RUNNING HEAD: BOUNDED ETHICALITY

Bounded Ethicality as a Psychological Barrier to Recognizing Conflicts of Interest

Dolly Chugh
Max H. Bazerman
Mahzarin R. Banaji

Harvard University

Address correspondence to:
Dolly Chugh
Harvard Business School
Boston, MA 02163
Email: dchugh@hbs.edu

The authors are grateful for the feedback of the highly engaged participants in the NSF/CBI Conference on Conflict of Interest, hosted by Carnegie Mellon University. We especially thank Ann Tenbrunsel and Don Moore for their useful reviews, as well as Matt Cronin, Bill Keech, Scott Kim, and Kent Womack for their thoughtful, written reactions to our conference presentation.

> But there is a more subtle question of conflict of interest that derives directly from human bounded rationality. The fact is, if we become involved in a particular activity and devote an important part of our lives to that activity, we will surely assign it a greater importance and value than we would have prior to our involvement with it.
>
> It's very hard for us, sometimes, not to draw from such facts a conclusion that human beings are rather dishonest creatures … Yet most of the bias that arises from human occupations and preoccupations cannot be described correctly as rooted in dishonesty – which perhaps makes it more insidious than if it were.
>
> - Herbert A. Simon, 1983, pp. 95-96

Herbert Simon's perspective (1983) is broadly compatible with Moore, Loewenstein, Tanlu, and Bazerman's (2003) recent research on the psychological aspects of conflict of interest in the context of auditor independence. Moore et al. (2003) focuses primarily on the work on self-serving interpretations of fairness. The current work broadens this theme, and develops a conceptual framework for understanding how unchecked psychological processes work against an objective assessment and allow us to act against personal, professional and normative expectations when conflicts of interest exist.

Our work pursues a more comprehensive treatment of Simon's informal notion through an integration of three critical psychological insights of the past century. We begin with Simon's own insight of bounded rationality, continue with subsequent insights offered in the work of Kahneman and Tversky regarding deviations from rationality, and then consider what we know today about the limitations of the conscious mind. In our

assessment, these three literatures together provide robust support for the view that conflict of interest that is not limited to explicit dishonesty. Rather, unconscious acts of ethically questionable behavior are more prevalent, more insidious, and as such, more in need of attention. The strands of these three insights weave together to form a powerful thread connecting what we know about basic human perception to cognitive, social, and ultimately, ethical consequences. Thus, we develop the argument that the computational bounds on human cognition stretch further than previously assumed – they can influence the quality of ethical judgments, leading us to extend Simon's phrase 'bounded rationality' to consider the possibility and consequence of 'bounded ethicality'. Bounded rationality refers to the limits on the quality of general decision-making, and bounded ethicality is a strand that is used to refer to the limits on the quality of decision-making with ethical import. In this chapter, we focus on the nature of bounded ethicality, and its psychological implications for recognizing conflicts of interest.

We propose that bounded ethicality places a critical constraint on the quality of decision-making. We focus on one consequence of bounded ethicality, the limitation in recognizing the ethical challenge inherent in a situation or decision, such as conflicts of interest. Specifically, we argue that individuals view themselves as moral, competent, and deserving, and this view obstructs their ability to see and recognize conflicts of interest when they occur. Thus, ethicality is not bounded in unpredictable or non-systematic ways, but in systematic ways that unconsciously favor this particular vision of the self in our judgments. The self is an important construct in our argument, and we do not challenge the individual's capacity to recognize conflicts of interest in the abstract, or in the situations facing others, but rather in the situations involving the self.

We argue that conflicts of interests are even more prevalent than the "visible" conflicts traditionally assumed by that term. For example, visible conflicts of interest include the firm that collects both auditing and consulting revenues from the same client, as well as the investment bank that seeks investment banking business from the same companies rated by the firm's equity analyst. In contrast to these visible conflicts of interest, "invisible" conflicts of interests are rarely viewed as conflicts at all. Rather, these situations are opportunities, and even obligations, to demonstrate loyalty and generosity for one's nation, or team, or ethnic group. We argue that these opportunities are, in fact, potential conflicts of interest, and even more so, when practiced by members of majority groups because of the large numbers of people within those groups who benefit.

Three Critical Insights of the Past Century

Simon offered bounded rationality as a "behavioral model (in which) human rationality is very limited, very much bounded by the situation and by human computational powers" (1983, page 34; see also Simon, 1957). Fundamentally, Simon challenged economists' assumption of humans as rational creatures. Boundedness has since come to represent the distinction between economists' normative and psychologists' descriptive views of human decision-making. Thaler (1996), for instance, extended Simon's thinking in describing the three ways in which "Homo Economicus" and "Homo Psychologicus" vary. People are "dumber, nicer, and weaker" than classical economic theory predicts (page 227, 230); that is, human beings have bounded rationality, self-interest, and willpower.

Building on Simon's work, Daniel Kahneman and Amos Tversky delineated the systematic patterns in which human beings demonstrate boundedness. From the 1970s to the present, the field of behavioral decision research has identified the systematic ways in which decision-makers deviate from optimality or rationality in the use of information (Kahneman & Tverksy 1973; 1979). This field has allowed researchers to predict, a priori, how people will make decisions that are inconsistent, inefficient, and based on normatively irrelevant information. The central argument of much of this literature is that people rely on simplifying strategies, or cognitive heuristics (Bazerman, 2002). While heuristics are useful short cuts, they also lead to predictable mistakes (Tversky & Kahneman, 1974). It is the systematic and predictable nature of these biases, and what they reveal about the human mind, that makes them so intriguing to researchers.

The roots of these traditions stretch back to cognitive psychology and basic visual and perceptual processes. Daniel Kahneman's acceptance speech for the Nobel Prize in Economics began, notably, with demonstrations of the primal limitations of our visual perception of lines and colors, followed by an extension of this limitation to more abstract forms of decision-making (2002). Boundedness begins in perception, and extends to cognition. Together, then, the insights of the bounded rationality and heuristics literatures have firmly established the universal computational limitations of the human mind.

In recent years, another important psychological insight has emerged, inviting us to consider boundedness from an even broader point view. That is, we have seen rapid accumulation of evidence both for the limitations of the conscious mind and the power of the unconscious mind. The weight of this insight is demonstrated in the most recent

Handbook of Social Psychology, which included a first-ever chapter about control and automaticity in social life (Bargh and Wegner, 1999). The limitations of the conscious mind are highlighted in Wegner's (2002) analysis of the role of consciousness in human thinking and action. He dramatically demonstrates "the illusion of conscious will" in which human beings not only claim responsibility, but also intention, for actions over which they had exactly no control. In a variety of tasks and contexts, humans tend to attribute their own behavior to premeditated intention, rather than to unconscious processes. Conscious will is consistently given more credit than is due, despite robust evidence about its limitations.

In parallel, the power of the unconscious mind in everyday life has become evident. In a growing, multi-method body of research, automaticity has been found to play some role in virtually every cognitive process studied, and its inevitability has been cleverly termed the "unbearable automaticity of being" (Bargh and Chartrand, 1999). The study of unconsciousness has been made possible by the growing commitment to the use and development of indirect measures (Greenwald and Banaji, 1995). Methodologically, unconscious processes present a challenge to observe directly, necessitating that researchers measure outcomes of those processes that are not directly accessible. Response latency is one of the most commonly used metrics in these methods[1], relying on the relationship between speed of response and strength of unconscious cognitive associations, and can be measured through millisecond-level response times thanks to computer-based tasks. Another important metric is ease of

[1] These have included, though not been limited to, the lexical decision task (LDT; e.g., Macrae, Bodenhausen, Milne, and Jetten, 1994), the sequential priming task (e.g., Fazio et al, 1995), word completions following unobtrusive priming (e.g., Gilbert and Hixon,

recall, which relies on the relationship between the accessibility of a thought and the strength of an unconscious cognitive association, and can be measured by observing how a participant completes a word when only a few letters are shown. Further, by exposing participants to particular stimuli subliminally (known as priming), researchers can compare response times or ease of recall under different conditions, such as stereotypical primes versus counter-stereotypical primes.

From these methods, data have emerged and converged that allows researchers to contrast implicit thinking with explicit thinking. Explicit processes are those of which the decision-maker is aware and can consciously endorse. Implicit processes are those of which the decision-maker is unaware, which are automatic, and which are not necessarily under the control of the decision maker. There is growing evidence that both types of mental processes have an impact on behavior, and growing evidence that we overstate the link between the conscious system and behavior, and understate the link between the unconscious system and behavior (Bargh, 1997; Chugh, in press). It is with this insight that we return to where we began, for a fresh look at bounded rationality.

The Case for Bounded Ethicality

We begin with the well-established knowledge that boundedness and heuristics offer computational speed, critical to the survival of human beings with less than infinite time for decision-making (Dawes, 1976; Bazerman, 2002). This "cognitive" perspective reflects humans' imperfections as statisticians and scientists (Dunning, 1999). In what has been presented as an opposing perspective by some (Dunning, 1999), the "motivational" perspective suggests that individuals' perceptions, judgments, and

---

1991), and the Implicit Association Test (IAT; Greenwald, McGhee, and Schwartz, 1998).

behaviors are biased towards the goal of maintaining self-worth, not just towards the more neutral goals of speed and efficiency. However, we see the two perspectives as complementary, not opposing, in the study of decision-making (see Kunda, 1990). The particular decisions we discuss here, ethical decisions, bring social forces, and thus motivational forces, to bear on decision-making.

So, we accept this motivational perspective as highly relevant to the domain of ethical decision-making and will argue that motivational and social forces are a less studied but important cause of boundedness. But our attention to the motivational perspective should not be interpreted as an abandonment of the cognitive, computational perspective. In fact, we believe both computational limitations and motivation towards self-worth are both at work in the domain of ethical decision-making, consistent with the thread connecting perceptual, cognitive, and social bounds on decision-making. Ethical decisions almost always involve consequences for self and / or others, and it is this social component that brings forth a surge of self-oriented motivations in ethical decision-making. Bounded ethicality represents that subset of bounded rationality situations in which the self is central and therefore, motivation is most likely to play a prominent role.

This particular feature of bounded ethicality brings us back to the roles of consciousness and automaticity in decision-making. In the bounded rationality and heuristics literatures, which emerged from the cognitive perspective, the researchers' assumptions about the limitations of consciousness and the power of the unconscious are neither articulated nor disputed. In the motivational perspective, the drive towards maintaining self-worth is assumed to be unconscious. So, while the existence of unconscious processes may have been assumed by researchers, we attempt here to make

such an assumption explicit, specific, and plausible. In fact, much insight into the nature and source of boundedness, and its role in ethical decision-making, can be achieved by making consciousness and automaticity a focal point of our argument.

The use of bounded rationality to address a particular type of ethical decision-making originated with Banaji and Bhaskar (2000). Arguing against the view that stereotyping is correct and rational, they linked the limitations of human cognition to memory and implicit stereotypes, demonstrating that such limitations lead to ethical failures. These ethical failures "reveal how the interaction of specific social experiences and a boundedly rational cognitive architecture jointly shape thought and behavior" (Banaji and Bhaskar, 2000, page 154). Our notion of bounded ethicality emerges from this perspective, and importantly, picks up on the importance of implicit mental processes.

Specifically, social and ethical situations are particularly likely to trigger bounds on conscious thinking and biases in unconscious thinking, allowing us to more fully describe the richness of Simon's original insight about boundedness and subsequent insight about conflicts of interest. In the remainder of this paper, we propose that bounded ethicality is a critical constraint on the quality of ethical decision-making. We propose that ethicality is bounded in systematic ways that unconsciously favor a particular vision of the self in our judgments. Just as the heuristics and biases tradition took bounded rationality and specified a set of systematic, cognitive deviations from full rationality, we endeavor to take bounded ethicality and specify systematic, motivational deviations from full ethicality. Similarly to the bounded rationality tradition, bounded

ethicality is characterized by computational speed that eases decision-making complexity, but in addition, motivational forces are at work as well.

In the bounded ethicality model, the self processes work, unconsciously, to protect a particular view and this view bounds ethical decision-making. Ethical decisions are biased by a stubborn view of oneself as moral, competent, and deserving, and thus, not susceptible to conflicts of interest. To the self, a view of morality ensures that the decision-maker resists temptations for unfair gain; a view of competence ensures that the decision-maker qualifies for the role at hand; and, a view of deservingness ensures that one's advantages arise from one's merits. An ethical blind spot emerges as decision-makers view themselves as moral, competent, and deserving, and thus assume conflicts of interest are non-issues. Thus, conflicts, particularly the Simon-esque variety mentioned at the start of this chapter, are unlikely to even be recognized as conflicts by the person at risk. The view of self that is preserved through bounded ethicality represents, in fact, exactly those qualities that one would require in order to be immune from conflicts of interest. In addition, it is this view of the self that prevents the decision-maker from even recognizing the ethical situation in which he finds himself. And yet, ironically, a decision-maker is made more susceptible to conflicts of interest because of the persistence of his or her self-image.

Further, the evidence suggests that we are both particularly unaware of data that contradicts this view of ourselves, and worse yet, particularly unaware of that unawareness. This unawareness is fundamental to the notion of the "totalitarian ego" (Greenwald, 1980). The ego (loosely equivalent to our use of "self" in this chapter) is an organization of knowledge, while the totalitarian ego displays three biases that

correspond to the thought control and propaganda devices of a totalitarian political system. In a totalitarian political system, "it is necessary to remember that events happened in the desired manner … and if it is necessary to rearrange one's memories or to tamper with written records, then it is necessary to forget that one has done so" (Orwell, 1949, p. 176). Similarly, the ego actively tampers and rearranges self-knowledge so as to ensure that a certain view is maintained, but retains no conscious belief that such tampering has taken place (Greenwald, 1980). Individuals are unaware of their unawareness. The limitations of the conscious mind are thus critical to the success of the totalitarian ego. Memory itself is distorted towards recollection of events "relevant to me" versus "not relevant to me," as well as a positive construal of those events.

The "egocentric ethics" (Epley and Caruso, in press) of the totalitarian ego, combined with the power of the unconscious mind, make conflicts of interest difficult to recognize. In the following section, we consider the susceptibility of individuals to conflicts of interest due to the persistent views of self as moral, competent, and deserving.

Self as Moral. People believe that they are more honest, trustworthy, ethical and fair than others (Baumhart, 1968; Tenbrunsel, 1998; Messick and Bazerman, 1996). We give ourselves more credit for our good behaviors and take less responsibility for our moral lapses than others would be likely to do (Messick and Bazerman, 1996**).** We are motivated to see ourselves as ethical, and rate ourselves as more ethical than the average person (Tenbrunsel, 1998). When we engage in ethically questionable behavior, we often justify it as self-defense (Shapiro, 1991).

However, research suggests that humans continue to maintain an "illusion of objectivity" (Armor, 1998). Across a series of five studies, participants consistently rated their own objectivity higher than that of their average peer. In fact, approximately 85% of the participants believed themselves to be more objective than their average peer. Given the statistical improbability of 85% of participants being above their group's average, the illusion of objectivity is evident. And, participants were not simply seeing themselves as relatively less subjective than their peers. Participants' ratings of their own objectivity reflected a belief that they are not only viewing themselves as more objective relative to others, but also as objective in the absolute. These data suggest that at least some percentage of human beings must be perceiving the world less accurately than they believe they are. Yet, the illusion is also persistent, as participants retained their belief in their own objectivity, even when made aware of the phenomenon taking place.

In one study, researchers explored the vulnerability of one's own objectivity by studying how perceptions of the world depend fundamentally on how the perception favors or disfavors the self (Kronzon and Darley, 1999). Participants observed an ethically questionable act of deception in a videotaped negotiation. Partisans who were randomly allied with the victim of the ethically questionable behavior perceived the act as more reprehensible than did either partisans randomly allied with the perpetrator or neutral observers. Despite the influence that the situation has on perceptions, research suggests that people underestimate differences in construal, and thus are overconfident in the objectivity of their predictions of the behavior of both themselves and others (Griffin et al., 1990). This bias exaggerates a conflict of interest as the decision-maker retains an unrealistic confidence in his or her perception of data about the situation.

In another study, researchers explored the conditions under which such unrealistically positive beliefs are maintained or loosened (Wade-Benzoni, Thompson, and Bazerman, 2003). Self-assessment of environmental sensitivity was found to depend on how much ambiguity surrounds the self-assessment. Specifically, individuals maintain unrealistically positive beliefs about their degree of environmental sensitivity when their self-evaluation is difficult to disconfirm, but possess more realistic assessments of themselves when they are constrained by the objectivity of the evaluation (consistent with earlier work, e.g., Allison, Messick, & Goethals, 1989; Kunda, 1990)}.

For example, assessments of general beliefs such as one's awareness of, concern for, understanding of, and interest in environmental issues and problems are difficult to confirm or disconfirm. In contrast, assessments of how well one performs on specific activities such as recycling, donating money to environmental organizations, and using energy-saving light bulbs can be checked against objective measures. If individuals define their environmental sensitivity in terms of general (not easily confirmable) behaviors instead of specific (objectively measurable) behaviors, their self-evaluations are likely to be inflated. Again, human beings maintain the illusion of objectivity, thus putting them at risk for not recognizing a conflict of interest when it presents itself.

Overall, this pattern of self-enhancement may provide people with an easy way out of engaging in more responsible societal behaviors. Thus, when the auditor hears of the Moore et al. (2003) concern that their audit might be biased in ways that they are not even aware, the auditor feels that her objectivity will make him or her immune from the problems. Babcock and Loewenstein (1997) demonstrated that even individuals' interpretations of these self-serving biases are self-serving. Study participants were

taught about these biases, and the participants demonstrated a clear understanding of the bias by shifting their expectations of others' objectivity. Yet, the participants maintained a commitment to their own lack of bias, even while adjusting their expectations of the objectivity of others.

The bias toward believing that we are more objective than reality dictates leads us to the conclusion that our objectivity will keep conflicts of interests from influencing our judgment. In fact, in 2000, this is exactly the argument that Joseph Berardino, the CEO of Arthur Andersen, made while testifying before the SEC Commission. He argued that the professionalism and objectivity of professional auditors solved the issue of auditor independence. The SEC commissioners appeared to be influenced by this argument, despite its inconsistency with psychological research. The self-as-objective argument carried the day, the SEC failed to act sufficiently, and the lack of auditor independence contributed to many corporate failures. Professionals commonly sell their professionalism as immunity against being affected by conflict of interest. We believe that professionalism provides only partial immunity against intentional corruption, and little immunity from the unconscious processes that lead decision-makers to succumb to conflicts of interest.

We also extend our idea of appropriate ethical behavior to others. Negotiators' expectations that their opponents will deceive them may be influenced by their own tendency to deceive. Tenbrunsel (1998) varied the amount of money participants could win for negotiating successfully. Participants who could win $100 expected significantly more deception from their opponents and were significantly more likely to deceive than those who could only win $1. However, participants' expectations of their opponents'

deception depended both on their own level of temptation, as well the level of temptation of their opponents.

Individuals' perceptions of a situation can vary dramatically, even when given identical information, depending on their roles. This difference occurs because individuals begin with their preference for a particular outcome, as motivated by self-interests, and then justify this view on the basis of fairness through a biased perspective on what attributes constitute fairness (Messick and Sentis, 1983). The ethical failure is not in the commitment to fairness but in the biased interpretation of information (Diekmann et al., 1997; Messick and Sentis, 1983).

These limitations of the conscious mind are described by Jon Haidt (2001) as the "emotional dog and rational tail", in which "moral judgment is caused by quick moral intuitions, and is followed (when needed) by slow, ex-post facto moral reasoning." The moral reasoning essentially occurs after the fact. This sequence suggests that "automatic egocentrism" precedes an evaluative moral judgment (Epley and Caruso, in press).

And, so, in such a tail-wagging-the-dog scenario, the view of oneself as moral is, at best, irrelevant (since morality occurs after the fact), and at worst, a psychological liability (since morality is rigged in our favor). The belief that the self is moral leads us to believe that conflicts of interests will not distort our judgment, thus bounding our ability to recognize the conflict when it occurs.

Self as Competent

People perceive themselves as being better than others on a variety of desirable attributes (Messick, Bloom, Boldizer, and Samuelson, 1985), causing them to have unrealistically positive self-evaluations across a wide range of social contexts. Broadly,

people have been found to perceive themselves as being superior to others across traits such as cooperativeness, decision making, negotiating, rationality, driving skill, health, and intelligence (Babcock and Loewenstein, 1997; Kramer, 1994).

Such inflated views are not based on abstract self-flattery. In fact, people tend to define concrete "performance standards" in ways that systematically favor their own unique set of attributes (Dunning, 1999). For example, Wade-Benzoni et al. (2003) found that people weight the environmental behaviors that they score high on to be more important than other environmental behaviors. In addition, a strong correlation exists between how subjects rate their actions regarding the environment and their judgments of the importance of that action to society. Positive illusions seem to enable people to believe that they are doing well relative to others on important activities, though they may admit to doing less well on activities they consider to be less important. These biases may cause individuals to think that their positive contributions to environmental issues are more important than the contributions of others. For example, an individual who puts effort into recycling, but refuses to take public transportation, may justify this decision by convincing him- or herself that recycling is the most important way of addressing the environmental crisis. Because individuals have the liberty to judge what they already do (which may be what is most convenient for them) as more important than behaviors that may call for inconvenient lifestyle changes, they are able to maintain positive views of themselves with minimal lifestyle adjustment.

By tilting performance assessments in favor of one's own competence, individuals who are paid to make sound decisions are unlikely to doubt their own competence in doing so. In many contexts, in fact, ethics and competence are

intertwined. The auditing executive who believes herself to be honest may also make the claim that her competence allows for the assurance of appropriate behavior. The physician known for astute clinical decision-making and deep commitment to patient well-being is likely to resist the notion that a pharmaceutical-funded trip to Hawaii might influence his clinical decision-making. In a conflict of interest, competence is often viewed as sufficient for avoiding sub-optimal decision-making.

But, Taylor (1989) provides significant evidence that most people view themselves to be more competent than reality can sustain. In some cases, the positive illusion may have benefits, as Taylor and Brown (1988) argue that positive illusions about oneself enhance and protect self-esteem, increase personal contentment, help persistence in difficult tasks, and facilitate coping with uncontrollable events. Taylor (1989) also argues that positive illusions are beneficial to physical and mental health.

However, such positive illusions also put the self at risk in ethical decision-making contexts. The ability to maintain unrealistically positive beliefs about oneself may be constrained to some degree by the objectivity of these beliefs, their credibility, and the potential to disconfirm them (Allison, Messick, and Goethals, 1989). Thus, people can more easily maintain the view that they are more honest than others than to maintain the belief that they are better tennis players or wittier cocktail party conversationalists. We rarely get accurate feedback on our comparative level of honesty. Allison et al. (1989) reason that it is harder to have optimistic illusions when they are inconsistent with easily available, objective data. In the same way, it may be easier for people to maintain the belief that they are fairer than other negotiators than to believe that they are more skillful at reaching profitable agreements.

Thus, while Taylor may be correct about certain advantages that positive illusions provide to the bearer of those illusions, such self-deception can also have less positive consequences. We argue that an additional harm that is created is that these illusions allow the illusion holder to act in his or her own self-interest, and against professional and normative demands. If our vision of self as competent is not always right, and if competence is intended to overcome conflicts of interest, then decision-makers face a serious ethical challenge.

Self as Deserving. In allocating resources, there exists a "tension between self-interest and the equality norm" (Diekmann et al., 1997). Allocators of resources and recipients of resources make sharply different fairness evaluations based on their role. Invariably, collaborators such as co-authors (Taylor, 1989), spouses (Ross and Sicoly, 1979), and joint Nobel prize winners (Harris, 1946) who are asked to quantify their contribution to a joint effort generate a sum greater than 100 percent (Taylor, 1989).

This tendency extends from the self to one's ingroup. In the now-classic "they saw a game" study, Hastorf and Cantril (1954) showed student football fans from Princeton and Dartmouth a film of a football game between the two schools. Both sets of fans watched an identical film, and yet, both sets of fans rated the rival's team as playing less fairly and with less sportsmanship. Assessments of which team was deserving clearly varied by in-group.

This tendency is not limited to football fans. World leaders show the same bias, as in a failed Cold War arms race negotiation where both leaders blamed the rigidity of the other side (Sutton and Kramer, 1990). President Reagan told reporters, "We came to Iceland to advance the cause of peace and although we put on the table the most far-

reaching arms control proposal in history, the General Secretary rejected it." Speaking about the same negotiation, General Secretary Gorbachev stated: "I proposed an urgent meeting here because we had something to propose. . . the Americans came to this meeting empty handed." Kramer (1994) finds evidence in these leaders' memoirs that these perspectives are more than political rhetoric, but reflect the leaders' unconscious commitments to a particular view of self.

Diekmann et al. (1997) examined how the feeling of deservingness affects judgment in a simulation containing many characteristics of real-life conflicts of interest. MBA students were asked to allocate resources across two divisions of a company, and then assess the fairness of the allocation. "Advantaged" allocation recipients assessed these allocations as more appropriate than similar allocations that favored their rivals. In fact, advantaged allocation recipients made such assessments even when the imbalance in their own favor exceeded their own original assessment of an appropriate distribution. In fact, they relied on the fact that another decision-maker had made the allocation to justify the favorable inequality. Finally, egocentrism in assessing fairness was greater when the information about the deservingness of various recipients was vague, leaving room for interpretations favoring the self. This study suggests that decision-makers who rely on their own assessments of who is or is not deserving are at great risk of falling prey to a conflict of interest without realizing it.

Distinguishing Visible and Invisible Conflicts of Interest

So far, we have argued that psychological barriers can prevent decision-makers from recognizing conflicts of interest. First, individuals view themselves as more

powerful than the situation (moral, competent), and then they view any gains incurred as appropriate (competent, deserving). The drive to maintain the view of oneself as moral, competent, and deserving is a barrier to recognizing otherwise visible conflicts of interest.

Visible conflicts of interest are those traditionally thought of by laypeople, economists, and regulators. In this view, the conflict is clearly in view (e.g. the auditor is charged with delivering a fair, potentially negative audit of the client, and simultaneously depends on the client for future earnings) and the decision-maker explicitly vows to remain unbiased by the conflict. Evidence suggests that this vow ignores our basic understanding of how the human mind works, as we overestimate the influence of our own intention and we underestimate the influence of the psychological forces outside of our consciousness. This first type of conflict of interest – the visible, yet dismissed, conflict of interest – is the type referred to in the types of disclosures required by many organizations (e.g. disclosing a financial interest in a client).

A second kind of conflict of interest, less commonly described, is the invisible kind. These more insidious, inadvertent, and self-supporting biases are still considered to be non-obvious and therefore unchecked. The human tendency to favor the self and ingroup creates a gravitational pull towards one set of interests, even when that pull is quite invisible, even to the self. For example, the conflict for an employer is his unconscious tendency to prefer a particular race or gender, yet his fiduciary commitment to shareholders to hire the best talent and his moral commitment to be egalitarian. This invisible conflict of interest is even more pervasive than the visible variety. Here, the conflict of interest is invisible, and therefore, dismissed.

As an example, consider the role of a scholar to be a fair and objective assessor of ideas. In citing work, the scholar's obligation is to cite colleagues who have contributed to the current state of the understanding, rather than to favor oneself or one's group. Tony Greenwald and Eric Schuh (1994) studied the citation tendencies of social scientists, finding that "author's [ethnic] name category [Jewish or non-Jewish] was associated with 41 percent greater odds of citing an author from the same name category." (page 623). This pattern even held up with the data set was limited to prejudice researchers. Presumably, these authors did not set out to exclude work by outgroup authors, but in essence, they did.

The insidious power of the self is evident in data captured on-line using the Implicit Association Test (IAT; Greenwald, McGhee, and Schwarz, 1998). A diverse 2.5 million tests have been taken through a publicly-accessible website (http://implicit.harvard.edu) in which participants are asked to make split-second categorization decisions of words and pictures. The task is presented in two versions, one in which the categories are paired together in an attitudinally "compatible" way (flower and pleasant, insect and unpleasant) as contrasted with the "incompatible" version (flower and unpleasant, insect and pleasant). The difference in the participant's speed in making decisions under the two conditions reflects the individual's implicit bias (in this case, in favor of either flowers or insects). More socially- and self-relevant versions of the test have examined implicit identity, using pairings such as "male and me" and "female and me" (Nosek, Banaji, and Greenwald, 2002). The results of test-takers' implicit identity tests are correlated with their results on other tests, such as implicit attitudes towards math. Implicit identity is shown to correlate highly with individuals'

implicit attitudes towards math, and implicit gender stereotypes about math. That is, test-takers with a strongly masculine implicit identity were more likely to show implicit gender stereotypes associating men (not women) with math, despite the fact that self-reported, conscious attitudes towards gender and math did not reveal such patterns (Nosek, Banaji, and Greenwald, 2002).

A similar pattern was found in a study of implicit racial attitudes and identity. There, two findings are relevant. First, test-takers' group membership (in a race) is related to test-takers' attitudes towards race, particularly for majority group (white) test-takers, most of whom show a bias favoring whites. Second, the test-taker's degree of implicit race identity (black or white) was correlated with the individual's implicit attitudes towards blacks and whites, and implicit attitudes towards self (Greenwald, Banaji, Rudman, Farnham, Nosek, and Mellott, 2002). The centrality of self and group membership is evident, then, especially at the unconscious level, where implicit biases towards oneself are related to other attitudes. Again, this preference for self has important implications for conflicts of interest as decision-makers are prone to invisible conflicts of interest in which their bias for themselves and their own group may distort their ethical decision-making.

The impact of group membership also applies to individuals in a particular professional role, individuals affiliated with a particular side of an issue, or individuals advocating for a particular group. As we cited at the start of our chapter, Simon (1983) noted that "if we become involved in a particular type of activity, we will surely assign it a greater importance and value than we would have prior to our involvement with it" (page 95). Moore et al. (2003) provide evidence that those in the auditing function are at

risk when making related financial assessments. This tendency toward biased information processing prevails even when people on different sides of an issue are exposed to the exact same information (Babcock et al., 1997). While many argue that professional auditors are less subject to these biases, research has found professionals to be vulnerable to the same motivated biases as are other people (Buchman, Tetlock, & Reed, 1996; Cuccia, Hackenbrack, & Nelson, 1995; Moore et al., 2003). When an auditor takes a partisan perspective, he is unlikely objectively assess the data, and is likely to see ambiguous data consistent with the preferences of his client (Babcock et al., 1997; Messick and Sentis, 1979).

The invisible conflict of interest is not only hard to see, but also deceptively easy to dismiss. In many instances, people are socially rewarded for explicit favoring of the ingroup, such as the support of sports teams (Banaji and Greenwald, 1995) or the willingness to do favors for similar others (Banaji, Bazerman, and Chugh, 2003). Human tendency towards such partisanship is strikingly powerful. The tendency to "take sides for no reason", or "implicit partisanship" (Greenwald, Pickrell, and Farnham, 2002) means that humans are always vulnerable to invisible conflicts of interest, even when performing altruistic acts.

Some organizations impose nepotism restrictions (e.g. immediate family members can not work in the same division, or the same company), to prevent conflicting family and organizational interests. While most conflicts of interest commentaries have been role specific, the logic in this paper also applies to situations in which individuals are claiming goods for their own group, selecting people for jobs, admitting students into school, and so on. Conflict of interest is a critical barrier to fairness in society.

Our claim that invisible conflicts of interest pervade every decision that involves our selves both buttresses, and challenges, the distributive justice notions of political philosopher John Rawls (1971). Rawls proposed that if an individual wore a "veil of ignorance" that cloaked his or her identity from himself or herself, the individual would make decisions as if to maximize the welfare of the worst-off member of society. This prediction represents the theoretical reverse of our empirical claim that individuals' decision-making is always influenced by the interests of the self. In this sense, we are making a claim about invisible conflicts of interest that is consistent with the essence of Rawls' view of the importance of imposing a neutral stance. However, Rawls positioned the veil of ignorance as a thought experiment, or theoretical condition, and in fact, the experimental evidence we have presented about the inescapability of the self suggests that the veil is only a theoretical, not actionable, construct. We ourselves have used the veil of ignorance as a powerful pedagogical tool (Banaji, Bazerman, and Chugh, 2003), but are less optimistic about the ability of individuals to truly don the veil. Psychologically, the veil of ignorance is inconsistent with our notions of human bounded ethicality.

Nonetheless, this is not to say that individuals from both advantaged and disadvantaged groups are equally susceptible to these invisible conflicts of interest. System Justification Theory (Jost and Banaji, 1994) demonstrates ways in which members of lower-status group may support, rather than resist, the status quo. In these cases, the tendency to favor one's own group may be less likely. That said, if the tendency of the individual is to be implicitly partisan towards members of other groups, the risk of a conflict of interest still remains, but in an ironically non-self-supportive way.

In our thinking about bounded ethicality, this scenario still represents a conflict of interest (or perhaps, it is better described as a "conflict of non-interest").

Conclusion

We have proposed that perceptual, cognitive, and social cognitive processes are bounded in similar, systematic ways that lead to gaps in observation and errors in decision-making. Despite this robust evidence about boundedness, humans tend to view their own ethicality as unbounded. In fact, decision-makers are psychologically motivated to maintain a stable view of a self that is moral, competent, and deserving, and thus, immune from ethical challenges. Because individuals view their immunity as more powerful than the situation (moral, competent), and view any gains incurred as appropriate (competent, deserving), this view is a barrier to recognizing and addressing conflicts of interest. So, ironically, decision-makers' persistent view of their own ethicality leads to sub-ethical decisions.

While we have limited our application of the bounded ethicality concept to conflicts of interest in this chapter, the concept can be applied to a broad set of ethical decisions. Instances of power and corruption can be explained by the phenomenon as well, as when Bargh and Alvarez (2001) consider the roles of both conscious and nonconscious causes of power abuse. In the related domain of sexual harassment, one researcher has found that three out of four harassers "simply don't understand that they are harassers" (Fitzgerald, 1993, page 22). Bounded ethicality limits the decision-maker's capacity to recognize a wide range of morally problematic issues.

As such, decision-makers are shown to be neither ethical, nor randomly unethical, nor fully aware of their unethicality. In distinctly different ways, three critical 20th

century insights point to the surprising limitations of the conscious mind and the surprising reach of the unconscious mind. In fact, consciousness may play a secondary role in determining judgments and decisions, while much of thought, feeling and motivation may operate in unconscious mode. Such pervasive operation of implicit or unconscious modes of thinking can compromise reaching intended ethical goals.

Our conception of conflicts of interest as instances of bounded ethicality implies that, unfortunately, many of the oft-discussed solutions are inadequate in the face of the robust psychological barriers to recognizing conflicts where they appear. Disclosure of interests addresses only visible conflicts, and even there, the conflict is not removed. Selecting better people is also unlikely to help, as the bias towards a particular view of self is not known to be easily pinpointed. Conventional approaches towards teaching ethics, borne of philosophical traditions, are also unhelpful, constrained by normative views of the ethicality rather than the more descriptive, psychologically based understanding of how the mind works.[2]

While the focus of our chapter has not been prescriptive, we offer that preventive measures represent one important path for redress. The best way to remove the tendency to favor oneself and one's in-group in a decision is to remove oneself from the conflict, whether it be visible or invisible. While such prevention may sometimes be impractical, we offer that the greater, immediate barrier to prevention is the illusion of objectivity that makes prevention seem unnecessary, rather than the practical difficulties of implementing the solution. Before solutions can truly be crafted, the need for a solution must be

[2] The philosophical tradition has begun, in some instances, to integrate the science of the mind. Owen Flanagan, for example, argues for "psychological realism" in ethics, which would constrain moral theories by what is psychologically possible.

recognized. Our argument in this chapter is that this recognition is unlikely to occur, and poses a threat to ethical decision-making in the face of conflicts of interest.

While human bounded ethicality is not an issue of honesty, it has implications for the trustworthiness of our decision-making. Simon (1983), we argue, was right in the quotation that opened this chapter: "Most of the bias that arises … cannot be described correctly as rooted in dishonesty – which perhaps makes it more insidious than if it were" (page 96). Conflicts of interest sometimes pit one's honesty against one's corrupt intentions. However, we have argued that honesty is not the critical bound on ethical decisions, such as those posed by conflicts of interest. Rather, decisions where the self is central are highly prone to self-serving biases that obstruct the recognition of imminent ethical risks. Motivated psychological processes put the decision-making of even "honest creatures" at risk.

References

Allison, S.T., Messick, D.M., & Goethals, G.R. (1989). On being better but not smarter than others: The Mohammad Ali effect. *Social Cognition, 7,* 275-296.

Armor, D.A. (1998). The illusion of objectivity: A bias in the perception of freedom from bias. [Dissertation Abstract]

Babcock, L., Loewenstein, G., Issacharoff, S. (1997). Creating convergence: Debiasing biased litigants. *Law and Social Inquiry-Journal of the American Bar Foundation 22 (4)*: 913-925.

Babcock, L., & Loewenstein, G. (1997). Explaining bargaining impasse: The role of self-serving biases. *Journal of Economic Perspectives 11*, 109–126.

Banaji, M.R., and Greenwald, A.G. (1995) Implicit gender stereotyping in judgments of fame. *Journal of Personality and Social Psychology. 68 (2),* 181-198.

Banaji, M.R. and Bhaskar, R. (2000). Implicit Stereotypes and Memory: The Bounded Rationality of Social Beliefs. In Schacter and Scarry (Eds.), *Memory, Brain, and Belief (pp. 139-175).* Cambridge, Massachusetts: Harvard University Press.

Banaji, M.R., Bazerman, M.H., and Chugh, D. (2003) How (Un) Ethical Are You? *Harvard Business Review, 81 (12),* 56-64.

Bargh, J. A. (1997). The automaticity of everyday life. In R. S. Wyer, Jr. (Ed.), The automaticity of everyday life: Advances in social cognition (Vol. 10, pp. 1-61). Mahwah, NJ: Erlbaum.

Bargh, J.A. and Alvarezin, J. (2001). The Road to Hell: Good Intentions in the Face of Nonconscious Tendencies to Misuse Power. In Bargh, J.A. and Lee-Chai, A.Y.

(Eds.), *The Use and Abuse of Power: Multiple Perspectives on the Causes of Corruption (pp. 41-55).* New York: Psychology Press.

Bargh, J.A. & Wegner, D. (1999). Control and Automaticity in Social Life. In Gilbert, D.T., Fiske, S.T., and Lindzey, G. (Eds), *Handbook of Social Psychology (pp. 445-496).* New York: Oxford University Press.

Baumhart, R. (1968). *An honest profit: What businessmen say about ethics in business.* New York: Holt, Rinehart and Winston .

Bazerman, M. H. (2002). *Judgment in Managerial Decision Making* (5th ed.) New York: John Wiley & Sons.

Buchman, T.A., Tetlock, P.E., & Reed, R.O. (1996). Accountability and auditors' judgment about contingent events. *Journal of Business Finance and Accounting, 23,* 379-398.

Chugh, D. (in press). Why Milliseconds Matter: Societal and Managerial Implications of Implicit Social Cognition. *Social Justice Research.*

Cuccia, A.D., Hackenbrack, K., Nelson, M.W. (1995). The ability of professional standards to mitigate aggressive reporting. *Accounting Review. 70 (2),* 227-248.

Diekmann, K.A., Samuels, S.M., Ross, L., & Bazerman, M.H. (1997). Self-Interest and Fairness in Problems of Resource Allocation. *Journal of Personality and Social Psychology, 72,* 1061-1074.

Dunning, D., Meyerowitz, J.A. and Holzberg, A.D. (2002) Ambiguity and Self-Evaluation: The Role of Idiosyncratic Trait Definitions in Self-Serving Assessments of Ability. In Heuristics and Biases, Eds. Thomas Gilovich, Dale Griffin, and Daniel Kahneman. Cambridge: Cambridge University Press.

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

Epley, N. and Caruso, E.M. (in press). Egocentric Ethics. *Social Justice Research.*

Fazio, R.H., Jackson, J.R., Dunton, B.C. & Williams, C.J. (1995). Variability in automatic activation as an unobtrusive measure of racial attitudes: A bona fide pipeline? *Journal of Personality and Social Psychology, 69,* 1013-1027.

Fitzgerald, L.F. (1993). Violence against women in the workplace. *American Psychologist, 48 (10),* 1070-1076.

Flanagan, O. (1993). *Varieties of Moral Personality: Ethics and Psychological Realism.* Cambridge: Harvard University Press.

Gilbert, D.T. & Hixon, J.G. (1991). The trouble of thinking: Activation and application of stereotypic beliefs. *Journal of Personality and Social Psychology, 60,* 509-517.

Greenwald A.G., McGhee D.E., Schwartz, J.L.K. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology. 74 (6),* 1464-1480.

Greenwald, A. G. (1980). The totalitarian ego: Fabrication and revision of personal history. *American Psychologist, 35,* 603-618.

Greenwald, A. G., & Banaji, M. R. (1995). Implicit social cognition: Attitudes, self-esteem, and stereotypes. *Psychological Review. 102,* 4-27.

Greenwald, A.G., Pickrell, J.E., Farnham, S.D. (2002). Implicit partisanship: Taking sides for no reason. *Journal of Personality and Social Psychology. 83 (2),* 367-379.

Greenwald, A.G., Schuh, E.S. (1994). An Ethnic Bias is Scientific Citations. *European Journal of Social Psychology, 24 (6),* 623-639.

Greenwald, A.G., Banaji, M.R., Rudman, L.A., Farnham, S.D., Nosek, B.A., and Mellott, D.S. (2002). A Unified Theory of Implicit Attitudes, Stereotypes, Self-Esteem, and Self-Concept. *Psychological Review, 109 (1),* 3-25.

Griffin D.W., Dunning D., Ross L. (1990). The role of construal processes in overconfident predictions about the self and others. *Journal of Personality and Social Psychology, 59,*1128-39.

Haidt J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review*. 108 (4), 814-834.

Harris, S. (1946). *Banting's miracle: The story of the discovery of insulin*. Toronto: J.M. Dent and Sons.

Hastorf, A.H., & Cantril, H. (1954). They saw a game: A case study. *Journal of Abnormal and Social Psychology, 49,* 129-134.

Jost, J.T., & Banaji, M. R. (1994). The role of stereotyping in system-justification and the production of false consciousness. *British Journal of Social Psychology, 33,* 1-27.

Kahneman, D. (2002). *Acceptance Speech for the Nobel Prize.* Retrieved June 1, 2003 from Nobel e-Museum Web site:http://www.nobel.se/economics/laureates/2002/.

Kahneman, D., & Tverksy, A. (1973). On the psychology of prediction. *Psychological Review, 80,* 237-251.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica, 47,* 263-291.

Kramer, R. M. (1994). *Self-enhancing cognitions and organizational conflict.* Unpublished manuscript.

Kronzon S, and Darley J. (1999). Is this tactic ethical? Biased judgments of ethics in negotiation. *Basic and Applied Social Psychology 21 (1),* 49-60.

Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin, 108,* 408-420.

Lee-Chai, A. and Bargh, J.A. (2001) Eds. *The Use and Abuse of Power.* Ann Arbor, MI: Sheldon Press.

Macrae, C.N., Bodenhausen, G.V., Milne, A.B., & Jetten, J. (1994). Out of mind but back in sight: Stereotypes on the rebound. *Journal of Personality and Social Psychology, 67,* 808-817.

March, J.G., & Simon, H.A. (1958). *Organizations*. New York: John Wiley & Sons.

Messick, D. M., & Sentis, K. (1983). Fairness, preference, and fairness biases. In D. M. Messick and K. S. Cook (Eds.), *Equity theory: Psychological and sociological perspectives* (pp. 61-64). New York: Praeger.

Messick, D.M. & Sentis, K.P. (1979). Fairness and preference. *Journal of Experimental Social Psychology, 15,* 418-434.

Messick, D. M., Bloom, S., Boldizer, J. P., and Samuelson, C. D. (1985). Why we are fairer than others. *Journal of Experimental Social Psychology, 21,* 480-500.

Messick, D.M., & Bazerman, M.H. (1996). Ethics for the 21st Century: A Decision Making Approach. *Sloan Management Review,* 37, 9-22.

Moore, D.A., Loewenstein, G., Tanlu, L., and Bazerman, M.H. (2003). *Auditor Independence, Conflict of Interest, and the Unconscious Intrusion of Bias.* Harvard Business School Working Paper #03-116.

Nosek, B.A., Banaji, M.R., & Greenwald, A.G. (2002). Harvesting intergroup attitudes and stereotypes from a demonstration website. *Group Dynamics, 6,* 1, 101-115.

Orwell, G. (1949). *1984.* New York: Harcourt, Brace.

Rawls, J. (1971) *A Theory of Justice*. Cambridge, Mass: Harvard University Press.

Ross, M., & Sicoly, F. (1979). Egocentric biases in availability and attribution. *Journal of Personality and Social Psychology, 37,* 322-337.

Shapiro D.L. (1991). The effects of explanation on negative reactions to deceit. *Administrative Science Quarterly. 36,* 614-30.

Simon, H. A. (1957). *Models of Man*. New York: Wiley.

Simon, H.A. (1983). *Reason in Human Affairs.* Stanford: Stanford University Press.

Sutton, R., & Kramer, R. M. (1990). Transforming failure into success. Impression management, the Reagan administration, and the Iceland arms control talks. In R. L. Zahn, & M. N. Zald (Eds.), *Organizations and nation-states: New perspectives on conflict and co-operation.* San Francisco: Jossey-Bass.

Taylor, S. E. (1989). *Positive Illusions*. New York: Basic Books.

Taylor, S. E., & Brown, J. D. (1988). Illusion and well-being: a social psychological perspective on mental health. *Psychological Bulletin, 103,* 193-210.

Tenbrunsel, A.E. (1998). Misrepresentation and expectations of misrepresentation in an ethical dilemma: The role of incentives and temptation. *Academy of Management Journal, 41,* 330-9.

Tenbrunsel, A.E. Justifying unethical behavior: The role of expectations of others' behavior and uncertainty (Dissertation).

Thaler R.H. (1996) Doing Economics without Homo Economicus. In Richard H. Thaler (Ed.), *How do Economists do Economics.* Warren Samuels.

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

Wade-Benzoni, K.A., Thompson, L.L. and Bazerman, M.H. *The Malleability of Environmentalism.* Unpublished Manuscript.

Wegner, D. (2002). *The Illusion of Conscious Will.* Cambridge: MIT Press.