity

The sexual domain is one of the most highly moralized content domains (Buss & Asao, 2013). People make moral judgments about which sexual acts are permissible or impermissible, who can perform those acts with whom, when and where those acts can take place, and why others decide to engage in sex acts (Shweder et al., 1997). This is because the consequences of other people's sexual behavior historically had dramatic consequences for one's own reproductive success (Symons, 1979).

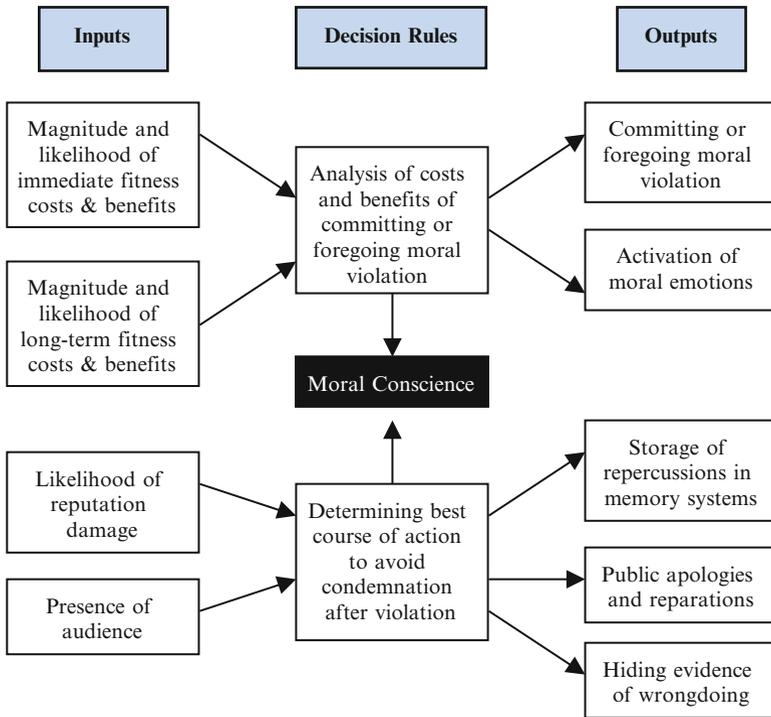


Fig. 3 Moral conscience

One commonly moralized sexual behavior is infidelity, which provides a useful illustration of the heuristic value of the proposed tripartite framework of Machiavellian morality. Sexual infidelity committed by one’s partner is a particularly costly behavior for both men and women. Sexual infidelity could lead to a variety of costs to one’s reproductive fitness, from termination of the existing relationship and all economic, social, and sexual benefits therein to sexually transmitted disease (Buss, 2000). Men face an additional cost of cuckoldry in the case of insemination by a rival male, whereas women face the diversion of partner’s investment to a rival female (Symons, 1979). Sexual infidelity by one’s partner is, therefore, an extremely cost-inflicting behavior, and changing the future likelihood of this behavior constitutes a recurrent adaptive problem that would have required a solution.

Moral judgments are often automatic, spontaneous, and associated with strong emotions (Haidt, 2001). However, the mechanisms responsible for the activation of these physiological, psychological, and behavioral responses need to be examined in greater detail. The inputs to the moral judgment mechanism would include (1) information that the infidelity occurred (e.g., proof of infidelity directly, indirect knowledge via gossip, or probabilistic cues to infidelity) (Shackelford & Buss, 1997), (2) potential costs associated with sexual infidelity (e.g., termination of a beneficial relationship, loss of sexual access to mate, loss of resources provided

by mate, reputation damage, sexually transmitted diseases, and cuckoldry), and (3) potential benefits associated with infidelity (e.g., termination of a detrimental relationship).

The decision rules would then weigh the evidence to determine the likelihood that the conspecific performed the moralized behavior and generate a net cost-benefit ratio of the behavior. For infidelity, the decision rules would calculate the likelihood that one's partner engaged in sexual infidelity and would calculate the large net cost associated with such infidelity. The output of the moral judgment mechanism could be emotional (e.g., moral outrage, disgust, or feeling betrayed), rational (e.g., identifying the cheater as a bad relationship partner), involve the memory system (e.g., encoding and storing the information about the infidelity, such as the identity of the mate poacher and the cues associated with discovery of the infidelity), or some combination of these outputs.

Moral influence mechanisms would then be activated to determine the best course of action for avoiding such large fitness costs in the future. One strategy would be to terminate the relationship, ensuring no possibility of sexual infidelity by this partner in the future. Another option would be to remain in the relationship but inflict some form of direct punishment on one's partner in the form of imposing physical or emotional costs or withdrawing benefits, such as withholding sexual access or resources (Buss & Duntley, 2011). This punishment would make sexual infidelity a prohibitively costly behavior, effectively preventing future instances of partner infidelity. These behavioral responses only represent two choices in a large array of options including indirect punishment (e.g., enlisting family members to inflict costs on an unfaithful partner), reputation damage (e.g., gossip to lower a partner's mate value and the likelihood of attracting future affair partners), and retaliatory affairs (Buss & Shackelford, 1997).

The influence strategy that an individual pursues depends on various inputs to the information-processing mechanisms. Some of these inputs will be unique to the adaptive problem of sexual infidelity, such as the relative mate value of self and the unfaithful partner and the likelihood of finding another sexual partner of equal or greater mate value in the future. Other inputs will be general to moral influence mechanisms across adaptive problem and content domain, such as the value and irreplaceability of the individual (i.e., association value), relative physical and social formidability of self and others, formidability of one's kin and alliance network, and likelihood of future moral violations by the individual. Empirically, one of the key motivators of intimate partner violence is suspicion or discovery of a sexual affair (Buss & Duntley, 2011; Daly & Wilson, 1988).

Unlike moral influence, moral conscience is independent of moral judgment. The distinction between moral judgment and moral conscience could help to explain moral hypocrisy, in general, and sexual double standards, in particular. Since moral conscience is designed to control and alter one's own behavior in response to past or probabilistic negative consequences of immoral actions, the inputs to the mechanism are orthogonal to the inputs to moral judgment mechanisms. In the case of one's own sexual infidelity, if either the benefits of sexual infidelity outweigh the costs or the chances of discovery are sufficiently low, then the mechanisms underlying moral

conscience could motivate an individual to engage in a clandestine affair. The conscience mechanism would take as input the likelihood that the sexual affair will be discovered by one's partner (e.g., suspiciousness of current partner, ease of coordinating or concealing the affair), the costs associated with discovery (e.g., loss of current relationship partner, reputation damage, risk of emotional or physical retaliation), and the benefits associated with engaging in sexual infidelity (e.g., mate switching to a higher-quality mate).

The decision rules would then weigh the positive and negative consequences of infidelity by their probabilities and generate a cost-benefit ratio. If the ratio is higher than alternative courses of action, then the mechanism could activate systems to avoid engaging in infidelity. If the ratio is lower than alternatives, then the mechanism could activate systems to motivate the affair, but only if one can conceal the affair with minimal costs to self, including avoiding the large cost of losing one's current partner.

Importantly, there are some circumstances in which the moral conscience mechanisms could motivate one's own sexual affair, while one's moral judgment mechanisms could judge another's sexual affair as morally wrong and worthy of punishment. This hypocrisy is possible due to the separate nature of the moral mechanisms. Specifically, the mechanisms that guide one's own behavior (i.e., moral conscience) are not necessarily those involved in judging other people's behavior (i.e., moral judgment). To the extent that the sexes differ in the costs and benefits of engaging in a given moralized behavior, the moral judgment and moral conscience mechanisms in men and women may provide diverging judgments of moral wrongness and deserved punishment of behaviors such as own and partner's infidelities (Buss, 2009).

In short, sexual infidelity illustrates the heuristic value of distinguishing among moral judgment, influence, and conscience. Additionally, highlighting the importance of sexual morality opens new lines of research. The next section applies our Machiavellian tripartite scheme to the domain of theft. Although theft is as ubiquitous as sexual infidelity, the two behaviors have little else in common. We chose such a divergent content area to examine which design features of the moral mechanisms operate across content areas and which are unique to the sexual domain.

Property Theft

Another problem that has been recurrently faced by humans over deep evolutionary time is the taking of one's personal belongings by conspecifics (Buss & Duntley, 2008; Duntley & Shackelford, 2008). Historically, personal property has taken numerous forms, from physical resources (e.g., animals, food, and modern currency) to intellectual property (e.g., ideas and written works). The negative ramifications of personal property theft are threefold: the actual loss of access to a valued personal item, the relative fitness costs associated with a rival gaining access to the valued item, and increased perceived exploitability due to the theft. Since these costs are

large, theft constitutes an adaptive problem of controlling or changing a conspecific's behavior. Theft, therefore, falls within the purview of the moral mechanisms.

To determine whether property theft actually occurred, moral judgment mechanisms will take as input certain information that is specific to the adaptive problem of theft prevention, such as local norms concerning ownership of property (e.g., documents or social rules that define what constitutes personal property and ownership in the culture) and information about previous and current possession of the property (e.g., who was first in possession, was the property lost or out of possession before it was taken by another). Other inputs will be the same across all contexts, such as whether the conspecific stole the property intentionally or as a result of accident, incompetence, or coercion. Similar to the inputs for sexual infidelity, moral judgment mechanisms will also take in information about the costs and benefits associated with having property stolen (e.g., how valuable and irreplaceable the property was and the probability of increased future exploitation by others) to calculate how morally wrong the violation was based on the costs and benefits to the burglarized individual. The higher the probability that a theft actually occurred and the more valuable or irreplaceable the item stolen is, the more morally wrong the thief will be judged. Other output of the moral judgment system includes activation of the moral emotions, and memory and avoidance of known thieves.

Once a moral judgment has been made, the moral influence mechanisms are activated such that the output of moral judgment is one input into the decision rules governing the type and magnitude of influence used to alter the behavior of others. In the case of an exploitative behavior, such as property theft, the influence would occur in the form of moralistic punishment. General inputs into the moral influence mechanism will include those mentioned previously for sexual infidelity (e.g., association value of the individual, relative formidability of self versus individual, and likelihood of future interactions with the individual). There will also be theft-specific inputs, such as whether the stolen property can be easily returned without damage or devaluation. The influence mechanism will then conduct a cost-benefit analysis for each potential punishment strategy to find the most efficient and economical means of deterring future theft either from the individual or from other conspecifics. If the stolen item is minimally valuable or one that can be easily returned without devaluation, the least costly strategy would be to simply ask the thief to return the property. However, if the piece of property were highly valuable or irreplaceable, if the item cannot be returned without devaluation, or if the theft increases one's reputation for exploitability (see Buss & Duntley, 2008), then the punishment should be more severe. Costly forms of punishment, such as direct physical, financial, or emotional punishment, could be worthwhile strategies if the long-term benefits of deterring future property theft by others outweigh the immediate costs of engaging in punishment.

The moral conscience mechanisms function to proactively deter theft, a costly strategy of resource acquisition, unless either the chances of detection are low or the benefits outweigh the costs weighed by their respective probabilities. For example, if an individual is on the brink of starvation and has no other feasible means of securing food resources, then food theft could be a useful strategy. In this case, the

benefits of immediate caloric intake necessary to sustain life outweigh the costs of potential detection and punishment. If the individual engages in property theft, the strategic damage control mechanisms should activate behaviors that decrease the likelihood of detection and punishment, such as leaving the site of the theft, avoiding the victim of theft, and denying accusations of theft unless the likelihood of detection is overwhelmingly high. If discovery of theft is inevitable, a different set of emotional, psychological, and behavioral mechanisms should be activated, including genuine feelings and confessions of guilt and remorse to appease the victim and victim's kin, verbal attempts to exculpate oneself from blame, and the return of or reparations for stolen property. The negative consequences associated with committing property theft should be encoded into memory systems to deter use of costly resource acquisition strategies in the future. The importance of circumstances in determining the output of the moral conscience mechanisms highlights the role of context in certain content domains. Some moralized behaviors are almost always condemnable when directed toward one's in-group members, for example, rape or torture. Other content areas, such as theft, may elicit moral decisions concerning wrongness, deserved punishment, and permissibility that are more context-dependent. Research currently being conducted suggests that violations involving theft may be more amenable to shifts in judgments of moral wrongness and deserved punishment than violations involving rape (Asao & Buss, 2014).

The examples of sexual infidelity and property theft illustrate the usefulness of the tripartite framework of Machiavellian morality. While some aspects of the moral mechanisms operate across content areas, other design features are content specific. By treating moral judgment, influence, and conscience as three related but separate suites of mechanisms, more nuanced predictions about the design features of those evolved psychological adaptations can be made. Specifically, the inputs, decision rules, and outputs can be examined in greater detail. This level of specificity is particularly helpful when dealing with a multidimensional and complex topic such as morality. This approach could help better define the borders between the different moral mechanisms and avoid the ambiguity, confusion, and polarization that have characterized research on morality in the past.

Ambiguity in Morality Research

Everyday discussions of morality are as ambiguous as they are ubiquitous. People show intense interest in the morality of other people. Gossip concerning the moral behaviors of others is commonplace. The media is dominated by information about the moral decisions of others. This is apparent from headline news stories such as the U.S. drone bombings in Yemen to small town reports of local heroes. Despite the widespread interest in morality, there are currently no agreed-upon definitions of what constitutes morality or the moral domain, nor is there even a consensus on whether such definitions would prove helpful when discussing morality.

Part of the conceptual confusion lies in the fact that the term “morality” is a broad term applied to at least three distinct adaptations designed to solve related but distinct adaptive problems. Some researchers have treated morality as synonymous with altruism (Wilson, 2012). However, altruistic behavior is only one subset of behaviors driven by moral conscience mechanisms. Although altruism is a theoretically important class of behaviors that can be partially explained by reference to moral conscience mechanisms, it is not the only behavioral output of those mechanisms. The function of moral conscience is to motivate people to strategically forego immediate fitness gains from moral wrongdoing in favor of delayed gains associated with maintaining a positive moral reputation and avoiding punishment. If correct, much of the function of moral conscience is to suppress exploitative tendencies out of concern for reputation and punishment, instead of promoting prosociality or group cohesion. A major component of moral conscience mechanisms will be to determine the circumstances in which one can effectively engage in exploitation without fear of negative fitness consequences. Thus, attempts to hide moral wrongdoing are important outputs of the moral conscience mechanisms. Shedding a light on the darker aspects of people’s moral conscience mechanisms may change the way scientists think about moral conscience.

Another complication with studying morality is that “morality” is an umbrella term that covers a wide spectrum of content domains. Moralized content areas include sexual activity, food taboos, physical harm, theft, property rights, cultural norms, and fairness. This diversity is further confounded because certain behaviors are moralized to solve a specific evolutionarily relevant problem and are nearly universally condemned (e.g., cheating in social exchanges, theft, and murder); however, moralization of other behaviors lacks grounding in solving an adaptive problem and is highly variable across culture and time (e.g., homosexuality and cultural norms). The term “morality” is applied equally to these two quite different categories of behavior. The approach advocated here (i.e., starting with conspecifics’ behaviors that would have constituted an adaptive problem for the individual that necessitated solving) leads to investigations of the former category of moralized behavior. However, once the moral mechanisms evolved to solve adaptive problems associated with avoiding and controlling cost-inflicting people, those mechanisms could have been used to moralize behaviors that were undesirable for reasons other than probabilistic fitness costs to promote one’s religious, social, or political ideology (Rozin, Markwith, & Stoess, 1997). Additionally, public moralization of behaviors could advertise one’s own moral goodness or to coordinate third-party side-taking during conflicts (DeScioli & Kurzban, 2013; Petersen, 2013).

Despite the diversity of moral content, there is evidence that people use similar underlying principles when making moral evaluations across contexts. The omission effect, in which actions are consistently judged more morally wrong than equivalent omissions (e.g., a lie is worse than an omission of the truth), is one such principle that operates across content areas (Cushman, Young & Hauser, 2006). Another is the doctrine of double effect which states that moral violations done to achieve another goal are less morally wrong than the same violation committed as a means to an end (Descioli, Asao, & Kurzban, 2012). Additionally, there is considerable agreement in

the relative ranking of moral violations across individuals and cultures, indicating a basic level of agreement in criteria used to assign severity of moral violations (e.g., Mikhail, 2007; Robinson, Kurzban, & Jones, 2007). Given the consistency in moral information processing, it is unlikely that moralization of different content domains evolved independently of one another. These surprising findings could be explained if each moral mechanism (moral judgment, punishment, and conscience) constitutes a unified set of adaptations that operate across distinct content areas by taking both content-general and content-specific factors as input, performing cost-benefit analyses, and coordinating behavioral, physiological, and psychological outputs to evaluate and control the behavior of a conspecific or oneself.

Discussion and Future Directions

The current framework builds on the extant body of research by integrating ideas from different models of morality and expanding these models to generate novel hypotheses about unexplored content areas. Baumard, André, and Sperber (2013) have emphasized the important role that partner choice plays in the evolution of fairness. We argue that evaluating potential relationship partners is the ultimate goal of moral judgment. We extend the idea of partner choice and switching into previously unexplored domains of morality, such as sexuality and property rights. Furthermore, we highlight the equal importance of moralistic rewards and punishment in shaping the evolution of moral norms. Partner choice was likely limited in the close group living that characterized ancestral hunter gatherers. Repeated exposure to most members of the group, lack of new available partners, and coercion could have made avoidance of exploitative individuals difficult. Therefore, the mechanisms for enacting moral rewards and punishments evolved to alter the behavior of others within one's social circle.

DeScioli and Kurzban (2009) raised crucial questions about the existence of third-party moral judgment and moralistic punishment. Additionally, they were the first to posit that moral conscience mechanisms are designed to avoid condemnation from others. Building on this work, the current framework offers a related, but different, account of the evolution of moral judgment and moralistic punishment. Furthermore, the model aspires to encompass the full spectrum of moral influence behaviors. In addition to moralistic punishment, the model sheds light on the equally puzzling phenomena of why humans praise, reward, and seek out as relationship partners those who help others. The media is filled with stories of everyday heroes, and people experience more positive emotions than negative emotions in everyday life (Algoe & Haidt, 2009). However, these positive aspects of morality and moral emotions are often overlooked in psychology. By emphasizing these positive aspects of moral influence, we hope to provide a more complete picture of morality.

The tripartite scheme of Machiavellian morality also recognizes the importance of perspectival shifts in moral decision-making (Duntley & Buss, 2004). There will be predictable differences in moral judgments depending on the

perspective of the evaluator. For instance, the same moral violation will elicit harsh negative judgments and punishment from the victim's close relationship partners, but may receive less harsh judgments from unrelated strangers or enemies. Furthermore, since moral judgment mechanisms are hypothesized to operate independently from moral conscience mechanisms, we expect that a moral violation committed by oneself will not be judged the same as a similar moral violation committed by another. Many moral double standards are accountable by the Machiavellian approach to morality. Thus, not all victims or perpetrators are created equal. A deeper understanding of the moral information-processing mechanisms will help to uncover potentially harmful biases in people's moral intuitions and behaviors. This information could then be used to guide public policy concerning morality to create more impartial and egalitarian policies.

Avoiding the Naturalistic Fallacy in Morality Research

The naturalistic fallacy is the logical error of assuming that because something is "natural" it must be morally right. We emphasize here that uncovering the evolutionary origins of people's moral intuitions and biases **does not** equate to condoning or endorsing those intuitions and biases. Importantly, this evolutionary psychological approach to morality, while theoretically useful, does not attempt to uncover what is objectively morally good or bad. The approach outlined here is concerned with understanding information-processing mechanisms housed in the brain and is agnostic about objective moral truth. Since our moral intuitions are shaped by evolution by natural selection, they were ultimately designed to increase reproductive success in ancestral environments, not to search for moral truths. The extent to which people's moral intuitions coincide with what any given philosophical perspective deems morally right is incidental.

Conclusion

Outside of the laboratory setting, questions of morality are rarely as simple and straightforward as the dilemmas commonly used to uncover people's moral reasoning. People infrequently conform neatly to a specific philosophical moral doctrine, such as utilitarianism, and instead use their biased, imperfect, and complex intuitions to guide their moral judgments and behaviors. Emotions are activated to coordinate among the various, competing subordinate mechanisms, while cost-benefit analyses are performed to determine the best courses of action. Prior relationships, reputation concerns, and formidability influence moral decision-making, despite people's desire for a justice system that is blind to those factors.

Further research using this tripartite evolutionary framework can help to recognize the biases in moral thinking. For example, people may be less impartial or egalitarian than previously believed. This approach aims to explore the role of social

relationships, formidability, physical attraction, and welfare trade-off ratios on moral decision-making. This knowledge can then be used to inform public policy, especially when our intuitions and behaviors do not coincide with what we deem objectively morally right.

Finally, the Machiavellian tripartite schema outlined above can help to organize the important advances that evolutionary psychologists have made in understanding the complexities of human morality. Moral judgment, influence, and conscience are adaptations designed to solve adaptive problems that span content areas. Starting from the adaptive problems that other people's behaviors pose, researchers can tease apart the content-specific and global design features of the moral mechanisms. This framework attempts to appreciate both the consistency and diversity within the moral domain.

References

- Algoe, S. B., & Haidt, J. (2009). Witnessing excellence in action: The "other-praising" emotions of elevation, gratitude, and admiration. *The Journal of Positive Psychology, 4*(2), 105–127.
- Asao, K., & Buss, D. M. (2014, June). *When justice isn't blind*. Poster presentation at CEU Summer University program on Morality: Evolutionary Origins and Cognitive Mechanisms, Budapest, Hungary.
- Baumard, N., André, J.-B., & Sperber, D. (2013). A mutualistic approach to morality: The evolution of fairness by partner choice. *The Behavioral and Brain Sciences, 36*(1), 59–78.
- Bloom, P. (2013). *Just babies: The origins of good and evil*. New York: Random House LLC.
- Brosnan, S. F., & de Waal, F. D. M. (2003). Monkeys reject unequal pay. *Nature, 425*, 297–299.
- Buss, D. M. (2009). *Sexual double standards*. Talk presented at annual meeting of the Human Behavior and Evolution Society.
- Buss, D. M. (2000). *The dangerous passion: Why jealousy is as necessary as love and sex*. New York: Simon and Schuster.
- Buss, D. M. (2006). *The murderer next door: Why the mind is designed to kill*. New York: Penguin Group.
- Buss, D. M. (2012). The evolutionary psychology of crime. *Journal of Theoretical and Philosophical Criminology, 1*(1), 90–98.
- Buss, D. M., & Asao, K. (2013, March). *The evolution of sexual morality*. Talk presented at Evolution of Morality conference at Oakland University, Rochester, MI.
- Buss, D. M., & Duntley, J. D. (2008). Adaptations for exploitation. *Group Dynamics: Theory, Research, and Practice, 12*(1), 53.
- Buss, D. M., & Duntley, J. D. (2011). The evolution of intimate partner violence. *Aggression and Violent Behavior, 16*(5), 411–419.
- Buss, D. M., & Shackelford, T. K. (1997). From vigilance to violence: Mate retention tactics in married couples. *Journal of Personality and Social Psychology, 72*(2), 346.
- Chagnon, N. (1988). Life histories, blood revenge, and warfare in a tribal population. *Science, 239*(4843), 985–992.
- Cosmides, L., & Tooby, J. (1992). Cognitive adaptations for social exchange. In J. Barkow, L. Cosmides, & J. Tooby (Eds.), *The adapted mind*. New York: Oxford University Press.
- Cosmides, L., & Tooby, J. (2000). Evolutionary psychology and the emotions. In M. Lewis & J. M. Haviland-Jones (Eds.), *Handbook of emotions* (2nd ed., pp. 91–115). New York: Guilford.
- Cushman, F., Young, L., & Hauser, M. (2006). The role of conscious reasoning and intuition in moral judgment: Testing three principles of harm. *Psychological Science, 17*(12), 1082–1089.
- Daly, M., & Wilson, M. (1988). *Homicide*. Hawthorne, New York: Aldine.

- Delton, A., Cosmides, L., Guermo, M., Robertson, T. E., & Tooby, J. (2012). The psychosemantics of free riding: dissecting the Architecture of a Moral Concept. *Journal of Personality and Social Psychology*, *102*(6), 1252–1270.
- Descioli, P., Asao, K., & Kurzban, R. (2012). Omissions and byproducts across moral domains. *PLoS One*, *7*(10), e46963.
- DeScioli, P., Christner, J., & Kurzban, R. (2011). The omission strategy. *Psychological Science*, *22*(4), 442–446.
- DeScioli, P., & Kurzban, R. (2009). Mysteries of morality. *Cognition*, *112*(2), 281–299.
- DeScioli, P., & Kurzban, R. (2013). A solution to the mysteries of morality. *Psychological Bulletin*, *139*(2), 477–496.
- Dunbar, R. I. M. (1998). *Grooming, gossip, and the evolution of language*. Cambridge, MA: Harvard University Press.
- Dunbar, R. I. M. (2004). Gossip in evolutionary perspective. *Review of General Psychology*, *8*(2), 100–110.
- Duntley, J., & Buss, D. M. (2004). The evolution of evil. In A. G. Miller (Eds.), *The social psychology of good and evil* (pp. 102–123). New York: Guilford.
- Duntley, J. D., & Buss, D. M. (2010). The evolution of stalking. *Sex Roles*. doi:[10.1007/s11199-010-9832-0](https://doi.org/10.1007/s11199-010-9832-0).
- Duntley, J. D., & Shackelford, T. K. (2008). Darwinian foundations of crime and law. *Aggression and Violent Behavior*, *13*(5), 373–382.
- Durrant, R., & Ward, T. (2012). The role of evolutionary explanations in criminology. *Journal of Theoretical and Philosophical Criminology*, *4*(1), 1–37.
- Frank, R. H. (1988). *Passions within reason: The strategic role of the emotions*. New York: WW Norton & Co.
- Gintis, H., Bowles, S., Boyd, R., & Fehr, E. (2003). Explaining altruistic behavior in humans. *Evolution and Human Behavior*, *24*(3), 153–172.
- Graham, J., Haidt, J., Koleva, S., Motyl, M., Iyer, R., Wojcik, S., & Ditto, P. H. (2013). Moral foundations theory: The pragmatic validity of moral pluralism. *Advances in Experimental Social Psychology*, *47*, 55–130.
- Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review*, *108*(4), 814.
- Haidt, J., & Joseph, C. (2004). Intuitive ethics: How innately prepared intuitions generate culturally variable virtues. *Daedalus*, *133*(4), 55–66.
- Haley, K. J., & Fessler, D. M. T. (2005). Nobody's watching? *Evolution and Human Behavior*, *26*(3), 245–256.
- Hamlin, J. K., Wynn, K., & Bloom, P. (2010). Three-month-olds show a negativity bias in their social evaluations. *Developmental Science*, *13*(6), 923–929.
- Keltikangas-Järvinen, L., & Lindeman, M. (1997). Evaluation of theft, lying, and fighting in adolescence. *Journal of Youth and Adolescence*, *26*(4), 467–483.
- Knobe, J. (2003). Intentional action in folk psychology: An experimental investigation. *Philosophical Psychology*, *16*(2), 309–324.
- Kohlberg, L., & Hersh, R. H. (1977). Moral development: A review of the theory. *Theory Into Practice*, *16*(2), 53–59.
- Krasnow, M., Cosmides, L., Pedersen, E. J., & Tooby, J. (2012). What are punishment and reputation for? *PLoS One*, *7*(9), e45662.
- Krebs, D. (2005). The evolution of morality. In D. M. Buss (Ed.), *The handbook of evolutionary psychology*. New Jersey: John Wiley & Sons.
- Krebs, D. (2011). *The origins of morality: An evolutionary account*. Retrieved from <http://www.ebilib.com>.
- Kurzban, R., Descioli, P., & O'Brien, E. (2007). Audience effects on moralistic punishment. *Evolution and Human Behavior*, *28*(2), 75–84.
- Lieberman, D., & Linke, L. (2007). The effect of social category on third party punishment. *Evolutionary Psychology*, *5*(2), 289–305.

- Lukaszewski, A. W., & Roney, J. R. (2010). Kind toward whom? Mate preferences for personality traits are target specific. *Evolution and Human Behavior, 31*(1), 29–38.
- Mikhail, J. (2007). Universal moral grammar: Theory, evidence and the future. *Trends in Cognitive Sciences, 11*(4), 143–152.
- Nelson, S. A. (1980). Factors influencing young children's use of motives and outcomes as moral criteria. *Child Development, 51*(3), 823.
- Pedersen, E. J., Kurzban, R., & McCullough, M. E. (2013). Do humans really punish altruistically? A closer look. *Proceedings of the Royal Society B: Biological Sciences, 280*(1758), 20122723.
- Petersen, M. B. (2013). Moralization as protection against exploitation: Do individuals without allies moralize more? *Evolution and Human Behavior, 34*, 78–85.
- Petersen, M. B., Sell, A., Tooby, J., & Cosmides, L. (2012). To punish or repair? Evolutionary psychology and lay intuitions about modern criminal justice. *Evolution and Human Behavior, 33*(6), 682–695.
- Powell, K. L., Roberts, G., & Nettle, D. (2012). Eye images increase charitable donations: Evidence from an opportunistic field experiment in a supermarket. *Ethology, 118*(11), 1096–1101.
- Range, F., Horn, L., Viranyi, Z., & Huber, L. (2009). The absence of reward induces inequity aversion in dogs. *Proceedings of the National Academy of Sciences of the United States of America, 106*(1), 340–345.
- Rigdon, M., Ishii, K., Watabe, M., & Kitayama, S. (2009). Minimal social cues in the dictator game. *Journal of Economic Psychology, 30*(3), 358–367.
- Robinson, P., Kurzban, R., & Jones, O. D. (2007). The origins of shared intuitions of justice. *Vanderbilt Law Review, 60*(6), 1633–1688.
- Rozin, P., Markwith, M., & Stoess, C. (1997). Moralization and becoming a vegetarian: The transformation of preferences into values and the recruitment of disgust. *Psychological Science, 8*(2), 67–74.
- Sell, A., Tooby, J., & Cosmides, L. (2009). Formidability and the logic of human anger. *Proceedings of the National Academy of Sciences of the United States of America, 106*(35), 15073–15078.
- Shackelford, T. K., & Buss, D. M. (1997). Cues to infidelity. *Personality and Social Psychology Bulletin, 23*(10), 1034–1045.
- Shweder, R. A., Much, N. C., Mahapatra, M., & Park, L. (1997). The “Big Three” of morality (autonomy, community, and divinity), and the “Big Three” explanations of suffering. In A. M. Brandt & P. Rozin (Eds.), *Morality and health* (pp. 119–169). New York, NY: Routledge.
- Sperber, D., & Baumard, N. (2012). Moral reputation: An evolutionary and cognitive perspective. *Mind & Language, 27*(5), 495–518.
- Stich, S. (2006). Is morality an elegant machine or a kludge? *Journal of Cognition and Culture, 6*(1), 181–189.
- Symons, D. (1979). *The evolution of human sexuality*. New York: Oxford University Press.
- Tomasello, M., & Vaish, A. (2013). Origins of human cooperation and morality. *Annual Review of Psychology, 64*, 231–255.
- Tooby, J., Cosmides, L., Sell, A., Lieberman, D., & Sznycer, D. (2008). Internal regulatory variables and the design of human motivation: A computational and evolutionary approach. In *Handbook of approach and avoidance motivation* (Vol. 15, p. 251).
- Trivers, R. (1971). The evolution of reciprocal altruism. *Quarterly Review of Biology, 46*(1), 35–57.
- Wilson, E. O. (2012). *The social conquest of earth*. New York: WW Norton & Company.
- Wright, R. (1995). *The moral animal: The new science of evolutionary psychology*. New York: Vintage Books.