Moral Alliance Strategies Theory

Jesse Marczyk

Published online: 4 March 2015
© Springer International Publishing 2015

Abstract The current paper puts forth a new conceptualization for the evolutionary function of morality. Moral alliance strategies theory proposes that our moral psychology can be understood as a tool for managing associations, using the observable behavior of others as an input. The behavior of others is a key input for this system, as it can be used to identify new and profitable alliance potential, to predict where other third parties might side in disputes and as a cue for when existing alliances are liable to become too costly to maintain. Through considerations of how people's behaviors have effects on their association value, a number of facets of our moral psychology—such as impartiality, proportionality, and moral praiseworthiness—can be better understood.

Keywords Alliance · Altruism · Morality · Punishment · Praise · Welfare

People often take sides in disputes that did not directly involve them for moral reasons (i.e., because someone behaved “immorally” or to defend an individual against claims of wrongdoing). This involvement brings with it the potential for incurring dispute-related costs, such as retaliation from condemned parties and disloyalty to existing allies, leading to a weakening or abandonment of alliances. In order to understand why people take sides in disputes on the basis of moral perceptions, one needs to consider explicitly what compensating advantages individuals might reap from their involvement in these disputes or, alternatively, what costs they might avoid from non-involvement.

J. Marczyk (✉)
Department of Psychology, MSC-3452, New Mexico State
University, P.O. Box 30001, Las Cruces, NM 88003-8001, USA
e-mail: jmarczyk87@gmail.com

First, this paper will briefly consider some of the adaptive problems people might be solving by taking sides in disputes more generally, how these adaptive problems might be solved, and what proximate inputs these side-taking mechanisms should be expected to use to solve them. Following that, moral side-taking in particular will be considered and differentiated from non-moral side-taking in terms of inputs. Some contemporary theories of the function of morality will then be discussed. Moral alliance strategies theory (MAST) will then be introduced and discussed. Finally, various features of moral psychology will be framed in terms of MAST to examine how well the theory can account for known findings.

Side-Taking Behavior

Disputes can be conceived of as strategic multiplayer games with different roles that individuals might fill at different times (DeScioli and Kurzban 2009a, b). These roles can be conceived of as actors (those who affect second parties through their behavior), second parties (those who are directly affected by actor behavior), and third parties (those who are not directly affected by actor's behaviors but might support either the actor or the second party). As a running example, consider the following: Person A (the actor) takes a resource from person B (the second party). Person B attempts to retaliate against person A for the action. Person C (the third party) could intervene to either protect A from B or to assist B in harming A.

The first question of interest is why third parties might become involved in disputes between actors and second parties. As there are potential intervention-related costs to third parties, a strategy that many third parties might find appealing is non-intervention. Indeed, non-involvement seems to be the strategy typical of most species (Tooby and Cosmides 1996). There are, however, a number of contexts

where third-party intervention on behalf of either the actor or the second party might prove to be adaptive.

One such context would be a mutualism scenario, where both the third party and the party they support benefit from the third party's side-taking behavior. As mentioned above, the roles in the side-taking game switch over time: the individual in the role of person C today might be in the role of person B tomorrow, contingent on A's actions. If person A were predicted to subsequently inflict costs on person C, it is in both B's and C's interest to defend against the costs inflicted by A through mutual side-taking. To the extent that inflicting costs on A to prevent the behavior alone are costlier than doing so together, B and C could work together to remove the threat posed by A.

Cognitive mechanisms designed for managing side-taking behavior in mutualistic scenarios should, at a minimum, use cues to shared interests as inputs. As side-taking in mutualistic contexts revolves around shared interest, the side-taking of C per se might have little predictive value when it comes to determining their future interactions with B if their mutual interests are subsequently de-coupled. It would only be when two or more individuals shared a substantially large overlapping set of mutual interests that side-taking today would be predictive of side-taking tomorrow, though it would not be the side-taking per se that is the input of importance.

Another such context favoring third-party involvement would be a reciprocity scenario (Trivers 1971). In this context, C might assist A or B in their current conflict to the extent that they predict A or B will help them in future ones. This scenario differs from the mutualistic one in that C need not have any immediate vested interest in whether A or B is benefited; it also differs from mutualistic scenarios in that the side-taking of C today should have predictive future value concerning future side-taking (Tooby and Cosmides 1996). Cognitive mechanisms designed for managing side-taking behavior in reciprocal scenarios should be expected to use a different set of inputs from mutualism, such as past side-taking history, to guide decisions. Any cues that hold predictive value for one's future side-taking likelihood could potentially serve as an input.

A third context which might favor third-party intervention would involve kin selection. As genes can improve their reproductive fitness by assisting other bodies that contain copies of them (Hamilton 1964), C might also take sides with A or B to benefit them contingent on shared genetic ties. Like mutualism, kin-based side-taking need not use reciprocity histories as inputs, though reciprocity in such relationships might often exist. Cognitive mechanisms designed for managing side-taking behavior in reciprocal scenarios should be attuned to inputs that cue genetic relatedness (like co-residency during early childhood in humans; Shepher 1971). As one's estimated relatedness coefficient to a given disputant increases, so too should the probability of siding with them in a dispute.

Moral Side-Taking

Third-party side-taking in moral disputes appears to diverge in some important ways from the three contexts outlined above. What makes side-taking on the basis of moral perceptions differ primarily concerns the inputs used by cognitive mechanisms in moral contexts: the behavior of actors (DeScioli and Kurzban 2013). Taking sides on the basis of who did what is markedly different from taking sides on the basis of cues of mutual interest, interaction history, or kinship. The next question of interest, then, becomes the matter of what benefits third parties can realize through using behavior as an input for determining their side-taking behavior.

Some theories of morality locate the function of this focus on behavior that improves aggregate welfare outcomes either in the individual or group sense (Baumard et al. 2013; Boyd et al. 2003; Darwin 1871; Fehr et al. 2002; Gigerenzer 2010; Haidt 2007). In order to achieve this welfare-benefiting function, we should expect moral mechanisms to use estimations of aggregate welfare outcomes as an input when considering what behaviors are morally acceptable.

There are a number of issues with the plausibility of these theories inasmuch as moral judgments often oppose aggregate increases in welfare (DeScioli and Kurzban 2009a, b, 2013). One classic set of moral dilemmas—the trolley and footbridge dilemmas—presents a notable example. In these dilemmas, redirecting a train to kill one person instead of five is judged to be far more morally permissible than pushing an individual in front of a train to save five (Mikhail 2007). A moral psychology designed to yield aggregate welfare benefits should see little difference between these two scenarios, owing to their identical welfare outcomes. Moral nonconsequentialism—the focus on features of moral dilemmas other than on aggregate welfare gains—poses a large theoretical hurdle for altruism models they do not seem able to overcome. There is also good reason for introducing a distinction between psychological altruistic and moralistic mechanisms (Kurzban et al. 2012).

A second contemporary theory locates the function for side-taking on the basis of behavior in more of a mutualism context. For instance, Tooby and Cosmides (2010) suggest that individuals might have been selected to attempt and leverage coalitional power to enforce their shared values on others or because of the coalition's downstream effects on rivals. While this suggestion certainly manages to outline some plausible benefits for involvement in moral disputes, it is unlikely to explain the full picture of moral condemnation. Specifically, this theory does not present a strong case for why people appear to take sides in disputes that seem to have little personal benefit.

As an example of this, in discussing the moral condemnation directed towards homosexual behavior, Tooby and Cosmides (2010) note that people's negative moral reactions

towards homosexuality might be the result of a personal, negative mental evaluation of it (i.e., “If I were to experience it, I would not enjoy it”) that is reframed in a moral light. While it is likely true that people can benefit from having some of their preferences moralized, it is not readily apparent that people engage in this kind of moralization for other preferences, such as whether one prefers to sleep on their side, enjoys the taste of carrots, or finds the idea of attending a country music festival pleasurable.

It is not clear why certain preferences might become moralized while others do not or that one’s personal evaluation of the act as unpleasant is what might be driving the moralization from mutualistic factors alone. Concerning homosexuality, for instance, people’s discomfort with homosexual individuals appears to be driven, in part, by the homosexual’s probability of contact with children (Gallup 1995). As it is unlikely that a homosexual’s probable contact with children is correlated with other individuals’ personal disgust toward engaging in homosexual acts, the personal preference model is at least incomplete.

Another contemporary theory locates the function of side-taking on the basis of behavior in coordination among third-party condemners (DeScioli and Kurzban 2013). The dynamic coordination account notes that third parties face risks of discoordination with other third parties when it comes to side-taking in disputes: if third parties are split relatively evenly on the matter of which side of a dispute they support, the evenly matched sides might need to escalate the conflict in order to determine a winner. Escalating a conflict can entail dispute-related costs that third parties are better off avoiding. Additionally, third parties might also need to avoid the costs brought on by excessively supported individuals (despots).

Accordingly, DeScioli and Kurzban (2013) posit that side-taking on the basis of the observable behavior of the disputants, rather than on their identity, can lead to disputes where these two sets of costs are minimized. If the majority of third-party support is given to one disputant, then the dispute can be settled without third parties incurring fighting costs. Further, if third-party support is organized around the disputant’s behavior, rather than identity, then no one individual can amass enough social support to initiate conflicts with third parties subsidizing the costs.

One potential concern with this account is that there seems to be no necessity for third parties to use the behavior of the disputants *per se* as the coordination device. As DeScioli and Kurzban (2013) note, the relationship of the nature of a coordination signal to the behavior it serves to coordinate could be arbitrary, referencing the example of a traffic light: there is no necessary relationship between the color of the light and driving behavior beyond an agreed upon meaning. Another such example used by DeScioli and Kurzban (2013) is trial by ordeal. The ability to survive a trial by combat could well be used as a coordination device for third parties when deciding

which side to take in a dispute, though it likely has little bearing on actual guilt or innocence. Other listed examples include coin flips, casting lots, and examining configurations of bones. This raises the question of why such practices are not more common and why people instead frequently opt to use behavior *per se* as the coordination signal. As an appreciable number of moral violations are committed without witnesses, and as many non-witnesses take sides on the basis of behavior they did not observe, behavior *per se* seems like a weak candidate for good a coordination device. In much the same way, a visually obscured traffic light would be a poorly designed coordination tool; it is not publicly and clearly observable. Publicly observable ordeals or even observable coin flips seem to be better coordination tools, relative to frequently non-observed behavior.

A second potential concern with the dynamic coordination account is that the benefits gained—or costs avoided—by third parties for their involvement in disputes, relative to non-involvement, are not made explicit. It is instead assumed that third parties are better off taking the side of one individual in a dispute, relative to spurning both sides. While that assumption may well be true, it is possible that making the benefits of involvement more explicit could also make clear some alternative functions for moral side-taking. Our knowledge of why people take a particular side in a dispute should be improved by our knowledge of why people become involved on any side.

While DeScioli and Kurzban (2013) do rightly suggest that expected personal benefits can influence moral stances, as Tooby and Cosmides (2010) do, they locate those benefits as being possible to achieve only *after* people are already taking sides on the basis of behavior. This is because there can be no adaptive benefits to favoring moral rules concerning particular behaviors unless people are already using behavior to take sides. This would suggest that benefits of side-taking behavior on the basis of actions could well have predated the favoring of personally beneficial moral rules.

Moral Alliance Strategies Theory

The goal of MAST is to take a conceptual step back and try to first provide an answer to the following question: what benefits might have been gained by third parties who took sides on the basis of the disputant’s behavior in the first place?

The answer to that question could be provided by returning to the reciprocity function for side-taking. As mentioned previously, a cognitive mechanism designed to manage side-taking on the basis of reciprocity should use past interaction histories as an input. While such an input is useful for guiding behavior on the basis of the past behavior of others, there are additional factors to consider: specifically, these relationships need to both be (a) initiated at some point and (b) only

maintained to the extent that they are likely to provide net benefits. Accordingly, MAST locates the answer to the question regarding side-taking on the basis of behavior in the fact that these behaviors generate needs in others, providing opportunities to both begin alliances by filling that need and to break off existing relationships that are too costly to maintain.

To understand MAST, it helps to first consider the banker's paradox (Tooby and Cosmides 1996): imagine that you happen to be a banker with some money to lend out and you are looking to make the best return on your investment. This means finding people who are both (a) willing to borrow money from you at the highest interest rate possible and (b) willing and capable of repaying that loan.

We could consider two potential loan prospects on opposite ends of a spectrum: those who have a lot of money on one side and those who have very little on the other. The former group would be quite capable of repaying the loan because they already have a lot of money, but their relatively low need for it would make them unwilling to accept a high rate of interest. The latter group has a great need for the money and so should be willing to accept a higher rate of interest on a loan; however, the fact that they need money might mean their ability to pay you back in the future is compromised.

The banker's paradox highlights the tension between various risks and rewards that can come with making investments. The logic does not just apply to bankers and money, though; social resources work as well, specifically friendship and other forms of social support. Given that I do not have an unlimited amount of time to spend with other people, and that I cannot take everyone's side in a dispute, I am unable to be everyone's friend. If I use my limited social budget pursuing people who would make useful allies but who want nothing to do with me, then I have wasted a resource I could have spent more profitably elsewhere. The same basic logic holds were I instead to pursue friendships with very needy individuals who would in turn offer me little to nothing as a social asset. These two types of friends would roughly correspond to the rich/poor loan targets from the banker's paradox. Given that there are reproductively relevant benefits from maintaining a healthy social network, selection should have been expected to shape cognitive mechanisms for managing our social investments adaptively.

One potential solution to the banker's paradox is to find individuals who are only facing temporary needs. Not only are such individuals currently in need of the investment—making it proportionately more valuable to them than someone not in need—but they are also more likely to repay it in the future than the chronically needy. Transient states of need do not necessarily make one a worse potential investment. As Tooby and Cosmides (1996) put it:

“...[I]f a person's trouble is temporary or they can easily be returned to a position of full benefit-dispersing

competence by feasible amounts of assistance...then personal troubles should not make someone a less attractive object of assistance. Indeed, a person who is in this kind of trouble might be a more attractive object of investment than one who is currently safe, because the same delivered investment will be valued more by the person in dire need.” (p. 132)

Encountering an individual with otherwise good investment potential in a temporary state of need might actually be a better investment than doing similarly when he is not facing that need.

Turning this example toward moral condemnation, consider a simple case of theft: person A steals from person B. In this case, person B might have a number of reasons to aggress against—or punish—person A, such as deterring future acts of theft or reclaiming his lost resources. However, enacting punishment against another individual is not without its risks: perpetrators might defend against such punishment and inflict further costs upon the punisher himself (Cinyabuguma et al. 2006).

Person B might thus be thought of as having a “need” to punish person A, but this need is costly to enact. Assistance from others can reduce that cost, as numerically superior groups tend to win in conflicts over smaller ones. By helping B fill that need by assisting in the punishment, third parties can potentially make themselves more valuable to B. Relatedly, if person B seeks to punish person A, person A now finds himself in need of support to defend against B and his allies. It follows that third parties can also make themselves more valuable to A by helping him avoid punishment. By mentally representing these behaviors that generate need as immoral or wrong, third parties can more effectively monitor which disputants offer the best alliance potential.

Any assistance which is offered by a third party might prove to be a valuable asset by either member of the dispute, and the provisioning of that asset might warrant future repayment. Third parties can thus demonstrate or improve their value to others by becoming involved in disputes on the basis of the behavior of the disputants.

According to MAST, the previous analysis explains why people become involved in moral disputes as third parties: moral support functions to signal one's value as a social asset to those involved in disputes. As such disputes generate transient states of need, they present opportunities for more lasting alliances to be formed, in much the same way that transient needs for money represent good opportunities for bankers to make money on loan interest.

To be clear, this type of reciprocity-building side-taking functions to result in partiality: one or more third parties are seeking to build lasting relationships with others involved in a dispute and are using a temporary state of need as a catalyst for the foundations of that friendship; what might be consider an

ingratiation strategy (Batson 1993). Despite the intended partial nature of these relationships, however, impartial moral mechanisms can arise from this dynamic for all members of the dispute. This is because avoiding the costs of condemnation, whether those condemners are partial or impartial, has potential adaptive value (see: impartiality).

It is also worth noting that moral disputes can quickly become complicated in this respect: when person C becomes involved in a dispute between persons A and B, person D may join the dispute to increase their association value to either A, B, C, or some combination of the three. What this means is that third parties might also join disputes to raise their value to other third parties, rather than the initial parties in the dispute. Nevertheless, the same logic should hold, regardless of which party a given individual is trying to raise his association value towards.

To get a sense for this complexity, one facet of third-party moral involvement that needs to be borne in mind concerns the nature of the moral interaction itself: *supporting one side in a dispute requires one siding against the other* (DeScioli and Kurzban 2009b). Let us say person C takes the side of person B in the above dispute. From the point of view of person B, person C is behaving altruistically (person C is enduring a cost to himself, by assisting in the punishment, in order to deliver a benefit to person B, which takes the form of reduced personal punishment costs); by contrast, from the point of view of person A, person C is acting spitefully (person C is enduring a cost to himself, by assisting in the punishment, in order to deliver a cost to person A).

Taking a moral stance requires that one is willing to trade off the welfare of one individual for another (inflicting a cost on A so that B might benefit). The willingness one has to trade off one person's welfare for another is called a welfare trade-off ratio or WTR (Petersen et al. 2010). For example, friends are expected to have a higher WTR toward each other than non-friends, so I should be willing to give up more of my welfare to benefit my friend or more of a non-friend's welfare to benefit my friend. Expanding this idea slightly, I should also be willing to give up more of a non-friend's welfare to benefit my friend. Morally condemning an individual can be conceptualized as a lowering of one's WTR with respect to the punished individual.

These modifications of WTRs extend beyond the initial parties themselves. By lowering one's WTR with respect to party A, you are by proxy inflicting a cost on A's allies: if person A is no longer receiving benefits, or is now enduring costs, his ability to provide benefits to others might be compromised. The allies of the condemned party might then lower their WTR towards their condemners and their condemner's allies, even if the allies of the latter were not initially involved in the dispute (Cushman et al. 2012; Uhlmann et al. 2012). Picking sides in disputes, then, should be expected to require a strategic assessment and management of the probable net changes of people's WTRs toward you as a result of taking one side or another.

These considerations allow us to more precisely define an individual's association value. One's association value to you comes in two parts: first, there is an individual-level component made up of how willing an individual is to trade his and other's welfare off for yours—his WTR with respect to you—and how able he is to enact that willingness. The second component is one's indirect association value, which is made up of how much *other people's* WTRs towards you will be modified by your associating with the target individual. For example, by siding with A over B in a dispute, the allies of A might increase their WTR with respect to, as you have aided their ally and, accordingly, your value to them, whereas the allies of B might do the reverse, as you have not inflicted costs on their ally. Consideration of these two abstract values should allow us to reconceptualize a number of features associated with moral perceptions and judgments.

Provided the preceding arguments hold true—that behavior generates need, which might serve as a catalyst for future, partial alliances—the matter of why third parties join disputes on the basis of behavior can be considered resolved to some extent: having alliances holds adaptive value, allowing allied individuals to better exploit others and avoid counter exploitation. With this argument in mind, various features of moral psychology can be framed in a new functional light that might aid our understanding of their form. Some of those features will now be considered, including why moral judgments are impartial, why moral judgments typically focus on harm, why moral condemnation is proportional, and why the positive end of the moral spectrum is worthy of consideration.

Explaining Some Features of Morality with MAST

Impartiality and the Moral Dimension

Appreciation of both the direct and indirect components of an individual's association value allows us to explain an otherwise-puzzling feature of moral condemnation: *impartiality*. Impartiality refers to the idea that a moral judgment is divorced from an actor's identity. Moral judgments strive for (at the least the pretense of) impartiality with some frequency (Kurzban et al. 2012; Lieberman and Linke 2007), and impartiality generally opposes friendship (Shaw and Knobe 2013). As MAST posits that the function of moral condemnation is alliance-building, this poses a conceptual hurdle.

The challenge posed by impartiality to MAST can be overcome by again considering the three-party dynamic of side-taking dilemmas. Beginning with third parties, there is a need to be able to represent what kind of acts tend to lead to opportunities for alliance-building—the ability to know which acts tend to generate need. Some of those acts will be relatively universal—such as murder and lying—while others will be

more specific to the individual in question—such as sexual promiscuity. As acts, not identities, are the things generating this need, the cognitive mechanisms in question should operate impartially at the level of behavior. Without these impartial representations, third parties would be unable to use behavior as a reliable cue for alliance potential.

Additionally, third parties need to be able to assess how much need behavior generates so as to manage their existing alliances. An ally who consistently transgresses against others, or who does so in egregious manners, will be frequently set upon by condemners, compromising their value to you as an asset to some degree; they will need frequent defense against rivals and devote their own resources to fighting off rivals that could be more profitably spent benefiting you. Not only will those who generate this need in others require assistance to defend against condemnation, they may also be less likely to receive altruism from others who share a vested interest in their well-being. As the value of an alliance might dramatically change with that ally's behavior towards others, this allows representations of behavior as immoral to guide both alliance-building and alliance-breaking behavior.

Accordingly, it may pay for third parties to represent behavior relatively impartially so as to accurately assess (a) who in the world currently offers the best alliance potential and (b) how other third parties might side in disputes. The latter portion is particularly important because, as previously noted, third parties can join in disputes to raise their association value to other third parties, not just to those initially involved in the dispute. Importantly, these two considerations hold *regardless* of whether the third-party support is ultimately in the service of generating partial alliances. So long as behavior generates social need and third parties are willing to step up and fill that need via assisting in condemnation, whether the third parties are doing so for reasons that are ultimately partial is irrelevant.

Turning to actors, as actors need to manage their own behavior in order to avoid condemnation (DeScioli and Kurzban 2009a, b), they also need to represent what acts tend to generate need and make others seem like good targets of social investment. This means that actors can use some of the same mechanisms to pick actions that third parties use to pick sides, as the two are inherently linked. Relatedly, actors, second parties, and third parties that have joined a dispute all face the need to recruit support to their side, making accurate knowledge of the factors which could recruit third-party support a vital piece of information. Thus, to the extent that people explicitly reason about what factors drive their moral judgments when trying to justify moral stances, this reasoning should only be effective to the extent that it is persuasive (Mercier and Sperber 2011). It follows, then, that if post hoc reasoning about moral stances frequently focuses on a particular topic, that topic should be expected to be an input relevant to the moral mechanisms.

MAST suggests that the degree of perceived need that is generated by a given act should account for an appreciable portion of the variance in ratings of *moral wrongness* or ratings of *immorality* (for criticisms of this perspective, see Section 2). In the most basic form of the above example, person A acts on B, after which B has a need to punish A. To get a fuller picture of how valuable this makes A and B to third parties, we should consider how great of a “need to punish” tends to get generated. For instance, A stealing B's car generates a much greater need than A stealing \$5 from B's wallet. The degree of this need ought to influence how much a third party could improve his association value to a given second party: the greater the need, the more valuable an investment in a given party should be, all else being equal. Accordingly, we might expect that ratings of immorality track how much need a given act tends to generate, on average. Since murder tends to generate more need than stealing, people rate the former as more immoral than the latter. This would help us understand why moral judgments are impartial: regardless of whether my family member or a stranger steals from someone, that someone will still experience some degree of need from the theft.

That said, since need is only one factor among many that people might use when considering whether to become involved in these disputes, or the extent of their involvement, we might expect that ratings of immorality, though impartial, correlate imperfectly with punishment and siding decisions, which are partial. This should be the case because moral impartiality is a threat to alliances, which require partiality. An individual who took sides on the basis of behavior *and nothing else* would not make a valuable associate, as their side-taking today would tell you nothing about their side-taking tomorrow.

Since we expect partiality from our allies in the form of relatively high WTRs, it follows that we ought to expect certain behavior on the part of close friends to be particularly moralized. According to MAST, it is readily understandable why Dante reserved the two innermost circles of hell for the acts of fraud and treachery in the *Inferno*: both reflect misrepresentations of association values. Committing a moral offense against a close friend, relative to a stranger, ought to be viewed as particularly morally condemnable due to the mismatch between the expected WTR inherent in the nature of the relationship and the actual low WTR displayed by the treacherous act. Indeed, this was precisely the pattern of results obtained by Haidt and Baron (1996): lying to a friend was rated as worse than lying to a stranger, and it mattered less whether such lies took place via omission or commission when they occurred between friends.

MAST can employ similar reasoning to explain another particularly puzzling facet of moral condemnation: why crimes against young children (Ward 2006) or other populations towards whom high degrees of altruism might be expected (perhaps women: Browne 2013; Glaeser and Sacerdote

2003; Mustard 2001) are often viewed as particularly heinous. Pinker (2002), for instance, writes about the usefulness of evolutionary theory for understanding “*why* we condemn prejudice, cruelty to children, and violence against women, and can focus our efforts on how to implement the goals we value most” (p. 422). The singling out of harm directed against certain groups requires an explanation. If people hold particularly high WTRs towards certain target individuals—like parents towards their dependent offspring—we might expect that transgressions against those targets should be viewed as particularly immoral, as the relative need or desire to punish generated by the transgressions is above average for the act itself.

Similarly, one might predict that certain individuals who would not be liable to return assistance should make poor targets of investment. Importantly, this should apply to individuals who are not likely to be able to return that assistance owing to certain uncontrollable factors. For instance, all else being an equal, one might predict that an elderly individual should be a less attractive target of investment than a younger individual, as the former has proportionately fewer remaining opportunities to repay. Indeed, this finding was reported by Callan et al. (2012): accidents befalling elderly individuals were rated as less unjust and the perpetrators less deserving of punishment, relative to younger individuals.

To summarize, an alliance-building perspective on moral side-taking highlights the importance of behavior, as the costs and benefits associated with behavior can provide cues as to others’ association values, in turn predicting where other parties might side. In order to better guide side-taking behavior, cognitive systems might represent behavior impartially. While behavior might be judged impartially, it is only one factor that third parties might attend to when picking sides in disputes, as behavior alone does not determine one’s association value.

Why Moral Judgments Are Often Harm-Centric

When asked to think of examples of immoral acts, those involving harm tend to be the prototypical examples that come to people’s mind (Gray et al. 2012), and when people rate acts as immoral, they also tend to represent victims of the acts, even when none are readily apparent or the claims of harm are of dubious credibility (DeScioli et al. 2012a, b). Pinker (2002) goes as far to suggest that all the good reasons for accepting or rejecting a moral proposition have to do with making people better or worse off. Intuitions about morality often seem to be centered on the idea of harms.

Why are harms so often the topic of moral thought? In principle, acts which are believed to not be harmful could be the subject of moral perceptions (DeScioli and Kurzban 2013), but they do not tend to represent the prototypical cases. According to MAST, as noted above, temporarily

disadvantaged individuals are likely to provide good social investment potential, all else being equal, due to their relative need. As inflicted harm tends to create people who are relatively needier (Petersen et al. 2010), these needier individuals should be expected to value the social support they receive more than less needy ones. It is likely for this reason that moral judgments are so often harm-centric: a credible claim of harm signals a transient state social need in the victim, which might provide the opportunity for alliance-building.

Conversely, an individual who consistently harms others will tend to generate many enemies. These enemies might present attractive assets to third parties due to the harms that the former group has suffered, leading the third parties to perceive punishing the initial perpetrator as a good investment. Individuals targeted for punishment, all else being equal, should make worse investments than those not so targeted. It follows that inflicting harms should generally be expected to lower one’s association value, while being harmed should generally be expected to raise one’s association value. When that harm is perceived to be inflicted by another individual, this creates a dispute in which sides could be taken and moral claims made.

Recognition of this idea allows MAST to account for interesting findings that the coordination and self-interest models have difficulty accounting for, such as why lying is viewed as less immoral when it benefits an individual other than the liar (Brown et al. 2005), and why telling the truth is not viewed as particularly honest when it causes harm, whereas lying is not rated as particularly dishonest when it results in benefits for others (Traifmow and Ishikawa 2012). From a coordination perspective, the behavior per se (truth-telling or lying) does not change, so third parties should coordinate against liars and not against truth-tellers regardless of the consequences the act generates. As one’s interest in not being lied to should not be expected to change as a function of whether lying benefits another individual in a given instance, the self-interest model should also predict that condemnation of lying not change on the basis of the consequences to others. According to MAST, however, as the need generated by an act changes, so too should our moral evaluations of it. While it is likely true that spreading false information tends to carry costs for others, situations in which this is not the case appear to reduce the moral response to the act. This is only the case, however, when it benefits someone other than the actor.

The importance of harm is further highlighted by a rather interesting finding: those who perceive acts to be immoral also overwhelmingly tend to represent victims for those acts, even if such victims are vague or ambiguous; those who did not represent an act as immoral did not tend to perceive victims (DeScioli et al. 2012a, b). If there is some truth to the matter, at least one group needs to incorrectly perceiving something. However, to the extent that people are persuaded to take sides in moral disputes on the basis of harm, these perceptions could

serve that function. Such a proposition finds support in experiments on moral reasoning where an appreciable minority of participants *did* reverse their moral judgments once their welfare concerns have been satisfied (16 %, Haidt et al. 2000; 30 %, Tetlock 2000). Forgoing the possibility that people's welfare concerns *had not* been satisfied by the experimenters as an explanation for why more people do not change their stances, if concerns about harm are not partially responsible for driving moral judgments, but rather be generated as post hoc justifications for intuitive or affective reactions (Haidt 2007), then it would be curious as to why attempting to satisfy welfare concerns should have any subsequent effect on moral judgments.

A rejoinder to the above points concerning the importance of harm would be to note that people frequently moralize harmless or even beneficial acts, such as cannibalism, consensual homosexuality, prostitution, masturbation, incest, or drug use (DeScioli et al. 2012a, b; Haidt et al. 2000; Haidt 2007). This criticism can be answered in a number of fashions, many of which turn on what precisely what is meant by "harmless."

The first answer to this criticism is that, in the moral sense, "harmful" and "beneficial" are always relative terms. A man paying a prostitute for sex might rightly be considered to be engaged in a mutually beneficial act in terms of himself and the prostitute. However, the availability of casual sex can have downstream consequences for those who wish to engage in monogamous relationships (Weeden and Kurzban 2014). As people are posited to be using behavior to take sides in disputes on the basis of behavior to generate partial relationships, what one counts as a harm and a benefit should, in some cases, be expected to vary as a function of which party it affects, rather than in the aggregate sense.

A second plausible answer to that criticism is that acts were historically associated with costs; in the case of incest, for example, offspring might suffer the proximate costs associated with inbreeding depression. That researchers can craft scenarios in which these proximate harms are avoided (Haidt et al. 2000; Tetlock 2000) might tell us little about the potential risk of harms to which the mind appears to be reacting (i.e., the expected value of the act). In a similar vein, people might still condemn a couple taking their life savings to a casino and betting it all on a round of roulette as imprudent even if they happened to have won (Jacobson 2012). People would also likely condemn drunk-but-lucky driving which does not result in a traffic accident.

A final point to make is that while harm caused might be one cue as to the probable availability of useful social alliances, it need not be the only one. Though some psychological researchers might not qualify taking offense or disgust as harm per se, one's tendency to offend others via their behavior (such as behaving disrespectfully towards others or in abnormal ways, such as by having intercourse with a chicken carcass before eating it) could well be a cue as to their association

value to the extent that they are predictive of one's motivations and future expected behavior (Jacobson 2012).

It is important not to lose sight of the larger function of moral psychology proposed by MAST: managing associations. There are cases in which one's association value can be threatened or improved even in absence of a dispute. For example, cleaning a toilet with your nation's flag could be used as a cue by others for predicting your loyalty: those who treat symbols of group membership with indifference or disdain will likely behave differently from those who treat them with respect or reverence. As disloyalty might be predictive of future costs and benefits you might provide as an ally, it can become the target of moral concerns. Using a flag to clean might thus not be considered directly harmful, but choosing an association with an individual who is less likely than others to provide you with benefits is costly.

Harm need not refer to acts that actually inflict costs on individuals; harm can also be conceptualized as failing to provide some expected level of benefits, even if benefits are still provided. This is because a positive WTR per se is not enough to build and maintain a relationship as not all relationships are equal in value while being zero-sum resources. Given a limited social budget to invest in relationships, accepting and investing in a relatively low-benefit friendship can be considered a cost when better alternatives are perceived to be available. Accordingly, MAST would predict that in certain behaviors—like altruism—will be moralized to different extents relative to the relationship between the individuals in question. A failure to provide useful benefits to socially close others should hold a different moral weight than a failure to provide those same benefits to strangers (Haidt and Baron 1996).

To the extent that other cognitive systems are also tracking association values using cues other than behavior that generates disputes, some of their outputs might occasionally be used by the moral system. Further consideration of this point should help explain why some preferences become the target of moral condemnation (such as various facets of sexual behavior) while others do not (whether one enjoys drinking milk).

It would also help explain another curious facet of moral judgments: people can be judged to be victims of themselves (DeScioli et al. 2012a, b). According to MAST, the representations of such behaviors as suicide and drug use as immoral, despite their ostensibly solitary nature, can be understood by noting that such behaviors have distinct effects on estimated association values: a dead individual is a poor associate, and drug use might make one behave in ways that generate negative externalities, either for third parties or friends and family (or engage in behaviors that fail to generate benefits for those parties). Morally condemning an individual for ostensibly solitary behaviors can be considered a recognition that such behaviors might threaten one's value as an ally, causing indirect harm to one's existing friends and family (via weakening the expected future value of such associations).

Moral condemnation directed towards an individual who harms himself might be a cue that he is compromising his association value to others. The threat of withdrawing investment from self-harming individuals might alter their behavior so as to reinstate their association value; if such efforts are unsuccessful, investment in those individuals might subsequently be abandoned. In other words, acts that display an overly negative WTR with respect to oneself without compensating benefits to others ought to draw moral condemnation, as they ultimately can reduce one's association value and, by extension, harm one's social allies.

Related to this idea is also the phenomenon of what has been labeled victim-blaming. Victim-blaming involves a victim of a crime either failing to receive appropriate third-party support or receiving third-party condemnation for either behaving in ways that put him in a position of elevated risk (such as traveling to a high-crime area alone at night) or failing to take adequate precautions against victimization (such as failing to lock his door when he left town). As engaging/failing to engage in these behaviors does not appear to be immoral *per se*, it is curious why they might affect the degree or target of third-party condemnation.

MAST is able to provide an explanatory framework for victim-blaming. Providing moral support as a third party is a costly endeavor, so it should follow that individuals who require less moral support should, all else being equal, appear to be more valuable associates. Accordingly, individuals who behave in ways that are predictive of a high probability of future victimization should be less appealing as associates, and people should be less inclined to support their moral claims. Returning to the banker's paradox, Tooby and Cosmides (1996) posited that individuals facing temporary need states might prove to be the best investments; those who signal that their needs will be chronic should not appear to be as appealing a target for investment. It is for this reason that people might blame victims for behaving without due caution. This line of reasoning could also explain why victims occasionally blame themselves (Perilloux et al. 2014): they are attempting to signal that their needs will be more transient, rather than chronic.

When harms are not readily observable, we should expect people to behave in ways that attempt to advertise them (Hagen 2003). This consideration might help explain why people occasionally engage in what is dubbed competitive victimhood (Sullivan et al. 2012). Competitive victimhood refers to the idea that individuals or groups often compete with one another to demonstrate who was subjected to more harm or injustice (Noor et al. 2012). If making oneself appear to be disadvantaged by another party serves to make one a more appealing target of social investment, we ought to expect some degree of competition between harmed parties to signal which is the more worthy cause.

MAST could similarly explain the finding that disadvantaged individuals tend to behave somewhat more selfishly (Xiao and Bicchieri 2010; Zitek et al. 2010) and punitively (Raihani, and McAuliffe 2012): if one is perceived as being unjustly victimized—and thus come to be viewed as a better potential target of social investment—he might also be more capable of avoiding the condemnation for selfish behaviors, at least temporarily (Gray and Wegner 2011). This is due to the fact that condemning an individual is unlikely to help build an alliance with him, as condemnation signals a negative WTR. Conversely, individuals who are relatively well-off and unharmed are less likely to be in need of social support, making them appear poorer targets of social investment.

Proportionality

Related to the notion of harms generating social needs, there is also the matter of proportionality in condemnation decisions (Petersen et al. 2012), summarized by phrases like, “an eye for an eye.” People do not typically support the death penalty for petty thefts or a week of imprisonment for murder as appropriate punishments. Punishments are frequently scaled in relation to the act's severity. Why should we expect that state of affairs?

One potential answer is that punishment might be enacted in proportion to the costs inflicted on others or the benefits reaped by the actor so as to make certain acts unprofitable to commit (Petersen et al. 2010; Pinker 2002). However, the upper limit on the effectiveness of such punishment is certainly higher than is frequently realized. For instance, if the death penalty were enacted for any and all acts of theft, that should highly deter stealing, as the expected costs of the act now far outweigh the benefits in most instances. Despite that, people do not frequently advocate for such extreme punishment, even if most people would be better off if they were not stolen from. While it is possible that enacting such extreme punishment might be prohibitively costly, people do not typically advocate against killing others as punishment for theft because killing others is difficult. The deterrence explanation for proportionality is thus an incomplete one.

The dynamic coordination model (DeScioli and Kurzban 2013) locates the explanation for proportionality in a need of third parties to coordinate with one another with respect to the magnitude of enacted punishment. This need, by itself, would not necessarily explain proportionality, though. It would seem that coordination around a single punishment value for all immoral acts would be less prone to discoordination, relative to attempts to coordinate around different punishment values for every possible immoral act. While this single-value punishment method of condemnation would be highly effective at achieving coordination, it would not yield proportionality.

The perspective advanced by MAST returns again to the notion that harms create social need. Importantly, this point

also applies to harms resulting from moral condemnation itself: individuals—and the allies of those individuals—set upon by moral condemners find themselves in need of social support to defend against the costs imposed by condemners. If the degree of moral condemnation is sufficiently greater than the harm inflicted by the perpetrator, then the relative value of investing in a given side of the dispute could plausibly shift. In other words, moral condemnation per se is capable of making victims of the perpetrator and of the perpetrator's kin and allies. In order to avoid making the condemned appear to be a more valuable target of investment potential, the punishment must be scaled appropriately to the perception of the harms initially inflicted.

This perspective would also explain why attempted crimes are typically treated differently than completed crimes (DeScioli and Kurzban 2013). On the surface, the only difference between attempted and completed crimes might be chance factors, making one wonder why condemnation of the acts should differ. However, attempted and completed crimes also have different consequences: a completed act generates a greater relative social need than an attempted one, and with greater need comes a greater degree of condemnation that can be enacted before the balance of association value shifts from the condemners to the condemned.

The Action/Omission Distinction

One interesting finding in the literature on morality is that outcomes that arise owing to omissions tend to be condemned more leniently than actions brought about through commissions (DeScioli et al. 2011a, b; DeScioli et al. 2011a, b). The dynamic coordination model (DeScioli and Kurzban 2013) has attempted to locate the explanation for that pattern of judgments in the idea that omissions do not generate evidence of wrongdoing and so do not generate a signal around which third-party condemners might coordinate.

To support that claim, DeScioli et al. (2011a, b) report a series of experiments in which participants were asked to deliver moral judgments about an individual who could divert a train or affect a demolition, the result of which would either kill or save an individual. When individuals had no causal effect on the killing (in one case, they failed to divert a train), those who opted out by doing nothing were rated as behaving less immorally than those who pressed a button that maintained the initial outcome. The authors suggest this result was obtained owing to the fact that pushing the button was a form of material evidence which might serve as a coordination signal for third parties.

However, it seems plausible that pressing a “do nothing” button has the effect of turning an omission (actually doing nothing) into a commission (pressing the button that maintains the outcome). In terms of the framework put forth by MAST, this ostensibly minor change should have a major impact.

While MAST would similarly argue that pushing the button can serve as a cue around which third parties might coordinate their side-taking behavior, what the action is signaling is the actor's WTR with respect to the victim: an actor who pushes the “do nothing” button could be perceived as demonstrating a negative WTR with respect to the victim, whereas the actor who does not do anything is simply failing to demonstrate any WTR.

To conceptualize this point, it is worthwhile to consider the discrepant responses obtained in response to the trolley and footbridge dilemmas: in the trolley dilemma, a train could be redirected away from the five hikers it is currently headed towards onto a side track where a single hiker would be killed instead; in the footbridge dilemma, a train is also headed towards five hikers, but to save their lives, one person has to be pushed in front of the train. A vast body of evidence has found that people tend to rate diverting the train in the trolley dilemma to be morally acceptable but pushing the man in the footbridge dilemma to be unacceptable (Hauser et al. 2007; Mikhail 2007).

To explain this discrepancy in moral acceptability, it is useful to reframe these dilemmas in light of problems humans might be more familiar with. Imagine instead that the trains in these dilemmas were not trains at all but rather intentional agents. The trolley dilemma can be understood as an agent of harm (i.e., a murderer) faced with the choice of harming five or one. The individual in this dilemma (the third party) acts on the agent of harm (the murderer) in such a fashion that causes it to choose a new target. In the footbridge dilemma, however, there is no initial choice: the agent of harm (murderer) is going to kill the five. It is then that the individual in this dilemma (the third party) trades off the welfare of one for the welfare of five himself. A key difference emerges here: in the trolley dilemma, the WTR is enacted by the agent of harm; the third party in it just provides some directional motivation. In the latter context, the WTR is enacted by the third party himself. Though both dilemmas involve the same WTR (trade off one life for five), the enactor of the WTR varies; this could explain why the former is not viewed with the same degree of condemnation as the latter.

Returning to the action/omission distinction, when an individual does not do anything, he is not directly signaling any negative WTR toward others through their actions. Third parties are not condemning omissions as strongly as commissions for the same reason my avoiding behavior that would kill someone is not praised as my saving lives. When such individuals press a “do nothing” button, though, their actions *are* sending a signal of a negative WTR with respect to the person about to be killed. In the experiments reported by DeScioli et al. (2011a, b) not pressing a button might be akin to not intervening in the actions of an agent of harm; pressing the button is akin to supporting the agent of harm directly.

Also of note is the finding that the action/omission effect can be reduced in cases where an actor either lies or fails to tell the truth to a friend, relative to a stranger (Haidt and Baron 1996). In that context, the available evidence for condemnation did not change nor was the omission transformed into a commission; the only change was the relationship between the individuals. Strangers are not expected to hold high WTRs with respect to one another, whereas friends are; we might even expect our friends to go out of their way to suffer costs to help us if necessary. Indications that they are unwilling to do so—whether through action or omission—represent betrayals of friendship and are morally condemned with greater force.

Guilt by Association

On some occasions, moral punishment does not appear to remain directed at the perpetrator of an act: sometimes, individuals who are perceived to aid or associate with the perpetrator might also be punished. This is the case even if the individual in question had no causal effect on the outcome.

Two empirical examples are worth mentioning. The first of these is a study examining revenge beanings in baseball (Cushman et al. 2012). Sometimes, a batter will be struck with a ball by the opposing team's pitcher; the pitcher from the former team may tend to strike a batter from the latter team in response. When asked about the moral acceptability of these revenge beanings (where a teammate of a perpetrator is hit with a pitch, despite not being the perpetrator), almost half of those surveyed suggested that the revenge pitch was morally acceptable; this percentage rose to a majority in cases where the person being surveyed was a fan of the team whose initial batter was hit. These intuitions concerning the moral acceptability of the punishment were held despite the fact that most participants did not view the batter being targeted for punishment as being morally responsible for the acts of their team's pitcher. Revenge beanings against other, non-offending teams, however, were deemed to not be morally acceptable by a wide margin. What this study seemed to suggest is that third parties appear to, at least in some situations, approve of punishing the associates of a wrongdoer, even if those associates were not themselves morally responsible for the act.

The second example of guilt-by-association concerns the willingness of third parties to hold kin somewhat responsible for the actions of their relatives. A study by Uhlmann et al. (2012) found that third parties tended to suggest that the biological grandchild of an exploitative factory owner should preferentially donate lottery winnings to the families of those impacted by their grandfather, even though the grandchild himself did not benefit from his grandfather's actions. Importantly, this effect was not observed to the same extent when the grandchild was related by marriage rather than genetics. A second study asked participants about whether or not two

individuals should be held in custody for a murder while police searched for more evidence or released until such evidence was discovered. The two individuals in question were either long-lost identical twins or exact lookalikes. Participants in this study expressed an increased willingness to hold the two suspects in custody when they were kin, relative to when they were strangers. In a final study, participants were asked about whether or not a child was morally tainted by the actions of his father, who had committed war crimes, despite the child never having contact with this relative. When his biological father was the war criminal, the son was perceived to be more morally tainted relative to when the criminal was simply the child's father by marriage.

It would seem that third parties express at least some willingness to inflict costs on the probable associates of a perpetrator at times, even if they do not believe that the associates have done anything wrong themselves. What explains this idea of guilt by association? According to MAST, guilt by association can be explained by referencing the indirect association value of an individual. By providing benefits to a perpetrator, an individual would be, *de facto*, making punishment of the perpetrator more difficult to enact. If one is attempting to punish a perpetrator, then one way of enacting this punishment would be to make associations with the perpetrator costlier. If people are disinclined from supporting the perpetrator or providing him benefits, this would not only inflict direct costs on the perpetrator (in the form of withdrawal of social support) but make enacting subsequent punishment easier, as the perpetrator would have fewer available allies available to assist in defense from punishment. Targeting biological kin of a perpetrator for condemnation, for instance, might represent an acknowledgment that such individuals are more likely to attempt to protect their kin from punishment, relative to non-kin.

Moral Praiseworthiness

DeScioli and Kurzban (2013) rightly caution against conflating explanations for morality with explanations for altruism, as adaptations for altruism—such as mammary glands—are not necessarily adaptations for morality. Nevertheless, countless writers on the subject from Darwin (1871) forward (Baumard et al. 2013; Gigerenzer 2010; Haidt 2007) appear to frequently discuss morality in the same breath as altruism, empathy, and compassion. There appears to exist an intuition that acts range from morally condemnable to morally praiseworthy, rather than just from morally condemnable to not morally condemnable (Miller 2007; Pizzaro et al. 2003; Wolf 1982).

MAST could explain the positive end of the moral spectrum: behaviors that are considered to be moral virtues (Miller 2007; Wolf 1982). Provided the function of morality is to manage one's associations, we should expect that moral

mechanisms should also pick up on certain acts that are reliable indicators of high association value: specifically, individuals who are relatively willing to sacrifice their own welfare to increase the welfare of others should prove to be highly valuable associates. Not only would such individuals tend to provide useful benefits, they might also be less likely to inflict costs or exploit their relationships. Additionally, there could also exist indirect benefits to being the friend of an altruist: inasmuch as others support them and they support you, morally praised individuals are more favorable allies than those not so favored.

One factor that differentiates moral praise from moral condemnation is that morally praiseworthy acts do not appear to generate disputes the way morally condemnable ones do. This is because, by and large, praiseworthy behaviors appear to be the ones that trade off an actor's welfare for the benefit of others (quite unlike behaviors that harm an actor's association value without benefiting others, like suicide). Without disputes, there can be no side-taking, rendering the dynamic coordination model unable to explain the positive end of the moral dimension. It is also important to differentiate moral praiseworthiness from the mere encouragement of an act. While people might be encouraged to save for retirement or to shop around for the lowest price, these acts are not typically considered to be worthy of any moral praise. This poses some conceptual problems for mutualistic models of morality, as people might wish to encourage mutually beneficial behavior without moralizing it. For instance, it might be both the interests of myself and a co-worker to commute to our job together, but it seems unlikely that we would morally praise those who commute for mutual benefit.

This point returns to the cautioning DeScioli and Kurzban (2013) provide against conflating morality with altruism, for not all altruism (or mutualism) is considered morally praiseworthy. For instance, a mother providing food or protection for her own child might not generate the perception of moral praiseworthiness, whereas the same mother providing those things for another individual's child might. This is because, owing to the logic of kin selection (Hamilton 1964), mothers should be expected to invest in their own children as such behavior is helping other copies of their genes. Accordingly, a mother's altruistic behavior toward her own children should not be a good indicator of how she would behave toward friends or strangers who do not share her genes (though her lack of altruistic behavior toward her offspring may well cue a lack of altruism toward strangers as well). By contrast, that mother's altruistic behavior toward unrelated individuals should serve as a much more reliable cue concerning her probable association value.

In more precise terms, MAST predicts that individuals should be perceived as acting in a morally praiseworthy fashion when they enact *unexpectedly high* WTRs with respect to those whose association values are low or unknown.

Examples of this might include a poor individual donating some of his very limited resources to those in need or a person jumping into a lake to try and rescue a drowning stranger. Importantly, this effect should be absent if the actor is perceived as behaving altruistically for his own benefit, as in such instances the actor would not be trading off his own welfare to help others, but rather trading off some of his own welfare at one time for more of his own welfare at another. Examples of the latter might include a mother feeding her own child or a man adopting a puppy in the hopes of attracting female sexual interest.

Also of note is that the target of this altruism might be of interest as well: donating one's time and money to helping the destitute rise out of poverty has a much different moral feel than donating the same time and money to help convicted rapists improve their station in life. One might predict that assistance delivered to those with *low* expected association values would be praised, whereas assistance delivered to those with *negative* expected association values would not. The former group are those who are simply in need of assistance, whereas the latter group include those who actively inflict harms on others. By assisting those who will harm others, one's altruism could plausibly be reframed as harm.

Discussion

Moral alliance strategies theory provides a theoretical framework for considering why third parties would take sides in disputes on the basis of behavior. In doing so, it helps reconceptualize the function of the moral dimension itself as a tool for managing associations, outlines how impartial moral mechanisms can arise from partial side-taking, explains the frequent emphasis on harm in moral judgments, and offers a novel perspective on the positive end of the moral dimension. While the full explanatory scope of MAST is not outlined here, it should be possible to expand these arguments to other facets of moral psychology, such as why some acts become moralized over time while other moral prohibitions slip away (Petersen 2013; Rozin 1999).

Like dynamic coordination, MAST posits that avoiding being on the losing sides of moral disputes is a cost to be avoided, requiring a focus on observable behavior. By making the benefits of using a behavioral strategy to pick sides in disputes more explicit, ostensibly strange moral behavior can be placed in new explanatory lights. To the extent that behavior can be predictive of shared interests, MAST is also compatible with mutualistic accounts of moral side-taking. Such an account does require that third parties are already taking sides on the basis of behavior, and MAST provides a framework for understanding why behavior-centric side-taking can make adaptive sense. Finally, MAST can help explain why so many past theories of morality have focused on

altruism when considering the moral dimension, as altruistic behaviors can help predict one's value as an associate.

References

- Batson, C. (1993). What is the difference? Communal and exchange relationships. *Personality & Social Psychology Bulletin*, 19, 677–683.
- Baumard, N., Andre, J. B., & Sperber, D. (2013). A mutualistic approach to morality: the evolution of fairness by partner choice. *Behavioral and Brain Sciences*, 36, 59–122.
- Boyd, R., Gintis, H., Bowles, S., & Richerson, P. (2003). The evolution of altruistic punishment. *Proceedings of the National Academy of Sciences of the United States of America*, 100, 3531–3635.
- Brown, J., Trafimow, D., & Gregory, L. (2005). The generality of negative hierarchically restrictive behaviors. *British Journal of Social Psychology*, 44, 3–13.
- Browne, K. (2013). Mind which gap? The selective concern over statistical sex disparities. *Florida International Law Review*, 8.
- Callan, M., Dawtry, R., & Olson, J. (2012). Justice motive effects in ageism: the effects of a victim's age on observer perceptions of injustice and punishment judgments. *Journal of Experimental Social Psychology*, 48, 1343–1349.
- Cinyabuguma, M., Page, T., & Putterman, L. (2006). Can second-order punishment deter perverse punishment? *Experimental Economics*, 9, 265–279.
- Cushman, F., Durwin, A. J., & Lively, C. (2012). Revenge without responsibility? Judgments about collective punishment in baseball. *Journal of Experimental Social Psychology*, 48, 1106–1110.
- Darwin, C. (1871). *Descent of man, and selection in relation to sex*. New York: D. Appleton and Company.
- DeScioli, P., & Kurzban, R. (2009a). Mysteries of morality. *Cognition*, 112, 281–99.
- DeScioli, P., & Kurzban, R. (2009b). The alliance hypothesis for human friendship. *PLoS ONE*, 4, doi:10.1371/journal.pone.0005802
- DeScioli, P., & Kurzban, R. (2013). A solution to the mysteries of morality. *Psychological Bulletin*, 139, 477–496.
- DeScioli, P., Bruening, R., & Kurzban, R. (2011a). The omission effect in moral cognition: toward a functional explanation. *Evolution and Human Behavior*, 32, 204–215.
- DeScioli, P., Christner, J., & Kurzban, R. (2011b). The omission strategy. *Psychological Science*, 22, 442–446.
- DeScioli, P., Asao, K., & Kurzban, R. (2012a). Omissions and byproducts across moral domains. *PLoS ONE*, 7, e46963.
- DeScioli, P., Gilbert, S., & Kurzban, R. (2012b). Indelible victims and persistent punishers in moral cognition. *Psychological Inquiry*, 23, 143–149.
- Fehr, E., Fischbacher, U., & Gächter, S. (2002). Strong reciprocity, human cooperation, and the enforcement of social norms. *Human Nature*, 13, 1–25.
- Gallup, G. (1995). Have attitudes toward homosexuals been shaped by natural selection? *Ethnology and Sociobiology*, 16, 53–70.
- Gigerenzer, G. (2010). Moral satisficing: rethinking moral behavior as bounded rationality. *Topics in Cognitive Science*, 2, 528–554.
- Glaeser, E., & Sacerdote, B. (2003). Sentencing in homicide cases and the role of vengeance. *The Journal of Legal Studies*, 32, 363–382.
- Gray, K., & Wegner, D. (2011). To escape blame, don't be a hero—be a victim. *Journal of Experimental and Social Psychology*, 47, 516–519.
- Gray, K., Young, L., & Waytz, A. (2012). Mind perception is the essence of morality. *Psychological Inquiry*, 23, 101–124.
- Hagen, E. (2003). The bargaining model of depression. In P. Hammerstein (Ed.), *Genetic and cultural evolution of cooperation*. Cambridge: MIT Press.
- Haidt, J. (2007). The new synthesis in moral psychology. *Science*, 316, 998–1002.
- Haidt, J., & Baron, J. (1996). Social roles and the moral judgment of acts and omissions. *European Journal of Social Psychology*, 26, 201–218.
- Haidt, J., Bjorklund, F., & Murphy, S. (2000). Moral dumbfounding: when intuition finds no reason. Unpublished manuscript, University of Virginia.
- Hamilton, W. (1964). The genetical evolution of social behavior I. *Journal of Theoretical Biology*, 7, 1–16.
- Hauser, M., Cushman, F., Young, L., Jin, R., & Mikhail, J. (2007). A disassociation between moral judgments and justifications. *Mind and Language*, 22, 1–21.
- Jacobson, D. (2012). Moral dumbfounding and moral stupefaction. *Oxford Studies in Normative Ethics*, 2.
- Kurzban, R., DeScioli, P., & Fein, D. (2012). Hamilton vs. Kant: pitting adaptations for altruism against adaptations for moral judgment. *Evolution and Human Behavior*, 33, 323–333.
- Lieberman, D., & Linke, L. (2007). The effect of social category on third-party punishment. *Evolutionary Psychology*, 5, 289–305.
- Mercier, H., & Sperber, D. (2011). Why do humans reason? Arguments for an argumentative theory. *Behavioral and Brain Sciences*, 34, 57–111.
- Mikhail, J. (2007). Universal moral grammar: theory, evidence and the future. *Trends in Cognitive Science*, 11, 143–152.
- Miller, G. (2007). Sexual selection for moral virtues. *The Quarterly Review of Biology*, 82, 97–125.
- Mustard, D. (2001). Racial, ethnic, and gender disparities in sentencing: evidence from the U.S. Federal courts. *Journal of Law and Economics*, 44, 285–314.
- Noor, M., Shnabel, N., Halabi, S., & Nadler, A. (2012). When suffering begets suffering: the psychology of competitive victimhood between adversarial groups in violent conflict. *Personality and Social Psychology Review*, 16, 351–374.
- Perilloux, C., Duntley, J., & Buss, D. (2014). Blame attribution in sexual victimization. *Personality and Individual Differences*, 63, 81–86.
- Petersen, M. (2013). Moralization as protection against exploitation: do individuals without allies moralize more? *Evolution and Human Behavior*, 34, 78–85.
- Petersen, M., Sell, A., Tooby, J., & Cosmides, L. (2010). Evolutionary psychology and criminal justice: a recalibration theory of punishment and reconciliation. In H. Oleson, H., MacMillian, P. (eds.), *Human morality & sociality: evolutionary & comparative perspectives*. New York.
- Petersen, M., 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, 682–695.
- Pinker, S. (2002). *The blank slate: the modern denial of human nature*. Viking.
- Pizzaro, D., Uhlmann, E., & Salovey, P. (2003). Asymmetry in judgments of moral blame and praise: the role of perceived metadesires. *Psychological Science*, 14, 267–272.
- Raihani, N., & McAuliffe, K. (2012). Human punishment is motivated by inequity aversion, not a desire for reciprocity. *Biology Letters*. doi:10.1098/rsbl.2012.0470.
- Rozin, P. (1999). The process of moralization. *Psychological Science*, 10, 218–221.
- Shaw, A., & Knobe, J. (2013). Not all mutualism is fair, and not all fairness is mutualistic. *Behavioral and Brain Sciences*, 36, 100–101.
- Shepher, J. (1971). Mate selection among second generation kibbutz adolescents and adults: incest avoidance and negative imprinting. *Archives of Sexual Behavior*, 1, 293–307.

- Sullivan, D., Landau, M., Branscombe, N., & Rothschild, Z. (2012). Competitive victimhood as a response to accusations of ingroup harm doing. *Journal of Personality and Social Psychology*, 102, 778–795.
- Tetlock, P. (2000). Coping with trade-offs: psychological constraints and political implications. In S. Lupia, M. McCubbins, & S. Popkin (Eds.), *Political reasoning and choice*. Berkly: University of California Press.
- Tooby, J., & Cosmides, L. (1996). Friendship and the banker's paradox: other pathways to the evolution of adaptation for altruism. *Proceedings of the British Royal Academy*, 88, 119–143.
- Tooby, J., & Cosmides, L. (2010). Groups in mind: The coalitional roots of war and morality. In Henrik Høgh-Olesen (Ed.), *Human morality & sociality: Evolutionary & comparative perspectives*. (pp. 191–234). New York: Palgrave MacMillan.
- Traifmow, D., & Ishikawa, Y. (2012). When violations of perfect duties do not cause strong trait attributions. *The American Journal of Psychology*, 125, 51–60.
- Trivers, R. (1971). The evolution of reciprocal altruism. *Quarterly Review of Biology*, 46, 35–57.
- Uhlmann, E. L., Zhu, L., Pizarro, D., & Bloom, P. (2012). Blood is thicker: moral spillover based on kinship. *Cognition*, 124, 239–243.
- Ward, C. (2006). Punishing children in the criminal law. *Notre Dame Law Review*, 82, 429–479.
- Weeden, J., & Kurzban, R. (2014). *The hidden agenda of the political mind: how self-interest shapes our opinions and why we won't admit it*. New Jersey: Princeton University Press.
- Wolf, S. (1982). Moral saints. *The Journal of Philosophy*, 79, 419–439.
- Xiao, E., & Bicchieri, C. (2010). When equity trumps reciprocity. *Journal of Economic Psychology*, 31, 456–470.
- Zitek, E., Jordan, A., Monin, B., & Leach, F. (2010). Victim entitlement to behave selfishly. *Journal of Personality and Social Psychology*, 98, 245–255.