Egocentric Ethics

Nicholas Epley[1,2] **and Eugene M. Caruso**[1]

Ethical judgments are often egocentrically biased, such that moral reasoners tend to conclude that self-interested outcomes are not only desirable but morally justifiable. Although such egocentric ethics can arise from deliberate self-interested reasoning, we suggest that they may also arise through unconscious and automatic psychological mechanisms. People automatically interpret their perceptions egocentrically, automatically evaluate stimuli on a semantic differential as positive or negative, and base their moral judgments on affective reactions to stimuli. These three automatic and unconscious features of human judgment can help to explain not only why ethical judgments are egocentrically biased, but also why such subjective perceptions can appear objective and unbiased to moral reasoners themselves.

Moral Philosophers of the Enlightenment generally assumed that objective moral principles existed—out there—in the world, and could therefore be divined with careful thought and clever argument. Although the subjectivity of human inference was clear even at that time, it was largely seen as an impediment to be overcome rather than the defining feature of mental life. Simple rules such as "act. . . in such a way that I can also will my maxim should become a universal law" (Kant, 1785/1964, p. 17) were seen to close the matter on moral ambiguities, as any clear-headed thinker would arrive at the same judgments regardless of status or circumstance. Those who did not could be dismissed as cloudy-headed thinkers who would eventually arrive at the "correct" conclusion once they set aside self-interest and overcame stupidity. Conclusions derived through these moral rules did not *feel* subjective, and thus appeared objective.

[1]Harvard University, Cambridge, Massachusetts.

[2]All correspondence should be addressed to Nicholas Epley, Department of Psychology, Harvard University, William James Hall 1480, Cambridge, Massachusetts 02138; e-mail: epley@wjh.harvard.edu.

Although dropping the penchant for pantaloons, everyday moral reasoners in the modern era seem to share this basic sentiment. Moral arguments in daily discourse often take on an objective sheen, and quickly devolve into shouting matches about who is right and who is wrong. The major problem for any objectively reasoned account of everyday ethical judgment, of course, is that everyday ethical judgments tend to be remarkably self-serving. Moral reasoners consistently conclude that self-interested outcomes are not only desirable but morally justifiable, meaning that two people with differing self-interests arrive at very different ethical conclusions. Such self-interested ethics often do not *feel* subjective, and are therefore perceived to be relatively objective.

Consider the recent dispute, for example, over ownership of Barry Bonds's record-setting 73rd home run baseball (Watercutter, 2002). The ball was hit deep into the right field stands, caught cleanly in the extended glove of Alex Popov, and lost into the welcoming hands of Patrick Hayashi in the ensuing skirmish. Popov held the ball first, Hayashi held it last, and both believed they were clearly the rightful owner for obvious ethical reasons. Ironically, both sides saw conclusive evidence for their position in the very same videotape (Luksa, 2003). A judge disagreed (or agreed?) with both and derived yet another position, deciding that the auction proceeds should be split evenly between them (Wilstein, 2003).

Stories like this are both common and predictable—diverging interests between two people, two groups, or two nations can lead to remarkably different ethical judgments. The most compelling demonstrations of egocentric ethics come in laboratory studies where self-serving judgments are based on diverging interpretations of identical information. For example, people in one study who were asked to decide on a fair allocation of wages claimed that they deserved, on average, $35.24 when they had worked 10 hours, but thought their partner deserved only $30.29 for the same work (Messick and Sentis, 1983). Similarly, subjects randomly assigned to the role of plaintiff or defendant in a hypothetical court case differed in their perceptions of a fair settlement by nearly $18,000 in the self-serving direction (Loewenstein *et al.*, 1993). Most important, however, is that the strength of these egocentric biases predict conflict and negotiation impasse between disputing parties (Babcock *et al.*, 1995; Thompson and Loewenstein, 1992). Clearly this conflict suggests that the subjectivity of moral reasoning is not especially clear to moral reasoners themselves.

As with most intuitive judgments, people making ethical judgments tend to be "naïve realists" (Robinson *et al.*, 1995), assuming that their perception of the world is a veridical representation of its actual properties rather than a subjective perception of the world as it merely appears to them. Others who perceive the world differently are therefore logically seen as motivationally distorted by self-interest, mentally crippled by stupidity, or both (Pronin *et al.*, 2002). It is these cynical attributions about others' motives and intentions that are especially problematic and lead to negotiation impasse, intransigence, and relationship dissolution.

Without denying that some differences of opinion are likely based on explicit, unabashed self-interest, the goal of this chapter is to sketch out a more benign possibility that explains why ethical judgments are consistently egocentrically biased, why they nevertheless feel perfectly objective, and why efforts to eliminate these egocentric biases have largely been unsuccessful. This possibility connects the dots between three distinct sets of empirical findings and suggests that egocentric ethics are produced by automatic and unconscious psychological mechanisms. First, people automatically interpret their perceptions egocentrically. This egocentric default is only subsequently (and insufficiently) adjusted if attentional resources are available, or if subsequent evidence makes it clear that one's initial position was in error. Second, people automatically evaluate stimuli and events as positive or negative, as good or bad. Coupled with automatic egocentrism, these evaluations are likely to determine whether an outcome or event is good or bad from one's own perspective—for *oneself*. Finally, moral judgments appear to be based on exactly these kinds of automatic evaluations. Positive automatic evaluations can lead to the perception that an ethical event is moral, whereas negative automatic evaluations can lead to the perception than an ethical event is immoral. Because egocentric evaluations happen rapidly, unintentionally, effortlessly, and without conscious awareness (i.e., automatically; Bargh, 1994), there is no trace of biased reasoning or ethical subjectivity to stimulate judgmental correction (Wilson and Brekke, 1994). Egocentric moral reasoners therefore feel that they have perceived the world as it actually is, rather than the way it simply appears to them. Although this three-step model does not prescribe easy remedies for alleviating egocentric ethics, it does lessen the sting of cynical attributions that arise in moral disputes. The words that follow describe the empirical evidence that led us to this conclusion.

AUTOMATIC EGOCENTRISM

People see the world through their own eyes, experience it through their own senses, and have more access to the others' cognitive and emotional states. This means that one's own perspective on the world is directly experienced, whereas others' perspectives must be inferred. Because experience is more efficient than inference, people automatically interpret objects and events egocentrically and only subsequently correct or adjust that interpretation when necessary (Epley *et al.*, in press a; Gilbert and Gill, 2000; Keysar *et al.*, 1998; Nickerson, 1999). The automatic default occurs rapidly but correction requires time and attentional resources, meaning anything that hinders one's ability or motivation to expend attentional resources will systematically hinder correction. As a result, many social judgments in the attention-demanding domains of everyday life tend to be egocentrically biased. For example, people tend to overestimate the extent to which others notice

and attend to their behavior (Gilovich and Savitsky, 1999), overestimate the extent to which their internal states are transparent to others (Gilovich *et al.*, 2000; Vorauer and Ross, 1999), and overestimate the extent to which others will share their attitudes, beliefs, knowledge, and emotional reactions (Keysar and Barr, 2002; Prentice and Miller, 1993; Ross *et al.*, 1977).

Several findings suggest that these egocentric biases are the downstream consequence of an automatic egocentric default. First, egocentric biases increase when the ability to expend attentional resources is compromised. For example, people tend to evaluate their abilities in comparison to others by egocentrically focusing on their own absolute abilities and insufficiently considering others' abilities (Klar and Giladi, 1997, 1999; Kruger, 1999). This leads to reliable above average effects in domains where absolute ability levels tend to be high (such as driving) and below average effects in domains where absolute ability levels tend to be low (such as juggling). What is more, these egocentric biases were especially strong in one experiment among participants who made their evaluations while simultaneously holding a six-digit number in mind (Kruger, 1999, Study 3). This cognitive load presumably precludes allocation of the attentional resources necessary to correct an automatic egocentric default.

Second, egocentric biases are reduced when participants are given financial incentives for accuracy (Epley *et al.*, in press a, Study 3). Presumably such incentives enhance motivation to expend the attentional resources described in the preceding paragraph, thereby producing greater correction of an automatic egocentric default.

Third, egocentric biases increase when people are asked to respond quickly (Epley *et al.*, 2003b, Study 2). This rapid responding presumably precludes the time required to correct or adjust an automatic egocentric interpretation, thereby leading to less extensive correction and stronger egocentric biases.

Fourth, egocentric biases are enhanced by manipulations that increase the likelihood of accepting values encountered early in the process of adjustment away from an egocentric default. Participants in one experiment, for example, were played a message that could be interpreted as either sarcastic or serious (Epley, 2001). Some participants were informed that the author intended the message to be serious, others that the author intended the message to be sarcastic, and all estimated the percentage of uninformed peers who would perceive the message as sarcastic. More important, approximately half of the participants made these estimates while nodding their heads up and down whereas the other half did so while shaking their heads from side to side. Previous research has found that people evaluate hypotheses more favorably while simultaneously nodding their heads up and down (in an affirmative fashion) than when shaking their heads from side to side (in a rejecting fashion; Brinol and Petty, 2003; Wells and Petty, 1980), and people nodding their heads up and down have been found to adjust less from an initial anchor value in judgment than people shaking their heads from side to side

(Epley and Gilovich, 2001). Similarly, participants in this experiment tended to assume that others would interpret the ambiguous message in a manner consistent with their own interpretation, but this egocentric bias was larger among participants who were nodding their heads up and down than among participants who were shaking them.

Finally, people make egocentric responses more quickly than nonegocentric responses. In one experiment, for example, those who indicated that others would interpret a stimulus in the same manner as they did responded more quickly than those who indicated that others would interpret the stimulus differently (Epley *et al.*, in press a, Study 2). In another study, participants were asked by an experimental confederate to move objects around a vertical grid (Keysar *et al.*, 2000). Some of the objects could be seen only by the participant, whereas others could be seen by both the participant and the confederate. On critical trials, the confederate made an ambiguous instruction that could refer to two objects, one hidden from the confederate and one mutually observable. Results showed that participants tended to look first at the hidden object suggested by an egocentric interpretation of the instruction, and only subsequently looked at the mutually observable object.

Collectively, these results demonstrate that people automatically interpret their perceptions egocentrically, and only subsequently adjust or correct that interpretation when necessary. Because such corrective procedures are notoriously insufficient (Epley and Gilovich, in press; Gilbert, 1989; Gilbert and Gill, 2000; Tversky and Kahneman, 1974), social judgments tend to be egocentrically biased. Although psychologists have traditionally considered egocentric judgment to be a stage outgrown with development, much like the ethical subjectivity observed by moral philosophers, these results suggest that egocentrism isn't merely outgrown with time but rather overcome in each social judgment. Indeed, in an eye-tracking paradigm using a vertical grid similar to that just described, children and adults did *not* differ in the speed with which they interpreted an instruction egocentrically (after correcting for baseline differences), but *did* differ in the speed with which they corrected that interpretation (Epley et al., in press b). Adults may not end up making completely egocentric judgments, but it appears that they usually begin there.

AUTOMATIC EVALUATION

Ethical judgments, however, are much more than matter-of-fact egocentric assessments. They are defined by an evaluative component, a sense of good and bad, of right and wrong, of positive and negative. Although these evaluations can be generated through careful deliberation and conscious reasoning, they can also be generated automatically—rapidly, effortlessly, unintentionally, and unconsciously (Bargh, 1994). Decisions about whether to approach or avoid a stimulus are among the most basic and important any organism can make, and the functional benefits of

rapid responses—especially in the presence of a personal threat—are fairly obvious
(Fazio, 1989). It should thus come as no surprise that evolution has fashioned
a neural system that quickly and efficiently evaluates virtually every stimulus
encountered. Coupled with an automatic egocentric default, this means that people
will likely be automatically evaluating whether a stimulus, event, or outcome is
good or bad for *them*. In fact, the most important dimensions of a concept's meaning
can be reliably captured by having people provide evaluative ratings on a series of
bipolar scales such as "good–bad" (Osgood *et al.*, 1957). It appears that the mere
process of perceiving a stimulus entails an evaluation of that stimulus.

Automatic evaluations are demonstrated through a variety of sources. First,
all organisms can exhibit rapid approach and avoidance behaviors in response to
stimuli (Schneirla, 1959). This includes bacteria and plants (Zajonc, 1998), whose
lack of higher order cognition seems fairly clear. The human brain evolved out of
these affectively based systems, and the resulting architecture served to correct or
override these automatic evaluative responses rather than to replace them. Basic
evaluative responses—such as fear—can even occur before any neural activation
in the centers of higher order cognition via a direct neural pathway through the
amygdala (Wilensky *et al.*, 2000).

Second, automatic evaluations can be seen in sequential priming paradigms
where affectively valenced words presented too quickly to be strategically evalu-
ated nevertheless activate similarly valenced words. In the most common version
of this paradigm (e.g., Fazio *et al.*, 1986), participants are presented with a positive
or negative attitude object (e.g., party or death), quickly followed by a positive or
negative target word (e.g., delightful or awful). Participants indicate whether the
target word is good or bad by pressing a computer key as quickly as possible.
Results typically indicate that participants are faster to respond to the target word
when it is preceded by a similarly valenced prime. That is, positive primes facil-
itate recognition of positive words, and negative primes facilitate recognition of
negative words.

Such results demonstrate automatic evaluation because they occur when the
target is presented too quickly after the onset of the prime to allow for conscious re-
sponding. In most experiments, the target word is presented approximately 300 ms
after the prime, when 500 ms appears to be the minimum time required for con-
scious responding (Neely, 1977). Variations on this procedure show similar results
even when the prime itself is presented subliminally (Greenwald *et al.*, 1995;
Krosnick *et al.*, 1992), when the prime is perceptually degraded (De Houwer
et al., 2001), and when participants are given no explicit goal to evaluate the
primes (Bargh *et al.*, 1996; Duckworth *et al.*, 2002). The effect also replicates
using a wide variety of prime stimuli, including faces of romantic partners (Banse,
1999), landscape pictures (Hermans *et al.*, 2003), musical sounds (Sollberger *et al.*,
2003), odors (Hermans *et al.*, 1998), spoken words (Duckworth *et al*., 2002), and
written words (Bargh *et al.*, 1992; Fazio *et al*., 1986).

Finally, people respond faster with behavioral actions that are consistent with the valence of a stimulus, highlighting the preparatory function of automatic evaluations. For example, participants in one experiment were asked to either push or pull a lever positioned in front of them to indicate whether a target word was good or bad (Chen and Bargh, 1999). Some participants were asked to pull the lever toward them (consistent with an approach motivation) to indicate that a target word was positive and push the lever away (consistent with an avoidance motivation) when it was negative. The other participants were asked to do the opposite. Results indicated that participants were faster to respond in a manner consistent with the evaluative connotation of the words— to pull faster when the target was positive and push faster when it was negative. A second experiment more clearly demonstrated automaticity by asking participants to simply push or pull as soon as a word appeared on a computer screen, rather than to evaluate it as good or bad. Although responses occurred too quickly for conscious responding to the stimulus, participants were nevertheless faster to pull the lever when the target word was positive (compared to negative) and faster to push the lever when the target word was negative (compared to positive).

Initial accounts of these automatic evaluations relied on the spreading activation of concepts stored in memory, whereby activation of a concept also activated its associated valence. Such automatic evaluations, however, would have little impact on most everyday ethical judgments because they tend to involve novel attitude objects. But recent evidence challenges this spreading activation account, because automatic evaluation effects are observed with both weak attitude primes (Bargh *et al.*, 1992, 1996) as well as novel attitude primes such as abstract polygons and Chinese ideographs (Duckworth *et al.*, 2002). This suggests that novel ethical dilemmas about which no preexisting attitude exists are completely open to automatic evaluation, and do not necessarily rely on previous experience with the particular object at hand.

Although little evidence directly links automatic evaluations with ethical judgments, recent research has shown that automatic evaluations are dependent on a perceiver's role and current goals—a critical finding for ethical judgments. In one experiment, for example, the word "dentist" facilitated recognition of a positive target when it was preceded by the word "doctor" but facilitated recognition of a negative target when preceded by the word "drill" (Ferguson and Bargh, 2004). In two other experiments, automatic negative evaluations of stereotyped outgroup members were weakened after exposure to positive exemplars of outgroup members (Dasgupta and Greenwald, 2001) or after exposure to positive stereotype contexts (i.e., a family barbeque versus a gang incident; Wittenbrink *et al.*, 2001). More important, these context-dependent attitudes appear to be relatively stable as long as the context remains constant (Dasgupta and Greenwald, 2001; Ferguson and Bargh, 2004).

These context-dependent results are of obvious importance to automatic egocentric ethics. Our thesis, after all, is that people on opposing sides of a moral dispute have automatic evaluative responses consistent with an egocentric evaluation of costs and benefits. Evaluations are not based on stable attitudes or preferences, but are constructed based on an egocentric assessment of what is good and bad from their own perspective. Outcomes that benefit the self invoke a positive automatic evaluation, whereas outcomes that hurt the self invoke a negative automatic evaluation. These speculations are completely consistent with the context-dependent nature of automatic evaluations. Notice also that the automatic nature of these egocentric evaluations leave no hint of subjectivity, attentional effort, or bias to stimulate judgmental correction (Wilson and Brekke, 1994), producing perceptions that appear to be caused by the stimulus itself rather than by the biased evaluations of the perceiver. These automatic egocentric evaluations are then seen as valid representations of reality, and opposing viewpoints as self-interested distortions. The intransigence of many moral disagreements may therefore stem directly from the automatic and unconscious evaluations upon which they are based.

EVALUATIVE MORAL JUDGMENT

Not wandering far from the sentiments of Enlightenment philosophers, moral psychologists have traditionally assumed that moral judgment involves a deliberate process of reasoning and reflection (Kohlberg, 1969; Piaget, 1932/1965). On this account, the emotional reactions associated with moral judgments are caused by moral reasoning, and can therefore be changed by altering one's reasoning. According to this logic, people only determine the morality of an act after they have consciously considered its consequences. Consistent evidence comes from structured interviews in which participants are presented with moral dilemmas and asked to resolve the conflict. Moral reasoning and moral judgment are often highly correlated within this deliberative paradigm, and become more cognitively complex and unconventional as a person ages.

Although a rationalist account of moral judgment has intuitive appeal because of its logical structure, Haidt (2001) points out that it has difficulty explaining several empirical findings. First, most judgments and behaviors appear to be made automatically, with little intention, awareness, or effort (for reviews see Bargh, 1994; Greenwald and Banaji, 1995; Wegner and Bargh, 1998). People form impressions of strangers (Ambady *et al.*, 2000; Devine, 1989; Higgins *et al.*, 1977; Uleman *et al.*, 1996), interact with others (Chartrand and Bargh, 1999; Chen and Bargh, 1999; Lakin and Chartrand, 2003), and make decisions (Dijksterhuis and van Knippenberg, 1998; Pelham *et al.*, 2002; Wilson and Schooler, 1991), for example, through psychological mechanisms that are unintentional, uncontrollable, and completely unavailable to conscious introspection. The ease and speed with which people make moral judgments in everyday life makes them a prime

candidate for similar unconscious mechanisms. Although the elaborate and deliberative interview method designed by Kohlberg may be perfectly reliable, it may also be completely unrepresentative of most moral judgments.

Second, conscious reasoning appears to be the consequence of these unconscious behaviors and judgments rather than the cause of them. People asked to explain the causes of their behavior, for example, often cite irrelevant causes and overlook relevant ones. Women in one experiment were asked to explain why they chose one particular brand of panty hose over another. In reality, the order in which the panty hose were presented dramatically influenced choices (women tended to choose the last pair considered), a factor not mentioned by a single woman (Nisbett and Wilson, 1977). The introspective search for the causes of judgment and behavior actually involves a process of inference based on culturally shared explanations for behavior, rather than a report based on direct access (Nisbett and Wilson, 1977; Wilson and Stone, 1985). Reasoning is also chronically distorted by motivational biases, such that people reason in ways that support a preexisting decision rather than analyze it logically or rationally. People reason in ways consistent with what they want or expect to see (for reviews see Dunning, 1999). There is little reason to believe that moral judgments are a marked exception to these general rules.

Third, asking people to consciously explain their preferences, judgments, and decisions can often change them. Difficulty in consciously justifying a particular decision can lead people to change it, sometimes leading to less satisfying or less optimal outcomes (Wilson and LaFleur, 1995; Wilson and Schooler, 1991). Decisions naturally made automatically or unconsciously are altered by reasoning about them deliberately, suggesting that the deliberate reasoning paradigm developed by Kohlberg may substantially alter moral judgments rather than systematically measure them.

Finally, there is, at best, only a weak relationship between moral reasoning and moral action. Children's attitudes toward cheating, for example, do not predict their actual likelihood of cheating (Corey, 1937; Hartshorn and May, 1932). Even when moral reasoning is correlated with moral action, the correlations are weak and appear to be almost completely explained by covariation with intelligence (Haidt, 2001). Low IQ is related to less impulse control and more negative morality, which are manifested in higher rates of crime and violence. Controlling for intelligence renders the relationship between moral reasoning and moral action weak, at best, and nonexistent, at worst.

While there is no question that people engage in moral reasoning, and that moral reasoning has the potential to alter moral judgment, these results suggest that moral reasoning in everyday life is unlikely to be the critical cause of moral judgments, but instead suggest that moral judgments may be guided by the automatic evaluations described earlier. Indeed, this possibility is explicitly proposed by Haidt (2001; see also Kagan, 1984), who argues that intuitionism characterizes moral judgment much better than rationalism. On this model, moral judgments are

based upon rapid and automatic emotional responses to morally relevant stimuli (i.e., moral intuitions), and moral reasoning is a post hoc explanation or justification of these emotional reactions. Moral intuition, then, is "the sudden appearance in consciousness of a moral judgment, including an affective valence (good–bad, like–dislike), without any conscious awareness of having gone through steps of searching, weighting evidence, or inferring a conclusion" (Haidt, 2001, p. 818).

To directly experience this intuition-based model, momentarily consider how you would feel about eating your pet dog after its accidental death. You will likely have an emotional reaction—almost certainly a strong and immediate one—to the mere thought of such a meal, and quickly conclude that it would be wrong to turn your Doberman into dinner. What is interesting, however, is that you might be hard pressed to explain exactly why it is wrong. Indeed, participants in one experiment who were asked to provide logical reasons to support their negative reactions to a variety of offensive actions (e.g., passionate kissing between a brother and sister, cleaning a toilet with the national flag) had considerable difficulty doing so. Nevertheless, these same participants remained steadfast that such actions are universally wrong (Haidt *et al.*, 1993). What is more, the extent to which participants believed they would be bothered by witnessing such acts predicted their moral judgments more strongly than their beliefs about the harmful consequences of such acts. Being unable to justify one's moral judgments doesn't change them so much as it simply leaves people "morally dumbfounded," highlighting the differential importance of affective and rational components to moral judgment (Haidt and Hersh, 2001; Murphy *et al.*, 2000).

These studies capitalize on preexisting affective reactions to demonstrate their importance in moral judgment, but affective responses to neutral objects can also be activated by simply asking people to adopt postures associated with approach or avoidance. For example, people evaluate unfamiliar Chinese ideographs more favorably when simultaneously pulling up on a table (i.e., arm flexion, consistent with approach movements) than when pushing down on a table (i.e., arm extension, consistent with avoidance movements; Cacioppo *et al.*, 1993). When evaluating people, similar positive impressions produce halo effects that also encompass moral evaluations—those who are liked, for example, are also perceived to be kind (Dion *et al.*, 1972). Even affective states that are unrelated to an ethical event can influence perceptions of morality such that ancillary positive emotions can lead to more positive moral evaluations than ancillary negative emotions (Van den Bos, 2003).

Perhaps the strongest existing evidence for an affective-based model of moral judgment, however, comes from the strong correlational and empirical link between emotions and moral actions. For example, true psychological altruism—behaving in a manner to benefit others without regard for one's own welfare—appears to occur only when a person can empathize with, and simultaneously experience the emotional reactions of, a person in distress (Batson, 1987). In one experiment,

those led to empathize with a person receiving painful electric shocks were willing to trade places and receive the shocks themselves if given a choice, even if given an easy opportunity to escape from the uncomfortable situation. Those who are not led to empathize with a person in need do not engage in similar altruism (Batson *et al.*, 1983, 1995).

Related conclusions also come from the disturbing descriptions of clinical psychopaths who show no decrement in reasoning abilities but generally do not experience emotional reactions to arousing stimuli, especially negative stimuli (Cleckley, 1955; Hare, 1993). Psychopaths do not feel sympathy for the suffering of others, do not feel remorse for inflicting pain on others, and do not feel embarrassment or shame when condemned by others. Psychopaths can recognize the consequence of their harmful actions, but they experience little or no inhibition from engaging in them. The presence of affective reactions therefore appears to be the critical determinant of moral action, and its absence the critical determinant of immoral action.

Collectively, these results suggest a repositioning of deliberate reasoning in the chain of moral judgment, as rationalist models appear to have placed the cart before the horse. Affective reactions to morally-relevant stimuli appear to occur automatically, creating a moral intuition that then guides subsequent moral reasoning, rather than the other way around. Given this causal sequence, it is now clear why ideological opponents find it so easy to derive what they perceive to be compelling evidence in support of their particular position from the exact same evidence. Automatic evaluations produce moral reasoners who are not empiricists reasoning dispassionately about a particular issue, but motivated partisans seeking justification for a preexisting intuition. The inherent ambiguity in almost any partisan issue is likely to ensure that people seeking supportive evidence for one position over another are likely to find some (Lord *et al.*, 1979), producing opposing positions that partisans each erroneously believe are a direct product of compelling rational arguments. Part of a recent newspaper headline on disagreements between the United States and Korea captures this experience well: "In Korean standoff, both sides claim reason" ("How U.S.," 2003). But arguing that the opposing side is unreasonable or illogical therefore completely misses the point. Egocentric ethics are not based on reason, but emotion.

CONCLUSIONS AND RECOMMENDATIONS

We have argued that egocentric biases in ethical judgments stem from three basic psychological processes. First, people are automatically inclined to interpret their perceptions egocentrically. Second, people are automatically inclined to evaluate those egocentric interpretations as good or bad, positive or negative, threatening or supporting. Finally, moral judgments about fairness and unfairness

are based upon these automatic evaluative responses. The unconscious and automatic nature of the first two steps in this process explains why one's own egocentric ethics are not perceived to be biased but relatively objective, and therefore why those who render opposing ethical judgments are perceived to be self-interested, stupid, or both.

More important, however, this model helps to explain why egocentric ethical judgments have proven so difficult to overcome. Researchers attempting to reduce conflict and bias have focused on altering partisans' cognitions by presenting them with the opposing sides' arguments (Lord *et al.*, 1979), by asking participants to generate the opposing sides' arguments themselves (Babcock *et al.*, 1996; see Babcock and Loewenstein, 1997), by encouraging full disclosure of conflicts of interest (Cain *et al.*, 2003), by having participants read about the impact and consequences of self-serving biases (Babcock *et al.*, 1996; see Babcock and Loewenstein, 1997), or by providing financial incentives for accuracy (Babcock *et al.*, 1995; Loewenstein *et al.*, 1993). These interventions have been completely ineffective or even counterproductive, sometimes producing more sharply polarized positions. Indeed, in one recent simulated negotiation on overfishing of the world's oceans, participants who represented fishing associations with competing concerns actually behaved *more selfishly* after being asked to adopt the perspective of other group members, compared to those not asked to think beyond their own egocentric perspective (Epley *et al.*, in press a). Follow-up analyses indicated that thinking about opponents' thoughts induced cynical, self-interested attributions of others' intentions that actually served to increase selfish behavior rather than to decrease it.

At present, the only effective debiasing strategies for egocentric ethics are to intervene before people have even developed a perspective to bias their judgments, or to make disputants actively generate and focus on the weaknesses in their own case (see Babcock *et al.*, 1996). Recall that simply assigning people—at random—to role-play a plaintiff versus defendant is sufficient to induce egocentric biases, but asking them to read the evidence for both sides *before* being assigned to a position effectively eliminates those biases (Babcock *et al.*, 1995). Social roles fundamentally alter people's perspectives, and therefore their perceptions. Once a person is given a particular perspective on the world, it appears inevitable that this perspective will influence one's judgments, behavior, and moral reasoning.

The model we have proposed has little trouble explaining such findings, however, as rational arguments will do little to alter judgments based on affective reactions. Research on attitudes and persuasion shows that attitudes formed through affective mechanisms can be changed most effectively by strategies intended to alter those affective reactions, while attitudes formed through cognitive mechanisms are relatively unaffected by altering one's affective reactions (Edwards and von Hippel, 1995; Fabrigar and Petty, 1999). What is more, affective reactions are more stable and change more slowly than cognitions, meaning that affective reactions

linger even after one's thoughts have changed substantially (Gilbert *et al.*, 1995). Manipulating participants' cognitions about partisan issues may temporarily alter their reported attitudes, but because the underlying affective reaction remains unchanged, those altered attitudes quickly "rebound" to their initial partisan positions (Lord *et al.*, 1979). Convincing participants to think about and listen to the weaknesses in their own case (Babcock *et al.*, 1996) may have been successful in reducing egocentric biases precisely because it created negative emotions about one's own perspective. Effective strategies for altering egocentric ethical judgments are therefore likely to be primarily affective in nature. As Jonathan Swift suggested, "You cannot reason a person out of a position he did not reason himself into in the first place."

Admittedly, however, we must end this paper on something of a flat note, as it is currently unclear which specific affective manipulations are likely to prove effective in reducing egocentric biases in ethical judgments. Specific prescriptions for reducing conflict must therefore wait for an empirical postscript. For now, we hope it is sufficient to suggest what egocentric biases in ethical judgments are not. Contrary to the opinions of those involved in partisan disputes, differences in moral judgments between groups are not always the result of stubbornness, stupidity, or blatant self-interest. In these cases, disagreements are not the product of mental shortcomings that can be overcome if only one shouts out his or her own arguments loudly enough. The differences of opinion run deeper, at an automatic, unconscious, and unintentional level. This message may not reduce the differences of opinion between partisan groups, but it might be enough to reduce the cynical attributions that produce anger and aggression between them.

ACKNOWLEDGMENTS

REFERENCES

Ambady, N., Bernieri, F., and Richeson, J. A. (2000). Towards a histology of social behavior: Judgmental accuracy from thin slices of behavior. In Zanna, M. P. (ed.), *Advances in Experimental Social Psychology*, Vol. 32, Academic Press, San Diego, pp. 201–271.

Babcock, L., and Loewenstein, G. (1997). Explaining bargaining impasse: The role of self-serving biases. *J. Econ. Perspect.* 11: 109–126.

Babcock, L., Loewenstein, G., and Issacharoff, S. (1996). *Debiasing Litigation Impasse.* Unpublished manuscript, Carnegie Mellon University.

Babcock, L., Loewenstein, G., Issacharoff, S., and Camerer, C. (1995). Biased judgments of fairness in bargaining. *Am. Econ. Rev.* 85: 1337–1343.

Banse, R. (1999). Automatic evaluation of self and significant others: Affective priming in close relationships. *J. Soc. Pers. Relat.* 16: 803–821.

Bargh, J. A. (1994). The four horsemen of automaticity: Awareness, efficiency, intention, and control in social cognition. In Wyer, R. S., and Srull, T. K. (eds.), *Handbook of Social Cognition*, 2nd ed., Erlbaum, Hillsdale, NJ, pp. 1–40.

Bargh, J. A., Chaiken, S., Govender, R., and Pratto, F. (1992). The generality of the automatic attitude activation effect. *J. Pers. Soc. Psychol.*, 62: 893–912.

Bargh, J. A., Chaiken, S., Raymond, P., and Hymes, C. (1996). The automatic evaluation effect: Unconditional automatic attitude activation with a pronunciation task. *J. Exp. Soc. Psychol.* 32: 185–210.

Batson, C. D. (1987). Prosocial motivation: Is it ever truly altruistic? In Berkowitz, L. (ed.), *Advances in Experimental Social Psychology*, Vol. 20, Academic Press, New York, pp. 65–122.

Batson, C. D., Klein, T. R., Highberger, L., and Shaw, L. L. (1995). Immorality from empathy-induced altruism: When compassion and justice conflict. *J. Pers. Soc. Psychol.* 68: 1042–1054.

Batson, C. D., O'Quinn, K., Fulty, J., Vanderplass, M., and Isen, A. M. (1983). Influence of self-reported distress and empathy on egoistic versus altruistic motivation to help. *J. Pers. Soc. Psychol.* 45: 706–718.

Brinol, P., and Petty, R. E. (2003). Overt head movements and persuasion: A self-validation analysis. *J. Pers. Soc. Psychol.* 84: 1123–1139.

Cacioppo, J. T., Priester, J. R., and Berntson, G. G. (1993). Rudimentary determinants of attitudes. II: Arm flexion and extension have differential effects on attitudes. *J. Pers. Soc. Psychol.* 65: 5–17.

Cain, D., Moore, D., and Loewenstein, G. (2003, September). The dirt on coming clean: Perverse effects of disclosing conflicts of interest. *Paper Presented at the Conference on Conflicts of Interest*, Pittsburgh, PA.

Chartrand, T. L., and Bargh, J. (1999). The chameleon effect: The perception–behavior link and social interaction. *J. Pers. Soc. Psychol.* 76: 893–910.

Chen, M., and Bargh, J. A. (1999). Consequences of automatic evaluation: Immediate behavioral predispositions to approach or avoid the stimulus. *Pers. Soc. Psychol. Bull.* 25: 215–224.

Cleckley, H. (1955). *The Mask of Insanity*. C.V. Mosby, St. Louis.

Corey, S. M. (1937). Professed attitudes and actual behavior. *J. Educ. Psychol.* 28: 271–280.

Dasgupta, N., and Greenwald, A. G. (2001). On the malleability of automatic attitudes: Combating automatic prejudice with images of admired and disliked individuals. *J. Pers. Soc. Psychol.* 81: 800–814.

De Houwer, J., Hermans, D., and Spruyt, A. (2001). Affective priming of pronunciation responses: Effects of target degradation. *J. Exp. Soc. Psychol.* 37: 85–91.

Devine, P. G. (1989). Stereotypes and prejudice: Their automatic and controlled components. *J. Pers. Soc. Psychol.* 56: 5–18.

Dijksterhuis, A., and van Knippenberg, A. (1998). The relation between perception and behavior, or how to win a game of Trivial Pursuit. *J. Pers. Soc. Psychol.* 74: 865–877.

Dion, K., Berscheid, E., and Walster, E. (1972). What is beautiful is good. *J. Pers. Soc. Psychol.* 24: 207–213.

Duckworth, K. L., Bargh, J. A., Garcia, M., and Chaiken, S. (2002). The automatic evaluation of novel stimuli. *Psychol. Sci.* 13: 513–519.

Dunning, D. (1999). A newer look: Motivated social cognition and the schematic representation of social concepts. *Psychol. Inq:* 10: 1–11.

Edwards, K., and von Hippel, W. (1995). Hearts and minds: The priority of affective versus cognitive factors in person perception. *Pers. Soc. Psychol. Bull.* 21: 996–1011.

Epley, N. (2001). *Mental Correction as Serial, Effortful, Confirmatory, and Insufficient Adjustment*, Unpublished Doctoral Dissertation, Cornell University.

Epley, N., Caruso, E. M., and Bazerman, M. H. (2004). Effects of perspective taking on judgments of fairness and actual behavior. Unpublished raw data.

Epley, N., and Gilovich, T. (2001). Putting adjustment back in the anchoring and adjustment heuristic: Divergent processing of self-generated and experimenter-provided anchors. *Psychol. Sci.* 12: 391–396.

Epley, N., and Gilovich, T. (in press). Are adjustments insufficient? *Pers. Soc. Psychol. Bull.*

Epley, N., Keysar, B., Van Boven, L., and Gilovich, T. (in press a). Perspective taking as egocentric anchoring and adjustment. *J. Pers. Soc. Psychol.*

Epley, N., Morewedge, C., and Keysar, B. (in press b). Perspective taking in children and adults: Equivalent egocentrism but differential correction. *J. Exp. Soc. Psychol.*

Fabrigar, L. R., and Petty, R. E. (1999). The role of the affective and cognitive bases of attitudes in susceptibility to affectively and cognitively based persuasion. *Pers. Soc. Psychol. Bull.* 25: 363–381.

Fazio, R. H. (1989). On the power and functionality of attitudes: The role of attitude accessibility. In Pratkanis, A. R., Breckler, S. J., and Greenwald, A. G. (eds.), *Attitude Structure and Function*, Erlbaum, Hillsdale, NJ, pp. 153–179.

Fazio, R. H., Sanbonmatsu, D. M., Powell, M. C., and Kardes, F. R. (1986). On the automatic activation of attitudes. *J. Pers. Soc. Psychol.* 50: 229–238.

Ferguson, M. J., and Bargh, J. A. (2004). *Liking is for Doing: Effects of Goal-Pursuit on Automatic Evaluation*. Unpublished manuscript, Cornell University.

Gilbert, D. T. (1989). Thinking lightly about others: Automatic components of the social inference process. In Uleman, J. S., and Bargh, J. A. (eds.), *Unintended Thought*, Guilford Press, New York, pp. 189–211.

Gilbert, D. T., Giesler, R. B., and Morris, K. A. (1995). When comparisons arise. *J. Pers. Soc. Psychol.* 69: 227–236.

Gilbert, D. T., and Gill, M. J. (2000). The momentary realist. *Psychol. Sci.* 11: 394–398.

Gilovich, T., Medvec, V. H., and Savitsky, K. (2000). The spotlight effect in social judgment: An egocentric bias in estimates of the salience of one's own actions and appearance. *J. Pers. Soc. Psychol.* 78: 211–222.

Gilovich, T., and Savitsky, K. (1999). The spotlight effect and the illusion of transparency: Egocentric assessments of how we're seen by others. *Curr. Dir. Psychol. Sc.* 8: 165–168.

Greenwald, A. G., and Banaji, M. R. (1995). Implicit social cognition. *Psychol. Rev.* 102: 4–27.

Greenwald, A. G., Klinger, M. R., and Schuh, E. S. (1995). Activation by marginally perceptible ("subliminal") stimuli: Dissociation of unconscious from conscious cognition. *J. Exp. Psychol.: Gen.* 124: 22–42.

Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychol. Rev.* 108: 814–835.

Haidt, J., and Hersh, M. (2001). Sexual morality: The cultures and emotions of conservatives and liberals. *J. Appl. Soc. Psychol.* 31: 191–221.

Haidt, J., Koller, S., and Dias, M. (1993). Affect, culture, and morality, or is it wrong to eat your dog? *J. Pers. Soc. Psychol.* 65: 613–628.

Hare, R. D. (1993). *Without Conscience*, Pocket Books, New York.

Hartshorn, H., and May, M. (1932). *Studies in the Nature of Character: Studies in the Organization of Character*, Vol. 3, MacMillan, New York.

Hermans, D., Baeyens, F., and Eelen, P. (1998). Odors as affective processing context for word evaluation: A case of cross-modal affective priming. *Cogn. Emotion* 12: 601–613.

Hermans, D., Spruyt, A., and Eelen, P. (2003). Automatic affective priming of recently acquired stimulus valence: Priming at SOA 300 but not at SOA 1000. *Cogn. Emotion* 17: 83–99.

Higgins, T. E., Rholes, W. S., and Jones, C. R. (1977). Category accessibility and impression formation. *J. Exp. Soc. Psychol.* 13: 141–154.

How U.S., North Korea turned broken deals into a standoff (2003, March 5). *The Wall Street J.* pp. A1, A10.

Kagan, J. (1984). *The Nature of the Child*, Basic Books, New York.

Kant, I. (1964). *Groundwork of the Metaphysics of Morals*, Paton, H. J. (trans.), Harper and Row, New York. (Original work published in 1785.)

Keysar, B., and Barr, D. J. (2002). Self-anchoring in conversation: Why language users don't do what they "should." In Gilovich, T., Griffin, D., and Kahneman, D. (eds.), *Heuristics and Biases: The Psychology of Intuitive Judgment*, Cambridge University Press, Cambridge, pp. 150–166.

Keysar, B., Barr, D. J., Balin, J. A., and Brauner, J. S. (2000). Taking perspective in conversation: The role of mutual knowledge in comprehension. *Psychol. Sci.* 11: 32–38.

Keysar, B., Barr, D. J., and Horton, W. S. (1998). The egocentric basis of language use: Insights from a processing approach. *Curr. Dir. Psychol. Sci.* 7: 46–50.

Klar, Y., and Giladi, E. E. (1997). No one in my group can be below the group's average: A robust positivity bias in favor of anonymous peers. *J. Pers. Soc. Psychol.* 73: 885–901.

Klar, Y., and Giladi, E. E. (1999). Are most people happier than their peers, or are they just happy? *Pers. Soc. Psychol. Bull.* 25: 585–594.

Kohlberg, L. (1969). Stage and sequence: The cognitive-developmental approach to socialization. In Goslin, D. A. (ed.), *Handbook of Socialization Theory and Research*, Rand McNally, Chicago, pp. 347–480.

Krosnick, J. A., Betz, A. L., Jussim, L. J., and Lynn, A. R. (1992). Subliminal conditioning of attitudes. *Pers. Soc. Psychol. Bull.* 18: 152–162.

Kruger, J. (1999). Lake Wobegon be gone! The "below-average effect" and the egocentric nature of comparative ability judgments. *J. Pers. Soc. Psychol.* 77: 221–232.

Kunda, Z. (1990). The case for motivated reasoning. *Psychol. Bull. 108*: 480–498.

Lakin, J. S., and Chartrand, T. L. (2003). Using nonconscious behavioral mimicry to create affiliation and rapport. *Psychol. Sci.* 14: 334–339.

Loewenstein, G., Issacharoff, S., Camerer, C., and Babcock, L. (1993). Self-serving assessments of fairness and pretrial bargaining. *J. Leg. Stud.* 22: 135–159.

Lord, C. G., Ross, L., and Lepper, M. R. (1979). Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence. *J. Pers. Soc. Psychol.* 37: 2098–2109.

Luksa, F. (2003, June 14). Auction can't heal wounds for Bonds home run ball. *The Mercury News*. Retrieved July 27, 2003, from http://www.bayarea.com/mld/mercurynews/sports/6088835.htm

Messick, D. M., and Sentis, K. (1983). Fairness, preference, and fairness biases. In Messick, D. M., and Cook, S. (eds.), *Equity Theory: Psychological and Sociological Perspectives*, Praeger, New York, pp. 61–94.

Murphy, S., Haidt, J., and Björklund, F. (2000). *Moral Dumbfounding: When Intuition Finds No Reason*. Unpublished manuscript, University of Virginia.

Neely, J. H. (1977). Semantic priming and retrieval from lexical memory: Roles of inhibitionless spreading activation and limited-capacity attention. *J. Exp. Psychol.: Gen.* 106: 226–254.

Nickerson, R. S. (1999). How we know—and sometimes misjudge—what others know: Imputing one's own knowledge to others. *Psychol. Bull.* 125: 737–759.

Nisbett, R. E., and Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychol. Rev.* 84: 231–259.

Osgood, C. E., Suci, G. J., and Tannenbaum, P. H. (1957). *The Measurement of Meaning*, University of Illinois Press, Urbana, IL.

Pelham, B. W., Mirenberg, M. C., and Jones, J. T. (2002). Why Susie sells seashells by the seashore: Implicit egotism and major life decisions. *J. Pers. Soc. Psychol.* 82: 469–487.

Piaget, J. (1965). *The Moral Judgment of the Child*, Gabain, M. (trans.), Free Press, New York. (Original work published 1932.)

Prentice, D. A., and Miller, D. T. (1993). Pluralistic ignorance and alcohol use on campus: Some consequences of misperceiving the social norm. *J. Pers. Soc. Psychol.* 64: 243–256.

Pronin, E., Puccio, C., and Ross, L. (2002). Understanding misunderstanding: Social psychological perspectives. In Gilovich, T., Griffin, D., and Kahneman, D. (eds.), *Heuristics and Biases: The Psychology of Intuitive Judgment*, Cambridge University Press, Cambridge, pp. 636–665.

Robinson, R., Keltner, D., Ward, A., and Ross, L. (1995). Actual versus assumed differences in construal: "Naïve realism" in intergroup perceptions and conflict. *J. Pers. Soc. Psychol.* 68: 404–417.

Ross, L., Green, D., and House, P. (1977). The "false consensus effect": An egocentric bias in social perception and attribution processes. *J. Exp. Soc. Psychol.* 13: 279–301.

Schneirla, T. (1959). An evolutionary and developmental theory of biphasic processes underlying approach and withdrawal. In Jones, M. (ed.), *Nebraska Symposium on Motivation*, University of Nebraska Press, Lincoln, pp. 27–58.

Sollberger, B., Reber, R., and Eckstein, D. (2003). Musical chords as affective priming context in a word-evaluation task. *Music Percept.* 20: 263–282.

Thompson, L., and Loewenstein, G. (1992). Egocentric interpretations of fairness and interpersonal conflict. *Organ. Behav. Hum. Decis. Process.* 51: 176–197.

Tversky, A., and Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science* 185: 1124–1131.

Uleman, J. S., Newman, L. S., and Moskowitz, G. B. (1996). People as flexible interpreters: Evidence and issues from spontaneous trait inference. In Zanna, M. P. (ed.), *Advances in Experimental Social Psychology*, Vol. 28, Academic Press, San Diego, pp. 211–279.

Van den Bos, K. (2003). On the subjective quality of social justice: The role of affect as information in the psychology of justice judgments. *J. Pers. Soc. Psychol.* 85: 482–498.

Vorauer, J., and Ross, M. (1999). Self-awareness and feeling transparent: Failing to suppress one's self. *J. Exp. Soc. Psychol.* 35: 415–440.

Watercutter, A. (2002, October 18). Fighting over Bonds' baseballs. *CBS News.* Retrieved July 27, 2003, from http://www.cbsnews.com/stories/2002/08/07/national/main517743.shtml

Wegner, D. M., and Bargh, J. A. (1998). Control and automaticity in social life. In Gilbert, D., Fiske, S. T., and Lindzey, G. (eds.), *Handbook of Social Psychology*, 4th ed., McGraw-Hill, New York, pp. 446–496.

Wells, G. L., and Petty, R. E. (1980). The effects of overt head movements on persuasion: Compatibility and incompatibility of responses. *Basic Appl. Soc. Psychol.* 1: 219–230.

Wilensky, A. E., Schafe, G. E., and LeDoux, J. E. (2000). The amygdala modulates memory consolidation of fear-motivated inhibitory avoidance learning but not classical fear conditioning. *J. Neurosci.* 20: 7059–7066.

Wilson, T. D., and Brekke, N. (1994). Mental contamination and mental correction: Unwanted influences on judgments and evaluations. *Psychol. Bull.* 116: 117–142.

Wilson, T. D., and LaFleur, S. J. (1995). Knowing what you'll do: Effects of analyzing reasons on self-prediction. *J. Pers. Soc. Psychol.* 68: 21–35.

Wilson, T. D., and Schooler, J. W. (1991). Thinking too much: Introspection can reduce the quality of preferences and decisions. *J. Pers. Soc. Psychol.* 60: 181–192.

Wilson, T. D., and Stone, J. (1985). Limitations of self-knowledge: More on telling more than we can know. In Shaver, P. (ed.), *Review of Personality and Social Psychology*, Vol. 6, Sage, New York, pp. 167–183.

Wilstein, S. (2003, June 26). Bonds' No. 73 ball: A story of greed. *NBC Sports.* Retrieved July 27, 2003, from http://stacks.msnbc.com/news/931529.asp

Wittenbrink, B., Judd, C. M., and Park, B. (2001). Spontaneous prejudice in context: Variability in automatically activated attitudes. *J. Pers. Soc. Psychol.* 81: 815–827.

Zajonc, R. B. (1998). Emotions. In Gilbert, D. T., Fiske, S. T., and Lindzey, G. (eds.), *Handbook of Social Psychology*, Vol. 1, McGraw-Hill, New York, pp. 591–632.