An Evolutionarily Informed Study of Moral Psychology

Max M. Krasnow

As this volume attests, the study of morality is a difficult task. Philosophers, research scientists and scholars across the academy have been wrestling with how to define, measure, and think about morality for thousands of years. Yet, despite this effort, answers to fundamental questions have been devilishly elusive, with researchers still debating what morality even is. Why is it that the study of morality is such a difficult task? Perhaps the answer is less a result of the subject material than of the minds of those studying it. While the human mind is not usually considered an impediment to scientific progress, it may present particular barriers to accurate models of the nature of morality and moral psychology.

This is not the first research question that has been hampered by the fact that science is done by humans. Often, the problem is that we have a powerful intuition or perception of how the world seems or ought to be that gets in the way of scientifically understanding how the world really is. For instance, unassisted by technology, our eyes look to the horizon and see the Earth stretching out as if in a plane. And indeed, flat Earth theory was held in many cultures around the world for hundreds of years. Scientists had been studying gravity for centuries before Einstein's formulation of gravity as a distortion of space-time geometry gave us a more accurate, though far less intuitive, theory: general relativity. Here and elsewhere, the fact that human intuition or perception does not well map the real world has made humans worse at science. But the psychological barriers to understanding morality may not merely be a problem of this kind.

Whatever the specific design of our psychology turns out to be, results have been collected that make it very hard to believe that this design is simply a machine for uncovering the objective truth of the world. Whether it is the emotional dog wagging the rational tail (Haidt 2001), our heuristics and biases (Gilovich et al. 2002),

M.M. Krasnow (✉)
Evolutionary Psychology Laboratory, Department of Psychology, Harvard University,
980 William James Hall, Cambridge, MA, USA
e-mail: krasnow@fas.harvard.edu

B.G. Voyer, T. Tarantola (eds.), *Moral Psychology*,
DOI 10.1007/978-3-319-61849-4_3

or our susceptibility to visual illusions (Gregory 1997) and illusory correlations (Whitson and Galinsky 2008), it very often seems to be that the inferences the brain reliably makes are not always targeted at the objective truth. But, should we expect the mind to be designed for discovering objective truths? The human brain is a product of natural selection just like any other organ of the body, and just like every other organ of the body, the design features it has are there because they solved adaptive problems in the past (Tooby and Cosmides 1992). The brain, like any other feature of an organism, is the result of generations of successive filters on potential designs, filters t hat p referentially m aintained t hose d esigns t hat o ptimized r eproduction against the backdrop of environments experienced by the organism's ancestors. We should only expect design for uncovering objective truth to be a reliable feature of the human brain if doing so was a reliable solution to one of these filters by contributing to reproduction or by hitchhiking along as the by-product of another design that did. For many of the designs that make up human psychology, neither of these options is very likely.

In contrast, it is eminently more plausible that in many cases designs that vied away from objective truth seeking in the direction of inferences and behaviors that reliably contributed to reproductive fitness were the ones that better survived the various filters. We should expect this for three distinct but convergent reasons. First, there are likely many inferences for which knowing the true state of the world caries absolutely no fitness gain. For example, for a terrestrial primate, perceiving gravity as a distortion of space-time and not merely a force that pulls objects down toward the Earth cannot plausibly have influenced anyone's fitness over ancestral conditions; this information is irrelevant in the extreme. If there is no selection pressure that would maintain the focus of an inference system on reliably picking out the true state of the world, then randomizing forces like mutation should degrade its precision and allow drift over time. Or, if information necessary to make a certain inference was not reliably available ancestrally, then there would not have been any filter on designs that picked up on this information and used it in the right way. For example, it is inarguable that humans have a richly articulated mating psychology that uses information about prospective mates (cues of age, health, status, trustworthiness, fertility, etc.) to make mating decisions (Buss and Schmitt 1993). For men at least, a lot of this design is aimed at picking fertile targets of mating effort out of the sea of nontargets. But, there was no selective filter on mechanisms that meta-represented that this was their function; designs could increase reproduction by assigning high mate value in decision-making to bearers of particular cues like youthful appearance, low fluctuating asymmetry, etc., but there was no gain from representing that these inferences were about reproduction per se. So, while male mating psychology has design for inferring fertility and using it to inform mating decisions, it does not meta-represent what it is about and so does not accurately assess relevant fertility across all situations. That is why plastic surgery can maintain youthful attractiveness divorced from objective age, why men don't line up at sperm banks competing to fertilize as many eggs as possible, and why men continue to find their female partners attractive even when they are contracepting. Many mechanisms are likely to be under a similar filter, where solving a problem can be

done without the solution necessarily explicitly considering what the problem is and how the solution solves it. For at least these mechanisms, objective truth seeking per se is simply not the problem to be solved.

Second, there are likely many inferences for which the costs of getting the inference wrong are asymmetrical—that is, the false positives are more or less costly than the misses (Delton et al. 2011; Haselton and Buss 2000; Johnson et al. 2013). Taking again the example of a terrestrial primate, mistaking a bit of ground-level motion at your peripheral vision for a snake and deploying an evasive response is minimally costly—regardless of whether you are actually avoiding a snake or a harmless breeze, the energy expended is relatively minimal. Alternately, failing to detect a poisonous snake if it actually is there and thus getting bitten is a very costly error. Many ancestral problems are likely to have this asymmetric costs profile, and so mechanisms designed to infer the state of the world relevant for these problems are likely to incorporate design that makes the expensive error relatively less likely than the cheap error. For a snake-avoidance mechanism, the goal isn't to be as accurate as possible but to be as unbitten as is reasonable.

Third, the social world is not a solitary game: my behavior can influence others' behavior which can then impact my fitness. The beliefs I hold, my motivations for action, the things I value, and how I act can all have consequences, and can be relevant to others and how they treat me. For example, both game theoretic modeling and simple intuition predict that we will like others who have a history of good deeds, preferring them as social partners and treating them better than we would otherwise (Axelrod and Hamilton 1981; Barclay 2013; Trivers 1971). As such, design to be seen to do good deeds (but probably not to be seen being seen) has very likely been under selection. Behaviors, attitudes, ideas, and opinions that signal this kind of disposition (or other dispositions of similar relevance) are all potential targets of selection for expressing them, especially when others can see. To the extent that these targets (particular kinds of beliefs, opinions, etc.) are different than the objective state of the world, we should for this reason not expect the respective psychologies to be designed to be aligned to the objective truth. Who among us hasn't genuinely felt and then told a romantic partner that they are the most beautiful man or woman in the world? We can't all be right, but it sure feels nice to have a partner say so. In the moral domain, the selection pressures responsible for our moral sentiments—our concern for the sick, our outrage at the oppressor, etc.—may be more about what these sentiments signal to others than anything to do with objective truth seeking.

Taking these points together—that the objective truth is often fitness irrelevant, that the right kind of error is often ecologically rational, and that the adaptive problem is at least sometimes about changing someone else's behavior—helps suggest a program for an evolutionarily informed study of human moral psychology. The first task is to identify the major filters—that is, the adaptive problems—that components of moral psychology have been designed to solve. Considering the ecology of ancestral hominins, more than a few adaptive problems stand out as both presenting substantial selection pressure and potentially producing morality-relevant psychological design, including but by no means limited to optimally allocating resources

between the self and others (Delton and Robertson 2016), attracting and keeping cooperative partners (Delton and Robertson 2012; Krasnow et al. 2012), marshalling allies and maintaining one's group membership (Pietraszewski 2016; Tooby et al. 2006), and preventing yourself and those you value from being exploited (Krasnow et al. 2016; Sell et al. 2009). The task is to get specific about the recurring features of the ancestral environment that a design solution could use to solve the problem, including what kinds of information a behaving organism would have access to and how the organism's behavior could affect its outcome. Importantly, the points above suggest that our moral intuitions and reasoning about moral problems are likely the result of mechanisms shaped by one or more of the above selection pressures and therefore that they may be systematically biased in directions away from what might otherwise be considered the objective truth or normative moral correctness and toward what were ancestrally fitness m aximizing conclusions.

There is a large and growing literature that can be analyzed (or reanalyzed) using the evolutionary framework suggested here. Rather than attempting an exhaustive review, below I sketch what I see as a major dividing line in the space of adaptive problems involved in morality and discuss research that exemplifies the distinction in ways I hope will be helpful to researchers going forward.

Inward- vs. Outward-Facing Mechanisms

Regardless of the other features of an adaptive problem, the solution is either to regulate my behavior (solved by what I will call here an *inward-facing mechanism*) or to influence the behavior of others (solved by what I will call here an *outward-facing mechanism*) or both. Typically, the former category has been the main focus of work in moral psychology. For example, a lot of work looks at what it means to do good or to be altruistic. Philosophers ask on what theory "good" is measured, be it a utilitarian calculus or some other set of ideals. Economists propose elements of subjective utility (e.g., the "warm glow") that could compensate for otherwise costly behavior (Andreoni 1990; Fehr and Schmidt 1999). Cognitive psychologists and neuroscientists ask what are the proximate mechanisms of this decision and ask whether it is reflexive or deliberative (Rand et al. 2012) and if neural reward centers are involved (Decety et al. 2004). But, misconstruing the target of the mechanism across the distinction of inward-vs.-outward facing can have major consequences for our understanding of the psychology. If doing good has been selected because being seen by others to be doing good resulted in being chosen for more or better cooperative relationships—that is, results from an outward-facing mechanism—then researchers have been looking for the benefits to balance the equation in the wrong place. You would never intuit your way to this answer by introspecting on the experience of doing good; you would just conclude that you do good because it is the right thing to do, because it feels good, because it triggers a dopamine pulse, etc. Problematically, when the target of a moral adaptive problem is to influence

another's behavior, one's own representation of one's motives is likely to be especially suspect. As discussed below, mistaking the target as inward rather than outward facing may be an especially likely mistake for humans to make.

I should note this inward- vs. outward-facing dimension is very similar to the distinction DeScioli and Kurzban (2009) made between what they term "condemnation" (moral adaptations for judging others' bad behavior) and "conscience" (moral adaptations for governing one's own behavior in order to preempt others' condemnation). DeScioli and Kurzban construe the ecology of moral problems as involving perpetrators, victims, and observers, with condemnation resulting from the interest of observers and conscience resulting from the preemptive response of potential perpetrators. Yet, for conscience to preempt condemnation, it can use outward-f acing expressions (contrast the mere private experience of guilt with a guilty expression). Condemnation can result from inward-facing design (contrast outwardly concerned rehabilitative punishment with the private orientation of ostracism). Dissecting the problem space as inward vs. outward facing, I believe, more cleanly aligns the adaptive problems with the mechanistic design features that solve them. Below I review a selection of morally relevant psychological mechanisms to hopefully illustrate the utility of this alternative inward- vs. outward-facing distinction.

Moral Sentiments Regarding When to Be Nice

Inward-Facing Mechanisms

Codes of morality around the world are filled with proscriptions concerning when and how to be nice to others, when and with whom to share, and who is entitled to being helped. But how to optimally share resources is not just an abstract moral question; who gets what is inherently fitness relevant. In a world where others in your environment may share genes in common with you by virtue of recent common descent (i.e., kin), traits that allocate resources in ways that maximize the likelihood of your genes reproducing will be favored by natural selection; this is Hamilton's theory of cooperation via kin selection (Hamilton 1964). In a world where there are gains in trade to be had by pooling or exchanging your resources with others, then traits that maximize these gains will be favored by natural selection; this is Trivers' theory of reciprocal altruism (Trivers 1971). In a world where others represent unique value to you—such as unique constellations of mutual interest—then traits that tend to keep them around and in good shape would be favored by natural selection; this is Tooby and Cosmides' theory of deep engagement (Tooby and Cosmides 1996). On these and other theories, the mind should embody design that, at least in some circumstances, favors giving resources away to others and being perfectly happy to do so.

Recent work has asked, "What kind of psychology could embody strategies that produce these other-favoring effects?" In answering this question, researchers have

considered features of the ancestral ecology that simple heuristics could exploit to produce, on average, a good approximation of a solution. While a great deal about the ancestral world cannot be known, certain features can be safely assumed. For example, the ancestral social world was filled with different kinds of relationships; some people were strangers to you, and others were your family, friends, or cooperative partners. While there were doubtlessly many features that discriminated these categories from each other, it can be safely assumed that our ancestors could not predict the future with perfect certainty. At least sometimes, someone who at first blush appeared to be an irrelevant stranger never to be seen again actually became a relevant social partner (Krasnow et al. 2013). Moreover, for a hunting and gathering hominid with a specialized division of labor, long periods of childcare, and the ability to both accumulate and transmit cultural knowledge, there were likely many gains in trade possible between our ancestors where the gains were potentially lucrative. A tendency to trust others on the chance that a mutually beneficial relationship could develop—that is, a psychology of default trust—would be optimal social foraging in such an environment. While default trust is risky, as some investments would not pay off, the long-term rewards should be higher than those of a safer, asocial strategy (Delton et al. 2011; Delton and Robertson 2012; Rand et al. 2014). There are many ways such a design could be implemented. Just as our mechanisms of animacy and agency detection seem to be hypersensitive, attributing these features even when the evidence is scanty or absent, our mechanisms of social foraging could be designed to err on the side of treating even strangers as if they could be long-term cooperative partners. And just as our mechanisms of animacy detection help coordinate our behavior without going through explicit cognition—jumping away from a rustle you thought hid a snake did not require you to explicitly represent the propositions "this is a snake," "snakes are dangerous," and "snake danger can be mitigated by avoiding proximity"—our mechanisms of social foraging could plausibly be designed to effect behavior in the absence of explicit representations like "I might see this person again," "this person may be able to help me out later," and "if I don't see them again, at least I'm not risking much."

Relatedly, it is safe to assume that not all social partners were created equal; some were more trustworthy than others. When presented with attractive outside options (a more profitable partner to trade with, a tempting reason to cheat, etc.), some partners would have been more likely to take the option than others. A partner who simply doesn't consider these outside options should be more trustworthy than one who does, and to the extent that we can perceive cues to this disposition—such as a friend immediately agreeing when asked for help—a psychology that was sensitive to these cues and found them appealing in others would be favored by selection (Hoffman et al. 2015). Just as mating mechanisms are built to accept cues of fertility—available information like low waist-to-hip ratio (Lassek and Gaulin 2008) that partially indexes information that is otherwise inaccessible—social mechanisms should use observable cues in a partner's behavior like loyalty or blindness to outside options to index the otherwise inaccessible information of a partner's association value.

This work informs the kind of mechanisms we should expect to underlie our moral intuitions of when to be nice. Humans and our recent ancestors have been

intensely social for millions of years, so these adaptive problems have been long-standing. Solving a problem like social foraging with a robust intuition may be a timeworn solution, one that doesn't suffer from failures of explicit reasoning to anticipate future benefits. But while fitness maximization may be the ultimate explanation for our moral intuitions, it does not minimize their sincerity or authenticity. Just as a mother's love and concern for her child, a genuine and passioned response if there ever was one, is the result of mechanisms designed to maximize reproductive fitness, there is every reason to expect that our affiliation to our friends, our feelings of genuine concern for others, our intuitions about who deserves help and when similarly result from mechanisms designed to maximize reproductive fitness via social behavior. Some moral phenomena may be by-products of these inward-facing mechanisms, like our mechanisms of parental care can spill over onto our pets. But what about the expression of these emotions, motivations, and decisions? What problems do they solve?

Outward-Facing Mechanisms

Taking the above mechanisms as a given immediately suggests a reciprocal set of adaptive problems to be solved: how do you best position yourself to be preferred or chosen by others? When others in your environment are distributing resources, allocating aid, and forming relationships nonrandomly with respect to the characteristics of the recipients, selection should act to increase the prevalence of designs that preferentially capture these benefits. Just as preferences in peahens select for plumage in peacocks, selection pressures for social foraging result in selection pressures on social display. The instantiation of the mechanisms that embody these solutions can take many forms. As above, there is little reason to predict a priori that the solutions should necessarily route through explicit reasoning. Just as babies don't smile at their parents because they consciously consider the benefits of smiling, the mechanisms that instantiate our outward-facing responses to social selection need not be proximately Machiavellian. In fact, we should expect them to not have this design. To the extent that a benefit was provided by someone who explicitly saw something in it for themselves, the beneficiary should not attribute the gift to an underlying disposition on the part of the actor to value the recipient (Tooby and Cosmides 1996; Tsang 2006). A "friend" who only helps you out when it is in their own best interest is not much of a friend. In contrast, it is precisely those who are (or appear) insensitive to their own proximate payoffs that should be the most trustworthy and dependable cooperative partners. Imagine asking your friend a favor only to find them ponder at length all of the possible consequences they would face. What would you think of them? A heuristic solution to this pressure of impression management is to simply cooperate yourself without considering alternative—potentially more appealing—options (Hoffman et al. 2015).

A growing body of work suggests that this dynamic is likely to extend beyond the case of cues indicating *whether* a partner considered outside options before

cooperating to the more general set of cues indicating how much a partner values you at all (Delton 2010; Krasnow et al. 2016; Petersen et al. 2012; Sell et al. 2009). One interesting place this design is turning up is in the psychology of charitable giving. Charity is widely viewed as morally good and intuitively about increasing the well-being of the recipient. But, if that were the concern of the mechanism generating our charitable impulses, we would probably do charity a lot differently than we actually do. Many have begun to point out that most of our charity is incredibly ineffective: We don't pick causes that present the biggest problems, we don't fund solutions that provide the biggest benefits, and in large part we don't seem to care (Money for good: Revealing the voice of the donor in philanthropic giving 2015). Why is this? An intriguing possibility is that our minds are actually designed to prefer giving to less efficient charities because of what they can signal about how much we value others. I needn't value a child very highly to spend a dime to feed her for a year; even if I cared for her very little, I would still prefer to give up the dime. But I must value her highly to spend a dime to give her just a grain of rice; for how little she benefited, I must value her highly to justify giving up the money. If our psychology of charitable giving is the product of mechanisms designed for this outward-facing target of value signaling, then we should in fact predict different designs than were the targets merely inward-facing: rather than giving benefits efficiently at low personal cost to provide large charitable benefits, a psychology designed to signal how much it values others should look for (but not be seen looking for) opportunities to pay large costs to provide comparably inefficient charitable benefits and have a chance to be seen doing so.

Moral Sentiments Regarding When to Be Mean

Inward-Facing Mechanisms

Often our moral concerns fuel anger, outrage, or indignation toward those who violate our moral code. Sometimes these emotions result in behaviors that harm these individuals, ostracizing them from benefits they would otherwise have access to or inflicting costs on them through punishment or more violent aggression. Many theories have been proposed to account for the evolution or expression of these kinds of motivations and behaviors. But, as above, I argue here that theories have a better chance of being right when we properly construe the target of the adaptation as either inward or outward facing. Some adaptive problems can be solved by reaching out and changing another individual's behavior; these adaptive problems select for outward-facing solutions. It is likely that our punitive responses often result from such an outward-facing mechanism. But, does it always? If an adaptive problem does not have this form, researchers looking for outward-facing solutions would be looking in the wrong place.

One adaptive problem that punishment can solve is the mere prevention of future bad actions. By ostracizing, incapacitating, or killing a bad actor, the punisher and those she cares about are no longer susceptible to the bad action (Duntley and Buss

2011). Especially in the case of killing, these responses don't require the targeted individual to change their mind about anything for their bad behavior to be prevented; the decision is unilaterally made by the punisher. But, taking this option also precludes enjoying any of the benefits that would have otherwise obtained if the punished person was still around. Optimally negotiating this trade-off involves design for reducing the motivation for harsh, incapacitating, or corporal punishment given cues that these forgone benefits would be substantial—that is, that the perpetrator has high association value. And the mind indeed shows this design, favoring rehabilitative sanctions more for high association value perpetrators and punitive sanctions more for low association value perpetrators when deciding on criminal sentencing (Petersen et al. 2012; Wilson and Rule 2015). This function can be accomplished merely by moderating the sanction a person metes out, though. The outward expressions of offense—including facial, postural, and vocal expressions—are big noisy signals that are superfluous to this function. If these are adaptations, their adaptive target is likely of the outward-facing variety.

Outward-Facing Mechanisms

The evolutionary function of punishment on most theories is to change another organism's behavior (Clutton-Brock and Parker 1995). For example, conflicts of interests abound in life and can sometimes be adjudicated by force. This situation can be modeled as an asymmetric war of attrition, where (1) two parties make costly bids for a contested resource, (2) both parties pay the cost of the lower bid (i.e., fight until one gives up), and (3) the higher bidder wins the resource (Hammerstein and Parker 1982). In this scenario, each party is incentivized to bid just up to their private valuation of the resource; any more would be entailing sure losses and any less would potentially leave value on the table. Imagine fighting with your sister over what to watch on television. You each have your own preferred show, but only one can be watched, and you can annoy each other into giving in. If you don't care very much about your show, it would be silly to put up too much of a fight as you would waste more in fighting than you cared about the show in the first place. And, if you don't put up enough of a fight, you might end up missing your show when you didn't have to. A costly strategy would be to actually keep fighting with your sister until it's not worth it anymore, just in case she backs down first. But, if you can predict being outbid and losing the fight, you can save your effort and avoid fights you would otherwise lose. You are likely to be outbid to the extent that she either (a) values the resource more than you do, or (b) faces lower costs of aggression than you do, or both. As such, mechanisms that outwardly express our valuation and formidability would be selected to cost-effectively deter aggressive conflicts with others.

Many aspects of the anger response in humans and other animals can be understood as components of this signaling architecture (Sell et al. 2010, 2012, 2014). Humans and other animals posture before fights to size up the competition to predict

if fighting would be worthwhile. During these prefight rituals, the potential combatants don't merely stand passively; they modify their visual and auditory appearance to seem bigger, stronger, and meaner than they usually do. By engaging in this signaling, individuals can reach into the minds of observers and manipulate their mental contents in ways that advantage the signaling individual, potentially earning the contested resource and more deferential treatment in the future.

This kind of outward-facing mechanism can be used in larger social contexts as well. Just as signaling to someone who offended against you can deter them from doing so in the future, signaling to someone who offended against others in your presence can signal that you would not tolerate such treatment yourself. The third-party punishment paradigm—where one participant can punish another for acting poorly toward someone else—has been widely used to model moral condemnation. Recent work has revealed that at least some of the third-party punishment we observe in experiments results from this kind of deterrence mechanism (Krasnow et al. 2016). Moreover, punishing on behalf of others has been found to signal cooperative value more broadly, such as that the punisher herself could be trusted to not act badly (Jordan et al. 2016). Third-party moral condemnation seems at least in part to result from two outward-facing mechanisms for regulating the behavior of others.

As these examples illustrate, the components of our anger, punitive, or condemnation psychologies that are geared toward outward-facing targets—like signaling to others—were under reliably different selective filters than those components that are merely inward facing. Outward-facing mechanisms of signaling, for example, are expected to be under arms-race dynamics (Dawkins and Krebs 1979). The value of a signal depends on the population of signals it competes with. If everyone but you exaggerates their formidability, by neglecting to exaggerate, you appear weaker by comparison. The same process should apply to our expressions of outrage or condemnation; if everyone but you exaggerates their outrage to some moral violation, by neglecting to exaggerate, you appear relatively less trustworthy, more exploitable, etc., than you otherwise could. In contrast to behaviors that result from merely inward-facing mechanisms, those with outward-facing components are expected to be prone to these dynamics.

An Evolutionarily Informed Study of Moral Psychology

Applying the lens of evolutionary psychology to the study of morality offers several unique insights. Most basically, analyzing the ancestral human ecology for morality-relevant selection pressures can help generate hypotheses of adaptations in moral psychology—design features in the mechanistic basis of our moral intuitions, motivations, and decision-making. Here I have argued that it is profitable to distinguish those selection pressures that can be solved by merely inward-facing mechanisms targeted at directing the organism's own behavior from outward-facing mechanisms targeted at changing the behavior of others. One reason this distinction may prove important is that outward-facing mechanisms (e.g., broadcasting cooperative

disposition by charitable giving or public moral outrage) are expected to be under selection to obscure their ecological rationality (e.g., obscuring "ulterior" motives) even from those attempting to study them from an objective perspective. Using our intuition as a scientific instrument and source of hypotheses is therefore likely to systematically mischaracterize the adaptive design of our moral psychology and especially those involving outward-facing mechanisms. An evolutionary perspective helps clarify why studying morality is such a difficult task and also helps guide our efforts so that we are at least looking in the right place for the answers.

References

Andreoni, J. (1990). Impure altruism and donations to public goods: A theory of warm-glow giving. *The Economic Journal, 100*(401), 464–477. doi: 10.2307/223413.

Axelrod, R., & Hamilton, W. D. (1981). The evolution of cooperation. *Science, 211*, 1390–1396.

Barclay, P. (2013). Strategies for cooperation in biological markets, especially for humans. *Evolution and Human Behavior, 34*(3), 164–175. doi:10.1016/j.evolhumbehav.2013.02.002.

Buss, D. M., & Schmitt, D. P. (1993). Sexual strategies theory: An evolutionary perspective on human mating. *Psychological Review, 100*(2), 204–232. doi:10.1037/0033-295X.100.2.204.

Clutton-Brock, T. H., & Parker, G. A. (1995). Punishment in animal societies. *Nature, 373*, 209–216.

Dawkins, R., & Krebs, J. R. (1979). Arms races between and within species. *Proceedings of the Royal Society of London B: Biological Sciences, 205*(1161), 489–511. doi: 10.1098/rspb.1979.0081.

Decety, J., Jackson, P. L., Sommerville, J. A., Chaminade, T., & Meltzoff, A. N. (2004). The neural bases of cooperation and competition: An fMRI investigation. *NeuroImage, 23*(2), 744–751. doi: 10.1016/j.neuroimage.2004.05.025.

Delton, A. W. (2010). *A psychological calculus for welfare tradeoffs*. Santa Barbara: University of California.

Delton, A. W., Krasnow, M. M., Cosmides, L., & Tooby, J. (2011). The evolution of direct reciprocity under uncertainty can explain human generosity in one-shot encounters. *Proceedings of the National Academy of Sciences of the United States of America, 108*(32), 13335–13340.

Delton, A. W., & Robertson, T. E. (2012). The social cognition of social foraging: Partner selection by underlying valuation. *Evolution and Human Behavior, 33*(6), 715–725. doi: 10.1016/j.evolhumbehav.2012.05.007.

Delton, A. W., & Robertson, T. E. (2016). How the mind makes welfare tradeoffs: Evolution, computation, and emotion. *Current Opinion in Psychology, 7*, 12–16.

DeScioli, P., & Kurzban, R. (2009). Mysteries of morality. *Cognition, 112*(2), 281–299.

Duntley, J. D., & Buss, D. M. (2011). Homicide adaptations. *Aggression and Violent Behavior, 16*(5), 399–410. doi: 10.1016/j.avb.2011.04.016.

Fehr, E., & Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation. *The Quarterly Journal of Economics, 114*(3), 817–868.

Gilovich, T., Griffin, D., & Kahneman, D. (2002). *Heuristics and biases: The psychology of intuitive judgment*. Cambridge: Cambridge University Press.

Gregory, R. L. (1997). Visual illusions classified. *Trends in Cognitive Sciences, 1*(5), 190–194. doi: 10.1016/S1364-6613(97)01060-7.

Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review, 108*(4), 814–834. doi: 10.1037/0033-295X.108.4.814.

Hamilton, W. D. (1964). The genetical evolution of social behaviour. *Journal of Theoretical Biology, 7*, 1–52.

Hammerstein, P., & Parker, G. A. (1982). The asymmetric war of attrition. *Journal of Theoretical Biology, 96*(4), 647–682.

Haselton, M. G., & Buss, D. M. (2000). Error management theory: A new perspective on biases in cross-sex mind reading. *Journal of Personality and Social Psychology, 78*, 81–91.

Hoffman, M., Yoeli, E., & Nowak, M. A. (2015). Cooperate without looking: Why we care what people think and not just what they do. *Proceedings of the National Academy of Sciences, 112*(6), 1727–1732. doi: 10.1073/pnas.1417904112.

Johnson, D. D. P., Blumstein, D. T., Fowler, J. H., & Haselton, M. G. (2013). The evolution of error: Error management, cognitive constraints, and adaptive decision-making biases. *Trends in Ecology & Evolution, 28*(8), 474–481. doi: 10.1016/j.tree.2013.05.014.

Jordan, J. J., Hoffman, M., Bloom, P., & Rand, D. G. (2016). Third-party punishment as a costly signal of trustworthiness. *Nature, 530*(7591), 473–476. doi: 10.1038/nature16981.

Krasnow, M. M., Cosmides, L., Pedersen, E. J., & Tooby, J. (2012). What are punishment and reputation for? *PloS One, 7*(9), e45662. doi: 10.1371/journal.pone.0045662.

Krasnow, M. M., Delton, A. W., Cosmides, L., & Tooby, J. (2016). Looking under the hood of third-party punishment reveals design for personal benefit. *Psychological Science, 27*(3), 405–418.

Krasnow, M. M., Delton, A. W., Tooby, J., & Cosmides, L. (2013). Meeting now suggests we will meet again: Implications for debates on the evolution of cooperation. *Scientific Reports, 3*. doi: 10.1038/srep01747.

Lassek, W. D., & Gaulin, S. J. C. (2008). Waist-hip ratio and cognitive ability: Is gluteofemoral fat a privileged store of neurodevelopmental resources? *Evolution and Human Behavior, 29*, 26–34.

Money for good: Revealing the voice of the donor in philanthropic giving. (2015). Retrieved from http://static1.squarespace.com/static/55723b6be4b05ed81f077108/t/56957ee6df40f330ae018b81/1452637938035/$FG+2015_Final+Report_01122016.pdf.

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. doi: 10.1016/j.evolhumbehav.2012.05.003.

Pietraszewski, D. (2016). How the mind sees coalitional and group conflict: The evolutionary invariances of N-person conflict dynamics. *Evolution and Human Behavior, 37*(6), 470–480. doi: 10.1016/j.evolhumbehav.2016.04.006.

Rand, D. G., Greene, J. D., & Nowak, M. A. (2012). Spontaneous giving and calculated greed. *Nature, 489*(7416), 427–430.

Rand, D. G., Peysakhovich, A., Kraft-Todd, G. T., Newman, G. E., Wurzbacher, O., Nowak, M. A., & Greene, J. D. (2014). Social heuristics shape intuitive cooperation. *Nature Communications, 5*, 3677. doi: 10.1038/ncomms4677.

Sell, A., Bryant, G. A., Cosmides, L., Tooby, J., Sznycer, D., Von Rueden, C., et al. (2010). Adaptations in humans for assessing physical strength from the voice. *Proceedings of the Royal Society of London B: Biological Sciences, 277*(1699), 3509–3518.

Sell, A., Cosmides, L., & Tooby, J. (2014). The human anger face evolved to enhance cues of strength. *Evolution and Human Behavior, 35*(5), 425–429.

Sell, A., Hone, L., & Pound, N. (2012). The importance of physical strength to human males. *Human Nature, 23*, 30–44. doi: 10.1007/s12110-012-9131-2.

Sell, A., Tooby, J., & Cosmides, L. (2009). Formidability and the logic of human anger. *Proceedings of the National Academy of Sciences, 106*, 15073–15078. doi: 10.1073/pnas.0904312106.

Tooby, J., & Cosmides, L. (1992). The psychological foundations of culture. In J. H. Barkow, L. Cosmides, & J. Tooby (Eds.), *The adapted mind: Evolutionary psychology and the generation of culture* (pp. 19–136). New York: Oxford University Press.

Tooby, J., & Cosmides, L. (1996). Friendship and the Banker's paradox: Other pathways to the evolution of adaptations for altruism. *Proceedings of the British Academy, 88*, 119–143.

Tooby, J., Cosmides, L., & Price, M. E. (2006). Cognitive adaptations for n-person exchange: The evolutionary roots of organizational behavior. *Managerial and Decision Economics, 27*, 103–129.

Trivers, R. L. (1971). The evolution of reciprocal altruism. *The Quarterly Review of Biology, 46*, 35–57.

Tsang, J.-A. (2006). The effects of helper intention on gratitude and indebtedness. *Motivation and Emotion, 30*(3), 198–204. doi: 10.1007/s11031-006-9031-z.

Whitson, J. A., & Galinsky, A. D. (2008). Lacking control increases illusory pattern perception. *Science, 322*(5898), 115–117. doi: 10.1126/science.1159845.

Wilson, J. P., & Rule, N. O. (2015). Facial trustworthiness predicts extreme criminal-sentencing outcomes. *Psychological Science, 26*(8), 1325–1331. doi: 10.1177/0956797615590992.