Social Cognition, Vol. 27, No. 5, 2009, pp. 786–812

A RATIONAL RECONSTRUCTION OF MISBEHAVIOR

Joachim I. Krueger
Brown University

Adam L. Massey
University of California, Los Angeles

Classic social psychological research suggests that humans are more susceptible to social influence than they should be. This discrepancy between observed and normative behavior is often taken as the footprint of irrational reasoning. As much of the studied behavior is also socially undesirable, there is a dual verdict of irrationality and immorality. Using concepts and analytical tools from decision and game theories, we propose a rational reconstruction of key instances of misbehavior. Conforming with one's peers, obeying legitimate authority, and being less helpful when in a group of bystanders can be modeled as the outcomes of intelligent information processing and strategic behavioral choice. Yet, a reconstruction of rationality does not entail a reconstruction of morality.

"You've got to change your evil ways, baby, before I stop loving you."
—Carlos Santana

Social psychological research has produced a wealth of data showing the influence of real or imagined others on people's thinking, feeling, and behaving. Some of the most striking studies showed that, given the right circumstances, many people will yield to social influence and act in ways that they, if left to their own devices, would not even consider. Such discrepancies raise questions about people's rationality. At the same time, the undesirability of these behaviors also raises questions of morality. Among the best-known examples of experimentally exerted social influence are Asch's (1956) studies on conformity, Milgram's (1963) work on obedience, and Darley and Latané's (1968) experiments on the bystander effect. These studies have obtained almost mythic status as they are thought to reveal the

I am indebted to Theresa DiDonato and David Funder for their thoughtful comments on a draft version of this manuscript.

Correspondence concerning this article should be addressed to Joachim I. Krueger, Department of Psychology, Brown University, Box 1853, 89 Waterman St., Providence, RI 02912. E-mail: Joachim_Krueger@Brown.edu.

tragic flaws of ordinary people. Their moral failures seem to emerge from deep-seated constraints on their ability to think rationally. It is somewhat ironic that these classic studies have not only afforded inferences about human limitations, but have also served as important building blocks of the situationist paradigm, which grants individual differences and intra-individual processes little explanatory power (Bargh, 2007; Zimbardo, 2007).

The more seriously one takes situationism, the less room one has for inferences about individual people (Kelley, 1972). Yet, social psychological variants of situationism have never been as strident as Skinner's (1955-1956) orthodoxy. The project of constructing situationist accounts of social behavior while still holding human actors responsible is open to the charge of trying to have it both ways. A more benign interpretation is that social psychology has kept the search from personal causation, agency, and responsibility alive.

Assuming that individual people have something to do with the behavior they emit, questions about morality and rationality are pertinent. The two questions are often tied up together. In a folk psychological adaptation of Plato, people would not behave despicably if they had their wits about them. Likewise, they would not act silly if they were thinking clearly. Allen Funt's "Candid Camera" explored the conjunction of immoral (or silly) and irrational behavior to great effect, and social psychologists took notice (Milgram & Sabini, 1979).

The classic studies on conformity, obedience, and apathy vitalized social psychology, while dealing a shock-and-awe blow to the conventional wisdom that ordinary people will act independently, that they will resist evil—though ultimately toothless—authority, and that they will help those in an evident and dire emergency. The "tu quoque" message of these studies left people wondering if they would muster the strength to do the right thing in a difficult situation. Yet, because many participants did not yield to the experimental demands, strict situationism cannot be right (Krueger, 2009). If some managed to resist, what does their behavior tell us about those who complied?

In this article, we revisit the famous studies by Asch, Milgram, and Darley and Latané to evaluate the rationality of participants' behavior. To do this, we take the perspective of decision-theory and game theory. Neither theory views people as mindless automata who reflexively respond to situational stimuli. Instead, the working hypothesis is that humans are potentially rational creatures who gather, interpret, and combine information to reach decisions and make choices. By evaluating the rationality of social behavior in light of explicit criteria, the temptation to infer irrationality from immorality is disabled. This decoupling recognizes the conceptual and empirical separability of these two dimensions (Krueger & Acevedo, 2007). Rational judgment can bring about horrible behavior, and conversely, benevolence and compassion may at times overwhelm rationality (Dawes, 1988).

CONFORMITY

How could individual socialists shout "Sieg Heil!" when their convictions revolted against this act? How could professed liberals hail the rush toward Baghdad when they opposed the war in the first place? Asch's (1956) experiments showed that even a unanimous majority of three can create a crowd mentality. About one third of the time, respondents called a short line long or a long line short when oth-

ers did the same. Asch created a psychological crisis by pitting the power of social consensus against the power of visual perception. The standard interpretation of his research is that people violate a norm that says visual perception trump social consensus. Campbell (1990, p. 45) asserted that "independence is productive from the social point of view, since it is the only way to correct errors and to steer the social process in accordance with felt requirements, [whereas] yielding is antisocial because it spreads error and confusion." The mandate of independence is justified on both moral and rational grounds. It is moral because it reflects the difficult, yet principled, high road; it is rational because it introduces the person's own perception as valid evidence into the process.

Asch's results are regarded as particularly striking because of his experimental task's objective simplicity. If conformity increases with stimulus ambiguity (Deutsch & Gerard, 1955), should it not be zero if there is no ambiguity? Yet, the experiment was presented as a study on visual perception. If individual certainty about the correct answer was the only concern, the study could have been designed as an experiment on analytical reasoning with questions such as What is the square root of 25? In either case, the judgment of lines or simple mathematical puzzles, any ordinary participant would be perplexed by peers confidently announcing wrong answers. But there is a difference. Whereas there are no experiments asking college students trivial math questions, there are experiments on visual perception. Asch's stage set-up simulated a research environment in which individuals might legitimately respond differently. However convinced participants were that their own perceptions were accurate, the billing of the study as an experiment on visual perception allowed the possibility of optical illusions. Even a small subjective probability that one's own perception may have been tricked is enough to produce a crack in the conviction that the judgments of others must be ignored.

Asch's (1956) findings of significant conformity under conditions that also encouraged individual autonomy revived the specter of crowd psychology (Rook, 2006) and especially notions of herd behavior (Tarde, 1895; Trotter, 1916). Later studies continued Asch's search for moderator variables of conformity (Bond & Smith, 1996), while accepting general the frame of individual irrationality (Kameda & Tindale, 2006).

In contrast to the prevailing view, the rational-actor perspective models conformity on the premise that behavior is the endpoint of a principled reasoning process. Conformity, and herding behavior in general, can be represented as informational cascades (Bikhchandani, Hirschleifer, & Welch, 1992). Cascade models assume that each person, except the first, has private information regarding the stimulus and information regarding the decisions of others who responded earlier. The first person has only private information, and is therefore most critical for the direction the cascade will take. The second person knows that the first person's decision is a reflection of private information. If the second person's private information is the same as the first person's (e.g., L for long), there is an LL sequence. If it is different (e.g., S for short), the model assumes that the second person flips a mental coin. In other words, people are assumed to rationally weight the judgments of others as much as their own. Knowing—or assuming—this, the third person can infer that the second person has the same belief as the first one with a probability of .75 if an LL sequence has occurred.

If the first two people agree, the third person falls in line without having to place any value on conformity per se. For all subsequent people, it is sufficient to think that the decisions of the first two individuals "reflect information that they have and we do not" (Banerjee, 1992, p. 798). An informational cascade sets in because "it is optimal for an individual, having observed the actions of those ahead of him, to follow the behavior of the preceding individual without regard to his own information [and] once the decision-maker disregards own information, his behavior is uninformative to others" (Bikhchandani et al., 1992, p. 994).

Consider a few implications of this model. If L is the correct judgment, LLL is a positive cascade, and SSS is a negative one. The former is more likely than the latter if private information has any validity, however modest, p[correct] > .5. The probability of a correct cascade after two individuals is

$$\frac{p(p+1)}{2},$$

the probability of no cascade is

$$p(1-p),$$

and the probability of a negative cascade is

$$\frac{(p-2)(p-1)}{2}$$

(Bikhchandani et al., 1992, p. 998). If a cascade has occurred, the probability that it is a correct one is

$$q = \frac{p(p+1)}{p(p+1)+(p-2)(p-1)}$$

The probability that a cascade is a positive one is larger than the probability that the private information, which was discarded to let the cascade continue, was correct. The difference $q - p$ is greatest for a value of p halfway between .5 and 1 (see Appendix for proofs).

As these mathematical implications show, an individual's decision to conform "reflects an element of wisdom" (Hung & Plott, 2001, p. 1519). Whether the same can be said from the collective's point of view is still a matter of debate (see Hung & Plott, 2001, for pro and Banerjee, 1992, for con). If private information is valid ($p > .5$), positive cascades are more probable than negative cascades. However, the number of people whose valid private information is subverted in a negative cascade is greater than the number of people whose invalid private information is subverted in a positive cascade. These two effects cancel each other out.

Although positive cascades facilitate collective accuracy and well-being, it is important to note that not all accurate (i.e., efficient) group decisions serve the best interest of all individuals. The beneficial effects of some positive cascades are self-limiting. In certain animal species, like grouse or guppies, some females identify the fittest males and seek to mate with them. Other females then follow suit without making their own assessments (Dugatkin, 2000). There is some evidence that humans do the same (Graziano, Jensen-Campbell, Shebilske, & Lundgren, 1993). Such a cascade creates a coordination problem. When all females focus their attention on a few males, they start crowding one another (Becker, 1991). Some

ftemales may not get to mate at all despite the availability of other males of acceptable fitness.

Three empirical patterns suggest that cascade models are applicable to the Asch paradigm. First, the rise and leveling-off of empirical conformity rates resemble S-shaped growth curves (Tanford & Penrod, 1984), which in turn resemble the cascading pattern. Second, and as noted above, conformity increases with stimulus ambiguity (Bond & Smith, 1996). Third, conformity does not decrease when participants' own decisions are made in private, as, for example, in the Crutchfield (1955) apparatus (Bond & Smith, 1996). The last two patterns miliate against the traditional distinction between informational and normative influence. Cascade models only recognize the former.

The scientific study of conformity contains an interesting self-referential irony as evidenced by the existence of scientific fads (Kuhn, 1962; Rozin, 2007). Campbell (1990) suggested that conformity researchers themselves are not immune to the very topic of their own studies.[1] One hopes that scientific activity is not variable because it operates like a beauty contest, but that it is self-correcting in the long run. In beauty contests, judgments do not have a static probability of being correct (Keynes, 1936). Instead, the probability of being correct is conditional on the predictions themselves. Investment markets that depend on this kind of second-guessing are notoriously volatile (Ottaviani, & Sørensen, 2000). If instead, the philosophy of "fallible ontological realism" (Campbell, 1990, p. 49) is applicable to social psychology, scientific judgments are constrained by truths lying outside of the judgments themselves.

Positive and negative cascades are polarized states, and it is difficult to predict in advance, which one will obtain and when a switch from one state to another will occur (Salganik, Dodds, & Watts, 2006). As easily as cascades form, so they can collapse. As "each person moves knowing the choices made by those before her but not the information these choices are based on" (Banerjee, 1992, p. 799), cascades can only become longer but not stronger. The problem is that cascades "prevent the aggregation of information of numerous individuals" (Bikhchandani et al., 1992, p. 998). A single dissenter, who is in possession of a particularly strong piece of private information, who is egocentric, or who responds randomly, can unravel unanimity. Ironically, it is irrational individuals that are most likely to stop or reverse a cascade (Huck & Oechssler, 2000). Such disruptions need not be a threat to the collective because positive cascades tend to recover more easily than negative cascades. Hence, it makes sense for the collective to encourage independence among individuals (Sunstein, 2003).

The Ally Effect is consistent with the fragility of cascades. A single dissenter placed prior to the naïve participant boosts the autonomy of the latter (Asch, 1956).[2] The standard interpretation of this effect is motivational. Participants are supposed to feel liberated from peer pressure. Cascade models suggest a more cognitive view. The participant now looks back on a sequence of broken unanim-

1. "Most of the hundreds of existing conformity studies have been done by researchers who are themselves very conformant to current fads in their discipline [and many of them, unlike Asch himself] implicitly created a deprecating social distance between themselves and those fellow human beings whom they have duped into 'conforming'" (Campbell, 1990, p. 41).

2. It is ironic that the ally effect should be seen both as an example of social influence and as an example of increased personal independence from social influence.

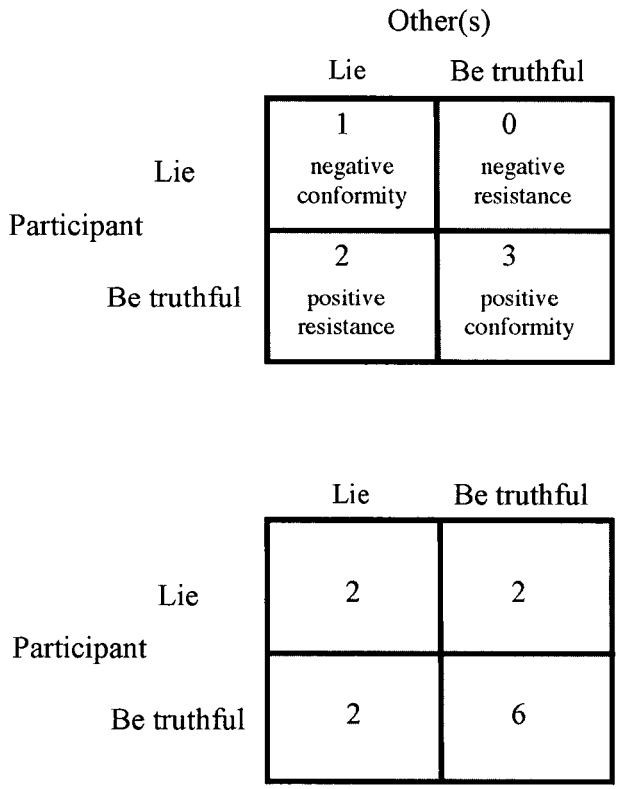

FIGURE 1. Preference ranking for a self-regarding person in the Asch situation (top panel) and a self- and other-regarding person (bottom panel).

ity. At most, there are four respondents (the first two in the entire sequence and the first two after the dissenter) whose judgments the participant regards as false, and there are two, the prior dissenter and the participant him- or herself, whose judgments the participant regards as correct. Suppose, for example, a participant looks back on a sequence of HHHHLHHH[L], where the final L is his or her private information. It might seem that this person would conform just like the third and fourth person did. However, this person may conclude that the prior dissenter's private information was unusually strong, and therefore put more weight on it than on the first person's decision.

The finding that prior dissenters facilitate the participants' decision to resist has an interesting implication. Neither the experimental work nor the formal models permit the possibility that people are mindful of how their judgments affect others. If people understood the fragility of cascades, and if they knew that others were looking to them for support, they could be more willing to resist. In the Asch paradigm, participants did not have this inducement for independence because they were seated at or near the end of the sequence.

Rational-actor models assume that people conform when they believe that conformity maximizes the expected value of the decision. Consider the Asch situation from a game-theoretic perspective (Luce & Raiffa, 1957). The top panel of Figure 1 shows the payoff matrix, where payoffs are presented as ordinal preferences. Overall, the participant prefers to speak the truth, but the strength of this preference is modulated by what others do. The most desirable outcome is a situation of positive conformity where everyone is being truthful (3 utils). The least desirable outcome is a situation of negative resistance where only the participant lies (0 utils). When others lie, the participant also prefers truth-telling, but the payoff discrepancy resulting from a change in behavior is less extreme. Own truth-telling in the face of others' lying is a case of positive resistance (2 utils), and own lying in the presence of others' lying is negative conformity (1 util).

According to classic game theory, a rational person will tell the truth if truth-telling is the dominating choice. Even when the critical participant—as was the case in the Asch situation—always acts after others do, the Theory of Moves (TOM; Brams, 1994) suggests the participant will select the best response. This rudimentary view of rationality cannot explain why some participants accept negative conformity some of the time. One incentive for negative conformity is fear of punishment. Small social groups tend to lack tolerance intolerant of deviants (Bernheim, 1994; Milgram, 1961; Schachter, 1951). Sanctions work because people are sensitive to social rejection, reprimand, and exclusion (Williams, 2007). Most people anticipate that the group's majority will try to bring them to heel for breaking rank (Monin, Sawyer, & Marquez, 2008), and they may therefore perceive conformity as a defensible strategy (Gigerenzer, 2008). Campbell (1990) saw the intolerance of groups as an evolutionary adaptation, arguing that "consensus clears the way for social action" (p. 46; see also Boyd & Richerson, 2005). The problem with this explanation is that Asch ruled it out in his design. Over 12 trials, his participants had ample opportunity to learn that the other group members—in contrast to naïve participants in an otherwise identical study design (Asch, 1952)—responded rather stoically to individual deviants.

A broader definition of self-interest brings a more potent incentive into play. Many people have social preferences that combine regard for the self with respect or concern for others. The preferences of prosocial individuals can be modeled as composites of selfishness and benevolence (van Lange, 1999). Consider a simple model, in which a person's preferences are the sums of own payoffs and other persons' payoffs weighted by 1/N, where N is the number of other people in the group. The bottom panel of Figure 1 shows the transformed preference matrix. Truth-telling is no longer the dominating choice. In a sequential game, where the behavior of others is already known to depart from the participant's truthful behavior, the choice is between positive resistance and negative conformity.

When the utilities do not yield a clear guide for action, additional considerations must be introduced. Although some people may respond randomly when the differences between their utilities are too close to call, researchers do not seriously entertain this possibility. Asch (1956) thought that people conform publicly and demur privately; Campbell (1990) thought they should do the opposite. Both realized that participants were trying to solve a complex epistemological problem that pitted truth against consensus.

Hodges and Geyer (2006) proposed an ingenious solution. Focusing on the multiple-trials property of the study's design, they noted that participants could

have used mixed decision strategies. They could conform with a certain probability and be independent with the complement of that probability. By this account, participants not only cared about the preferences of others, but they used their own intermittent acts of conformity to signal their respect for the group. "Agreeing with the incorrect majority occasionally might not be an error (as Asch called it), but a creative strategy to communicate unity" (Hodges & Geyer, 2006, p. 6). This is an appealing idea, although it runs into the same problem as the fear-of-sanctions hypothesis. It must have been odd for participants to notice that they were the only ones being strategic. There was no indication in Asch's (1956) work that the probability of conformity decreased over trials. Another possibility is that intermittent dissent was not a strategy to communicate respect for others and allegiance to the group, but a strategy to signal one's own strength and uniqueness. Men in particular, and especially those motivated to attract a mate, choose to oppose majority opinions (Griskevicius, Goldstein, Mortensen, Cialdini, & Kenrick, 2006). In sum, these analyses suggest that social behavior is at its most rational and adaptive when it strikes a balance. Asch's participants seem to have come close to this goal.

OBEDIENCE

Many destructive human acts are committed in the name of obedience. One of the enduring contributions of social psychology is evidence for the power of authority to make people perform acts of cruelty. Social psychology has rallied to the call of Question authority! while struggling to separate legitimate from illegitimate authority. On the one hand, humans gain advantages from respecting powers that are legitimate or too strong to be opposed. Deference to expert judgment, for example, is a powerful heuristic for managing one's attitudes (Kruglanski et al., 2005). On the other hand, obedience has the potential of multiplying violence and harm far beyond what it would be if the authorities had to do themselves what they ask of others.

Milgram had worked with Asch, and he had conducted conformity experiments in Norway and France (Milgram, 1961). At Yale University, he turned his attention to the study of destructive obedience (Milgram, 1963). An insistent experimenter, who exerted explicit social influence, replaced the majority of stoic peers. The delivery of electric shocks to a presumed victim replaced the mundane psychophysical judgments. With the simultaneous reduction of the number of influence agents and the increase in the severity of the behavioral consequences, Milgram's studies became an experimental parable for 20th-century violence. Whereas a participant could walk away from the Asch situation with a "no-harm-no-foul" attitude, a participant in the Milgram study had to wonder if a homicide had occurred.

The standard interpretation of Milgram's (1963) work is that it validated the Eichmann defense. Eichmann claimed in Jerusalem that his role in the Holocaust had to be understood in hierarchical and bureaucratic terms (Arendt, 1963). He followed orders in a context that did not tolerate dissent. The Eichmann defense is a transparent maneuver, which was tried before, and without success, at Nuremberg. In the eyes of the prosecution and the public, such gambits are deceitful and contemptible. Against this background, Milgram's finding that about two thirds of ordinary people are capable of homicide was disturbing. If everyone is at risk of

surrendering to authoritarian pressure, a personal sense of invulnerability seems sanctimonious and foolish.

Why does obedient behavior appear to be irrational? The critical feature of Milgram's design is the gradual growth of an accommodating behavioral pattern coupled with a gradual increase of the behavior's consequences. The teacher-participant begins to administer small shocks to the learner-confederate when the latter makes mistakes in a memory tasks. With each mistake, the shock increases by 15 volts. The crisis of conscience emerges slowly, coming to a head at about 150 volts when the learner-confederate demands to be released. It is at this point that disobedience, if it occurs at all, is most likely to be seen (Packer, 2008).

The slippery-slope pattern of incremental obedience recalls the foot-in-the-door effect in compliance (Freedman & Fraser, 1966) and the sunk-cost effect in decision-making (Arkes & Blumer, 1984). These effects are social traps, in which escalating commitments seduce people into doing what they don't want to do, in other words, into acting irrationally. In social traps, behavioral consistency is not rational, but the sign of a deeper incoherence. The concept of *akrasia*, introduced by Plato in his Protagoras and elaborated by Aristotle in his Ethics, survives in modern theories as a Weakness of Will (Ainslee, 2001). Addictions and other failures to delay gratification fall under this type of après-moi-le-deluge reasoning. By this account, Milgram's participants were being irrational; they chose the cold comfort of obedience over the hard work of resistance.

There is an alternative perspective, according to which obedient behavior is not necessarily irrational. This analysis also turns on the sequential nature of behavior. In this view, repeated acts of obedience amount to a within-person informational cascade or Self-Herding (Ariely, 2008). With each passing trial, the participant integrates a private signal with the growing shadow of his past behavior. As in a typical group cascade, the incentive to conform increases with the length of the sequence, and as in a group cascade, the increments in these incentives become smaller. In contrast to the group cascade, the private signal grows stronger. If and when the two lines cross, the individual should dissent (and some do). Like group cascades, individual cascades are fragile. Like Asch, Milgram showed that when dissent is introduced, a cascade quickly collapses. Participants disobeyed more readily if they found themselves team-teaching with others who stood up to the experimenter (Milgram, 1974; but see Burger, 2009, for a failure to replicate this effect).

If the cascade analogy is insufficient for the reconstruction of rationality, a second aspect of sequentiality can be noted. Each act of obedience is preceded by an explicit instruction. Each command-and-response unit can be modeled as a social game. After each of the learner's failures to perform the memory task, the experimental protocol demands punishment; if the teacher-participant hesitates, the experimenter urges the participant to go on. "You have no other choice. You must go on" (Milgram, 1974, p. 29).

It is then the participant who makes the first move. He obeys or resists. He knows that the experimenter has the second move, which he can imagine as either relenting or punishing. This yields four combinations of outcomes. (A) The participant obeys and the experimenter relents (i.e., lets the participants off the hook, ultimately). (B) The participant obeys and the experimenter punishes (i.e., calls security or produces other institutional mechanisms to intimidate the participant).

(C) The participant resists and the experimenter relents (i.e., accepts his decision).
(D) The participant resists and the experimenter punishes.

To assess whether a strategy is rational, one needs to make assumptions about both players' preferences and accept them as common knowledge (Brams, 1994). Then, the player who moves first can anticipate how the other player will respond given that player's preferences. The argument that it was rational for Milgram's participants to obey implies that participants expected the experimenter to relent after obedience (which turned out to be true), and that this outcome would be preferable to any outcome resulting from resistance.

To illustrate, consider an ancient precedent. In the Hebrew Bible, God commands Abraham to take his son Isaac to Mount Moriah, and sacrifice him (Genesis, chapter 22). Abraham obliges, but before he can complete the act, God relents and produces a ram to be slaughtered in the boy's stead. Whereas the traditional reading treats the story as a test of faith, Brams (2003) sees it as a test of wits. He explores several plausible preference rankings, but let it be sufficient to consider those that leave the greatest room for Abraham's uncertainty. These scenarios are the most conservative tests of the rationality hypothesis, and the ones that most plausibly transfer to the Milgram situation (see Brams, 2003, pp. 37-45, for his full exposition).

Treating God as player, who is omnipresent and emotional, but not omnipotent, Brams attributes preferences to Him. God prefers an obedient servant (O) over a disobedient one (D), and if the servant obeys, He would rather build His reputation as a merciful (M) rather than the punitive God (P). If the servant is spiteful, however, He prefers to be feared to being seen as one who can be duped. In other words, God's preference ranking is (OM) > (OP) > (DP) > (DM).

Abraham's preferences are different. According to the first scenario, Abraham prefers a merciful god (M) over a punitive one (P). If there is mercy, he prefers obedience (O) over disobedience (D), but if there is punishment, he might as well disobey. His preference ranking is therefore (OM) > (DM) > (DP) > (OP). Figure 2 (top) shows these rankings, running from 3 down to 0, in the normal-form representation of the game. Brams (2003) argued that Abraham knows enough about God to infer His preferences (e.g., by remembering that God promised him many descendants). Even without such knowledge, Abraham can simulate God's preferences by projecting the preferences he thinks he would have if he were God (Krueger, 2007). Then, as Abraham moves first, all he needs to realize is that God's dominating strategy is tit-for-tat. If Abraham obeys, God will be merciful; if he disobeys, God will be wrathful. Hence, Abraham obeys.

The second scenario, shown in Figure 2 (bottom) involves a slight change in Abraham's preferences, making it harder for him to obey. Here, Abraham still prefers a merciful over a punitive god, but he values his son's life more. In this scenario, Abraham prefers the outcome of DM over the outcome of OM. That is, he places no value on obedience per se. Still, anticipating God's preference, he obeys and realizes his second ranked outcome and allows God to realize His first preference. The game again ends to everyone's satisfaction because the first player knows that the second player will use a tit-for-tat strategy.

Consider another game-theoretic reconstruction, in which obedience is cast as a cooperative offer in a trust game (Rosenthal, 1981). In a trust game, player 1 has an endowment that, if she transfers it to player 2, multiplies in value. Player 2 can then keep it all or return half. Classic game theory stipulates that betrayal of trust

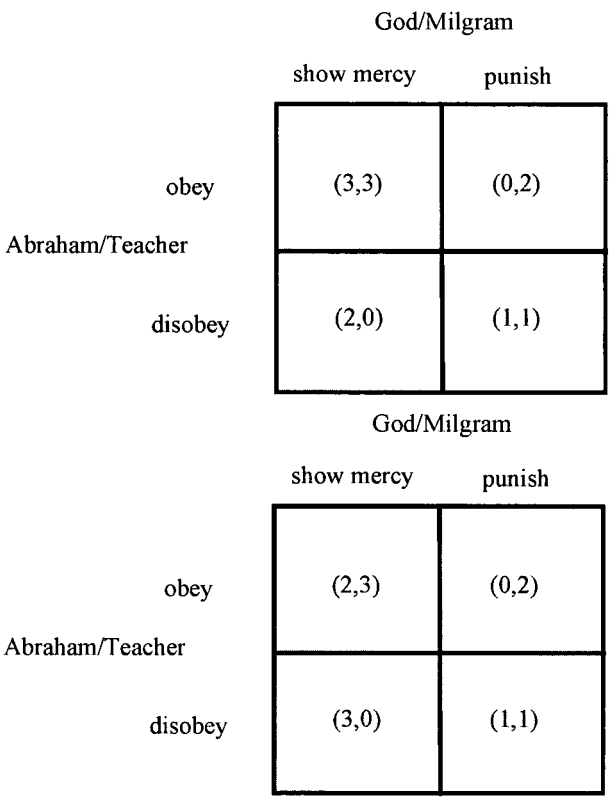

FIGURE 2. Preference rankings in a sequential game between Abraham ("The Teacher") and God (Milgram).

is the rational, Nash-equilibrating strategy, but most human players reciprocate trust. When the game is played repeatedly, trust and trustworthiness yield more efficient outcomes than distrust and betrayal. Note that in the Theory of Moves analysis, player 2 (God/experimenter) gets his most prized preference, whereas he gets only his second best in the trust game analysis. This is so because the former already incorporates psychological assumptions about what players care about. Recall that God was construed as *wanting* to be seen as merciful, as caring more about His image than about shekels. Human players honor trust for the same reason. They are most likely to adhere to norms of reciprocity when their reputations are at stake (Krueger, Massey, & DiDonato, 2008).

Cialdini (2001, p. 185) recognized that the ancient tale of human sacrifice "might be the closest biblical representation of the Milgram experiment." But the analogy has limits. Consider the following differences. First, the Milgram situation is a repeated game, whereas Abraham and God played a one-shot game. Because Milgram's experimenter is not God or seen as such, he cannot tell the teacher-

3. Dershowitz (2000, p. 125) attributed the mutual-test hypothesis to Elie Wiesel and concluded that "No God should ever ask a father to kill his child, and no father should ever agree to do so." The same may be said about the Milgram situation.

participant to deal a lethal blow right away. The conflict must build gradually (as discussed above).

Second, the teacher-participant acts by flipping the shock switch, whereas Abraham stopped short of killing his son. The biblical game had an element of mind-reading that was absent in the Milgram situation. God intervened once Abraham convincingly displayed his intention to obey. Again, if God is construed as a player, He cannot be cast as omniscient. Sometimes, mind-reading goes wrong, and Abraham may have outsmarted God by signaling an insincere intention. If so, the Isaac story can be read as a game of chicken, in which God blinked first (Schelling, 1960). The game could have been Abraham's test of God as much as it was God's test of Abraham.[3] The suggestion that, by analogy, Milgram's participants may have been testing the experimenter breaks a taboo. Experimenters are supposed to test their subjects, not be tested by them.

Third, Milgram's teacher-participant received strong information about the confederate's state, but no irrefutable proof of his demise. The learner-confederate could be injured, unconscious, or even dead. In contrast, Abraham's game ended as soon as he had elected his response strategy with Isaac's fate known without a doubt. The prevailing view is that the learner must have been convinced that the confederate was dead. It is this assumption that qualifies Milgram's study as an analogue for atrocities committed during the Holocaust. This view can be questioned. Like Asch, Milgram allowed uncertainties. He presented his shock generator as a delivery device of voltage. However, the severity of the punishment—if it had actually occurred—would have depended on the strength of the current as measured in ampères. When the current is weak, even a high voltage yields little electrical power. It may stun, but not kill. Perhaps this is what the experimenter was hinting at when he assured the participant that "although these shocks may be painful, there is no permanent tissue damage" (Milgram, 1974, p. 27). To assume that obedient shocking is irrational if people know Ohm's Law, then regarding shocking behavior as irrational begs the question of people's knowledgeability.[4] These uncertainties are critical for any evaluation of rationality, and because such uncertainties existed in Milgram's experiments, they cannot explain confirmed killings committed in the name of obedience. Once such confirmation is available, the actor cannot bet on the authority to raise the dead. Whether the authority relents or punishes is now irrelevant, as judgment passes to courts and juries. Hence, the game-theoretic analysis does not extend to Eichmann, although a more narrow, coherence-based reconstruction can be attempted (see Dawes, 1988, on Rudolf Höss, the Kommandant at Auschwitz).[5]

Fourth, Milgram ran many repetitions of his experimental protocol, one participant at a time. His findings thus involved individual differences. In contrast, the Hebrew god used only one subject, whose behavior cannot be compared with

4. Sheridan and King (1972), who replicated Milgram's work with a puppy as victim, had their participants deliver real shocks. While varying the voltage, the kept the amperage low to avoid damage.

5. When shocks exceeded 300 v, participants in most of Milgram's studies must have wondered if the learner was unconscious or dead. In the proximity experiment, however, the confederate and the participant were in the same room. The participant could assess the learner's state and presumably see that he was not dead. Milgram (1965) was uncharacteristically vague in his description of the procedures. Interestingly, the rate of obedience was *lower* in this experiment than in experiments that allowed participants to assume the learner had died.

that of others. This final difference is only tempered by the fact that the Hebrew writings also include a diverse set of legends. According to one legend, Abraham actually did kill Isaac (before God resurrected him; Spiegel, 1967). From a TOM point of view, the dynamics and the outcome of the game are the same, although less intriguingly cast than in the Genesis version of the story. There is no legend, as far as we know, according to which Abraham disobeys.

THE BYSTANDER EFFECT

The third instance of social misbehavior is the tendency to rely on others to help when encountering a person in need. In 1964, the murder of Kitty Genovese made headlines because it was presumably witnessed by 38 of her neighbors. No one intervened before it was too late, and a variety of theories sprang up to explain their apathy. Most of these theories were *ad hoc*, and most sought the causes of inaction in the personality dispositions of the neighbors. Darley and Latané (1968) proposed a social-psychological alternative, suggesting that the number of witnesses played a critical role. The more potential helpers there are, they argued, the more the individual feels released from a personal responsibility to help. Although a recent analysis of archival records suggests that the number of witnesses was smaller than previously thought (Manning, Levine, & Collins, 2007), the question remains how the size of the social group does affect and perhaps should affect an individual's decision to act altruistically.

Darley and Latané (1968) had a confederate simulate an epileptic seizure, which was overheard, though not seen, by a naïve participant who believed to be the only witness or to be among a group of witnesses. As expected, the individual's probability to intervene decreased with the number of potential helpers. Compared with conformity and obedience, the bystander effect inspired more ambivalent reactions. On the one hand, it has been said that Darley and Latané had "painted a sympathetic picture of the unfortunate bystander, forced to choose among courses of action hurriedly, on the basis of incomplete information, and under unfavorable cost and reward schedules" (Latané & Nida, 1981, pp. 308-309). On the other hand, Aronson (2003, p. 40) felt that the behavior of "the participants in the Darley-Latané experiments projects a rather grim picture of the human condition." The experimenters themselves seemed to doubt the diffusion of responsibility can be rational "by drawing analogies to crowd behavior" (Manning et al., 2007, p. 560).[6]

To consider the diffusion of responsibility irrational is to suggest that it violates the assumption of invariance (Dawes, 1998; Kahneman, 2003). Rational behavior is

6. Milgram and Hollander (1964) offered a more subtle analysis, but they also remained ambivalent with regard to the bystanders' rationality. They suspected that certain forces "inhibited rational action [for why would] people choose a course of action that probably shames them in retrospect?" (p. 603). Their tentative answer was that people considered direct, physical intervention to be the normative mandate, but that they shrank from intervention because of legitimate fears for their own safety. Considering a call to the police a second best, almost cowardly, alternative, they ended up doing nothing, which violated their preference ranking, according to which making a call would still be better than doing nothing. Yet, Milgram and Hollander also claimed that there "are risks even in minimal forms of involvement, and it is dishonest to ignore them" (p. 604). If so, preference rankings may not have been violated after all.

unaffected by irrelevant factors. Preference reversals (Lichtenstein & Slovic, 1971) and framing effects in decision-making (Tversky & Kahneman, 1981) are considered irrational because they involve judgments that are systematically contradictory. What should an individual do if the number of bystanders is irrelevant? A radical solution is to help with certainty; a relaxed solution is to allow some diffusion of responsibility, but not as much as has been observed empirically. Before these alternatives can be accepted as normative, their implications need to be examined.

The normative outcome of an emergency is typically construed from the victim's point of view. Ideally, the victim receives help from as many others as necessary, but no more. The bystander's rationality cannot, however, be evaluated from the victim's perspective, but only from the bystander's own perspective. An individual bystander can ensure that the victim receives help if there are no other bystanders. If others are present, and if they all help as if they were alone, effort is wasted and helpers might get in one another's way. The costs of overhelping nullify the radical alternative. If everyone intervened regardless of group size, the outcome would likely be counterproductive, if not catastrophic.

To avoid inefficiency, each individual helper must reduce the probability of helping, q, as the number of potential helpers, N, increases. One heuristic is to help with probability $1/N$. This possibility may seem attractive, but it does not pass scrutiny. There is no justification for the probability that at least one person helps, $p = (1-(1-q)^N)$, to become smaller as the group becomes larger. If the nature of the emergency is a constant, the probability of receiving help should also be a constant. If neither 1 nor $1/N$ are workable solutions for q, what is? Darley and Latané (1968) grant that most emergency situations have an element of uncertainty. Even the lone participant sometimes does not intervene because intervention might turn out to be a false alarm. A fight between a woman and a man might be a lovers' quarrel, and a person in distress might simply be experiencing a temporary fit of nervousness. Hence, it is reasonable to expect that $q_1 < 1$, and if so, $p < 1$. Furthermore, it follows that q_N must decrease with N so that $p = q_1$. This equality holds if

$$q_N = 1 - (1 - p)^{1/N} .$$

Darley and Latané's (1968, p. 380, footnote 3) data are consistent with this analysis. The individual's probability of helping decreased from .85 in the lone condition to .62 when one other bystander was thought to be present, and to .31 when the putative number of other bystanders was five. The victim's probability of receiving help remained at $p = .85 +/- .01$. This remarkable result stood the test of time. Averaging over 14 studies, Latané and Nida (1981) reported that $q_1 = .84$ and that $p = .89$. The trend of help being more probable in larger groups was not statistically significant. It seems that participants were attuned to the implications of the presence of others. They experienced a diffusion of responsibility, and they translated it into action in such a way that the probability of the victim receiving help was consistently high regardless of the size of the group.

From a game-theoretic perspective, the participants in the bystander experiments faced a volunteer's dilemma, where a person's decision is associated with

costs and benefits depending on what other people do (Rapoport, 1988). Passive bystanders receive the payoff T (Temptation) if at least one bystander intervenes. T can be understood as a psychological benefit to the passive bystander seeing the victim is being helped. An active bystander receives the payoff R (Reward), which captures the psychological benefit derived from helping. R is smaller than T and the difference is the cost of helping as that portion of T that is voluntarily sacrificed. Formal models of the volunteer's dilemma assume that costs and benefits are invariant with respect to group size (Diekmann, 1985; Franzen, 1999). If no one intervenes, everyone receives the payoff P (Penalty), which is smaller than R. Hence, the payoffs are ranked T > R > P. To simplify analysis, let only T vary, while R and T are constant with values of 1 and 0, respectively.

A rational person volunteers if the expected value of volunteering is greater than the expected value of defecting. To estimate expected values, the person needs to consider the payoff T, the probability q_e with which another individual will volunteer, and the number N of these individuals. The expected value of volunteering is greater than the expected value of defecting if $(1-(1-q_e)^N)T < 1$. This inequality can be solved for each of the three variables. First, the finding that volunteering is attractive if

$$q_e < 1-\left(\frac{T-1}{T}\right)^{1/N}$$

reveals the fundamental dilemma. A person cannot succeed by positively coordinating with others, but only by doing the opposite of what she believes others will do. To the extent that she believes others are likely to volunteer, her own defection becomes more attractive; to the extent that she believes others are unlikely to volunteer, her own volunteering becomes more attractive. Second, the finding that volunteering is attractive if

$$N < \frac{\ln(T-1)-\ln(T)}{\ln(1-q_e)}$$

confirms the observation that individuals are less prone to help as the group becomes larger.[7] Third, the finding that volunteering is attractive if

$$T < \frac{1}{1-(1-q_e)^N}$$

is consistent with the idea that a person is less motivated to defect if the marginal benefit of defection is small. The high probability of intervention in the bystander studies implies that participants perceived a low cost/benefit ratio. Supposing that participants correctly estimated the probability of individual others to intervene—which they appear to have done—their implied benefit of defection was T = 1.17, which in turn implies a cost/benefit ratio of .15, i.e., (1.17 – 1)/1.17.

Earlier models (Diekmann, 1985; Franzen, 1999) derive the probability of volunteering from costs, benefits, and group size, whereas the present approach also includes the person's estimated probability that someone else will intervene. These other models assume that $q_1 = 1$ because R > P, whereas the present model assumes

7. There is a point at which $\ln(T-1)-\ln(T) < \ln(1-q_e)N$. Since $0 < 1-q_e < 1$, $\ln(1-q_e) < 0$. Therefore, dividing both sides by $\ln(1-q_e)$ gives the inequality presented here.

that q_1 reflects the lone bystander's uncertainty regarding the true nature of the emergency. Despite these differences, all models lead to the conclusion that rational people become less likely to volunteer as the group becomes larger, while the probability that at least one person will volunteer remains relatively constant.

In the classic bystander studies and in the present analysis, the costs and benefits of helping are only estimated from the probability of helping. This approach presupposes that people become more willing to perform a costly act as the costs become proportionately smaller relative to the benefits (Weesie & Franzen, 1998). Other studies from the bystander paradigm suggest that people are rational in this sense. Latané and Darley (1968; see also Latané & Rodin, 1969) asked whether participants would report an apparent emergency if others were present. While participants were filling out surveys, a vapor seeped into the room. Participants had to decide whether this vapor was dangerous smoke. When three naïve participants were in the room, the probability of anyone reporting to the experimenter was lower (p = .38) than the probability of a single participant reporting (q = .75). This situation was not a pure volunteer's dilemma because participants could see one another; hence, they ran a greater risk of embarrassing themselves if they were to sound a false alarm. Darley and Batson's (1973) Good-Samaritan study also provided evidence for the impact of psychological costs. Bystanders (or rather bywalkers) were less likely to stop and assist a prone man with a cough the more they were under time pressure to accomplish an academic goal.

The decision to help should be reached more easily as the costs of helping decrease, as the rewards for helping increase, and as the costs of not helping increase. Each of these three assumptions has empirical support (Franzen, 1999; Levy et al., 1972; Penner, Dovidio, Piliavin, & Schroeder, 2005). Staging emergencies in New York City subways, Piliavin, Rodin, and Piliavin (1969) found that bystanders were more likely to help once someone else had intervened. This release from inhibition may have reflected a steeper decrease in the perceived costs of helping than in the perceived benefits of not helping. Piliavin and Piliavin (1975) drew an important connection between the effect of the cost/benefit ratio and the effect of group size. They found that the bystander effect was stronger when costs were high than when costs were low.

From our quantitative model, we know that the probability of a single bystander to intervene, q_1, decreases with rising costs. The diffusion of responsibility effect is the difference between q_1 and q_N under the assumption that p, the probability that anyone will help, remains equal to q_1. The diffusion of responsibility effect, q_1-q_N, is curvilinear. When costs are low to begin with, increases in costs first yield larger and then smaller bystander effects at first. The maximum effect occurs when

$$q_1 = 1 - N^{N/(1-N)}.$$

For example, for N = 2, the bystander effect is largest when q_1 = .75; for N = 10, the effect is largest when q_1 = .93. Under most conditions, increases in the personal costs of helping make the bystander effect smaller instead of larger, although under specific conditions, the opposite can occur as the Piliavins showed.

In a pure volunteer's dilemma, all players choose at the same time and without knowledge of one another. The experiments on the bystander effect were not pure in this sense. The seizure study and its replications were terminated once the person volunteered or once a preset maximum window of time had closed. Believing

that others were contemplating the same choice between intervention and passivity, the participant could not be indifferent to the passage of time. One interpretation is that each passing second reduced the cost/benefit ratio, putting greater pressure on participants to act in order to avoid further erosion. Franzen (1999) suggested that in a time-sensitive volunteer's dilemma, the probability of at least one person helping increases with the size of the group. This did not happen in the seizure study, which suggests the presence of a countervailing force. A participant noticing that others have not intervened might conclude that these others do not perceive the situation as a genuine emergency. This dynamic introduces the possibility of an informational cascade as discussed in the context of the conformity experiments.

A common assumption underlying the bystander paradigm is that although participants may experience legitimate uncertainties regarding the nature of the incident, the outside observer knows that a real emergency is on hand—or rather, a bogus but well-staged one—and that the participant should respond. In contrast, not all apparent emergencies turn out to be genuine. A full model of rational behavior must include the possibility that both, informational cascades suggesting action and cascades suggesting inaction, may turn out to be correct.

In the bystander literature, participants who fail to act because others are passive, are portrayed as victims of pluralistic ignorance (Latané & Rodin, 1969). There are two attempts to cure this presumed cognitive illusion. One remedy is for the participant to ignore others and to act as if alone. As shown above, this strategy is inefficient because it imposes greater costs on the collective than a single volunteer would. Another remedy is social projection (Krueger, 2002). Participants could be advised to ignore the inaction of others, and to assume that these others interpret the situation as they themselves do. This strategy is successful in that it keeps p from becoming smaller with increasing N. A person who has not chosen yet between action and passivity will assume that others are also stuck in this predecision phase. Their apparent inaction does not signal a final decision, and therefore does not feed into a negative cascade. In other words, social projection turns an impure, time-sensitive volunteer's dilemma back into a pure one.

When bystanders are able to communicate, the TOM (Brams, 1994) suggests a different solution to the time-sensitive volunteer's dilemma. Treating the dilemma as a multi-person game of chicken, a rational player refuses to help, but only after making sure that others know this. Then, the others must find another helper. The problem here is twofold. First, the player must find a way to signal to the others that she cannot volunteer (e.g., by throwing the telephone out of the window in the Genovese situation; see Schelling, 1960, for examples of acting crazy for rationality's sake). Second, if the first player acts crazy enough to be excluded from the pool of volunteers, the problem continues in the remaining pool until only one bystander is left who is free to act. The risk is that by the time the TOM finds a solution, the victim is dead.

There are two main reasons for why the studies on the bystander effect became morality tales. One reason is that the data have been misunderstood. According to a common misconception they showed that "someone in need of emergency aid would have a better chance of survival if a single bystander, rather than a crowd, were present" (Cialdini, 2001, p. 115). This misperception is a powerful one; it trapped even one of the original authors (see Latané & Rodin, 1969, p. 190).

The other reason is also perceptual. Specifically, the cost of helping is difficult to assess from the outside. Observers can be tempted to trivialize the helper's costs to the point of making them virtually zero. If helping were costless, there would be no volunteer's dilemma, and the situation would be trivial. The victim, the experimenter (or the public), and the participant in a bystander study have different interests. Only the interests of the latter are relevant for a rational reconstruction of bystander behavior. The victim's preference for a high probability of being helped regardless of group size is irrelevant, and so is the experimenter's wish that the probability of help should increase with the size of the group. What matters is the participant's conflict between wishing to keep his or her own probability of helping as low as possible, while simultaneously not letting the victim be hurt by increases in group size. The participants in Darley and Latané's (1968) managed to do this. It makes no sense to demand from the participant to select a probability of helping that allows the victim's probability of being helped by someone to increase at a certain rate with the size of the group. Which rate should that be? Participants who follow this line of reasoning might as well decide to help for sure. As we have seen already, this would be rational only if the cost of helping were zero.

RATIONALITY OR RATIONALIZATION

Caving in to peer pressure, obeying destructive authority, and relying on others to help are widely seen as embarrassing human failures that define the field of social psychology. It is easy to condemn these acts—or failures to act—and to rationalize one's moral outrage (Haidt, 2001). One form of rationalization is the belief that immoral acts must also be irrational (Ketelaar, 2006). Such an inference is itself irrational. It is a simple heuristic that leads to systematic error. The charge of irrationality is a serious one, which deserves a rigorous defense.

The first line of defense is to show that the claim of the prosecution does not overcome the threshold of reasonable doubt. In none of the three behavioral domains can it be shown that the participants' behavior leads to systematic incoherence. Yet, showing that a behavior is not irrational may not convince an audience that it is, in fact, rational. To do this, a second line of defense is needed, and to provide one was the purpose of this article. The use of explicit criteria for rational behavior reveals that in all three areas of study, behavior follows coherent patterns. In the Asch paradigm, conformity is sensitive to the size of the group and the presence of allies; in the Milgram paradigm, obedience is sensitive to the distances among the players and the institutional prestige; in the Darley and Latané paradigm, helping is sensitive to the size of the group and various cost and benefit factors.

Milgram (1979, p. 139) noted that "it is certainly legitimate to accept the behavioral facts and carry out arguments in regard to the psychological attitudes that lie behind them" (p. 139). Yet, any reconstruction of rationality must face the question of whether it is a mere exercise in post-hoc modeling. The answer lies again in the patterns. Since it is the patterns that speak to the behavior's rationality, it is also the patterns that reveal what irrational behavior would look like. Asch's participants would have behaved irrationally if their willingness to conform rose nonmonotonically with the size of the group, or if their probability of yielding

changed over trials.[8] Milgram's participants would have behaved irrationally if they had shown more than one point of indifference along the voltage scale, or if they had never obeyed or never disobeyed at all. Darley and Latané's participants would have behaved irrationally if they helped no matter what, never helped, or helped with a probability that made the victim's chances dependent on the number of bystanders.

As all of these irrationalities could have occurred, but did not, one may conclude that the observed behavioral patterns corroborate the rationality hypothesis (Roberts & Pashler, 2000). Given that the situations created for the participants were complex, demanding, and fraught with emotional tension, one must conclude that the participants acquitted themselves rather well. Indeed, they acquitted themselves well enough to permit the hypothesis that their behavior was the result of sophisticated strategic reasoning. This hypothesis was proposed and evaluated in three specific contexts. It is to be hoped that the generalizability of this hypothesis will be the object of future theoretical and empirical work. Like irrationality, rationality is not a foregone conclusion. For example, bad cascades do occur (according to the model, they must occur with a knowable probability). Being rational now does not protect one from looking foolish later. Coming to the defense of rationality on a case-by-case basis is a different endeavor than categorically claiming that human thought and behavior is rational. Such claims are currently *de rigueur* among proponents of social intelligence (Kihlstrom & Cantor, 2000) and adaptationism (Kameda & Tindale, 2006). Yet, human do make errors and mistakes, and some of them are avoidable (Ariely, 2008).

SOCIAL BEHAVIOR WITH EYES WIDE SHUT

The orthodox social-psychological view is unsympathetic to decision-theoretic and game-theoretic analysis because it has little use for strategic reasoning. When it is considered at all, strategic reasoning tends to take the form of participants being wise to demand characteristics, which then undercut all substantive conclusions (Orne, 1962). In orthodox analysis, the causes of bad behavior are typically sought in a Lewinian force field of bad dispositions and corrosive social situations (Zimbardo, 2007). Here, orthodoxy tends to be incompatibilist, claiming that any evidence for situational causation is evidence against personal causation (see Krueger, 2009, for a critique of the hydraulic person-situation model).

In recent years, the situationist assault on human rationality has promoted the view that social behavior is overwhelmingly automatic, and that therefore the famous instances of misbehavior must also be understood in terms of automaticity. In their Handbook chapter on "control and automaticity in social life," Wegner and Bargh (1998) claimed that "the classic studies [including the one's reviewed here] highlight automatic forms of human responding" (p. 484). "As a rule," they say, "people in these powerful situations don't acquit themselves very well, as they succumb to pressures that make them do things ranging from merely uncharitable to frighteningly robotic" (p. 447).

8. Ariely and Levav (2000) found that when patrons in a restaurant order in sequence, they—irrationally—tend to counterconform and end up eating meals they do not like.

This claim is not true and it is not new. In a previous edition of the Handbook of Social Psychology, Moscovici (1985) had strong words to say about Asch's research. He claimed that Asch's findings defeated his goal to refute the doctrine of prestige-suggestion. Instead, "the Asch experiment exemplifies an experiment whose value lies in the fact that it falsified what it set out to verify and clearly invalidated his theory. It serves, on the contrary, as one of the most dramatic illustrations of conformity, of blindly going along with the group, even when the individual realizes that by doing so he turns his back on reality and truth" (p. 349).

More recently, Cialdini (2001) argued that yielding to social influence often takes the form of "fixed-action patterns [that is] mechanical behavior sequences" (p. 17). When called on by an authority figure to act, "we rarely agonize" (p. 185) over a decision. "Once we realize that obedience to authority is mostly rewarding, it is easy to allow ourselves the convenience of automatic obedience" (p. 186).

According to the automaticity paradigm, even higher-level behaviors can be elicited without conscious mediation (Bargh, 2007). The new, smart, unconscious can produce behavior that appears to be strategic (Bargh & Morsella, 2008). Hence, some misbehavior, such as conformity (Epley & Gilovich, 1999) or bystander inaction (Garcia, Weaver, & Moskowitz, 2002) only requires appropriate environmental primes. Although we are impressed with these findings, we do not believe that the study of minimally sufficient conditions of a behavior necessarily account for the entire category of such behaviors. One problem is that the automaticity hypothesis has difficulty explaining striking individual differences in the behavior of people exposed to the same situation (Cesario, Plaks, & Higgins, 2006; Krueger, 2009). Another problem is that the classic social influence experiments involved conditions that forced conscious participation and emotional reactivity. The respondents' ultimate decision to yield or to resist could not be reached by mere schema activation. We agree with Milgram (1979) who noted that "obedience occurs not as an end in itself, but as an instrumental element in a situation that the subject construes as significant and meaningful" (p. 146).

The casting of socially influenced behavior as exclusively automatic has the hallmarks of an informational cascade. The more often this view is presented, the more credible it will seem. There is reason to believe, as we have argued, that this cascade is a bad one. For a restoration of balance in social psychology, the notorious fragility of cascades gives some reason for hope.

RECONSTRUCTING MORALITY?

Assuming that we have made the case for the rationality of the individuals who participated in classic social psychological research, the question remains of how to evaluate their actions in moral terms. The view presented here is a rational reconstruction of certain behaviors has no necessary implications for judgments of morality. Rational behavior can be moral, but it can also be immoral. Many people share this compatibilist view. They hold perpetrators of heinous crimes responsi-

9. I am borrowing this phrase from the academic impostor Ward Churchill against my own better judgment. A case of *akrasia* if you will.

ble even when believing that all actions are fully determined by causes preceding personal deliberation and intention (Nichols & Knobe, 2007). Even some situationists agree. Zimbardo (2007), for example, insists that psychology is not excusiology. He suggests that the situation explains why United States Army personnel committed atrocities at the Iraqi prison Abu Ghraib in 2004; yet, some personal responsibility remains. As we noted earlier, a dual attribution of behavior to both the situation and the person is inconsistent with the premise of situationism itself (which is incompatibilist; Krueger, 2009).

Our compatibilist approach does not face this problem of contradiction. We can contemplate the possibility that some yielding to authority is not immoral. Milgram's participants may have been well-intentioned in the sense that they accepted a commitment to the idea that they would help science and education. In the course of the experiment, their moral imperatives were split. The moral responsibility to help the experimenter was gradually overshadowed by their moral responsibility to do no harm. Those who continued to obey remained true to one of their moral mandates, which arguably turned out to be the less important one.

POSTSCRIPT ON EICHMANN

For better or for worse, modern social psychology evolved in the shadow of the Holocaust. To understand how millions of Germans and other Europeans could participate in mock-medieval rituals, carry out the inhumane orders of their superiors, and turn a blind eye when their neighbors were taken away to certain death, is to understand a critical element of human nature. Social psychology tackled this difficult question with ingenious research programs, which collectively suggest that anyone can act despicably given the right circumstances.

The idea that social misbehavior (and "evil") arises from irrational psychological processes is oddly comforting because it suggests that with a little clear thinking, matters can improve. This is an illusion. Clear thinking and bad behavior can coexist. In fact, bad behavior that is also rational is far scarier than bad behavior that is irrational. To change the former kind of behavior, it may be necessary to engage nonrational forces, such as emotions or social norms.

Adolf Eichmann was the only Nazi war criminal that was tried and executed in the state of Israel. His ghost haunts social psychology. Are we all "Little Eichmanns,"[9] ordinary people who can become monsters if the circumstances demand it? Hannah Arendt's (1963) famous "report on the banality of evil" suggested that we are, and Milgram cited it approvingly. Contemporary research questions this view. Cesarani (2006) concludes that Eichmann was not a shadowy bureaucrat who committed crimes of obedience from the safety of his desk. Instead, "Eichmann was a forceful personality who acted with zeal and initiative" (p. 358). To the extent that he followed orders, he did so creatively, not blindly. He showed no signs of "Kadavergehorsam" (obedience to the death), although this thanatological value was drilled into a younger and more impressionable generation. When Nazi rule came to an end, Eichmann did not go down with the ship, but took care to preserve his own life.

Eichmann is the wrong model for the participant in a psychological study on social influence. In the Milgram paradigm, he resembles the experimenter more closely than the participant-teacher. Why did the experimenter not resist Milgram's instructions to prolong the participant's suffering? Arendt (1963) noted that Eichmann's guilt was compounded by his physical distance from the death camps. The verdict "took cognizance of the weird fact that in the death camps, it was usually the inmates and the victims who had actually wielded 'the fatal instrument with [their] own hands' " (p. 246). Quoting from the judgment, Arendt states that "to the extent to which any one of the many criminals was close to or remote from the actual killer of the victim means nothing, as far as the measure of his responsibility is concerned. On the contrary, in general the degree of responsibility increases as we draw further away from the man who uses the fatal instrument with his own hands" (p. 247). Unlike many social psychologists who feel confused about what their experiments say about human responsibility, Eichmann knew that he was guilty. As quoted by Cesarani (2006, p. 244), he realized "that I cannot wash my hands in innocence, because the fact that I was exclusively a receiver of orders is today meaningless." In Eichmann's case, there was little room to argue that he was moral in some other, less important, way (e.g., loyalty to the group).

If the present analysis is correct, the project of trying to account for immoral behavior with reference to irrational thought has failed. Human destructiveness and apathy are best confronted on their own plane, which is a moral one. To let go of the hope that rationality can be a servant to morality might be the wise choice. It has been suggested many times, but ignored just as often.[10]

10. Of those who have counseled compassion, my favorites are Spinoza, Schopenhauer, Russell, and the Dalai Lama.

APPENDIX

To prove that $q > p$ if $.5 < p < 1$, we note that $q > p$ if and only if $(p+1)/[p(p+1) + (p-2)(p-1)] > 1$, or $(p+1) > [p(p+1) + (p-2)(p-1)]$, or $(p+1)-p(p+1) > (p-2)(p-1)$, or $(1-p)(p+1) > (p-2)(p-1)$, or $p+1 > -(p-2)$, or $p+1 > 2-p$, or $2p > 1$, or $p > 0.5$. Therefore, $q > p$ if and only if $.5 < p < 1$. As these inequalities are already assumed, it follows that $q > p$.

To find the maximum value for the difference q-p, we take the derivative of $f(p) = (p(p+1)/[(p(p+1)) + (p-1)(p-2)]) - p$, which is

$([(2p+1)(2p^2 - 2p + 2) - (4p-2)(p^2 + p)]/[(2p^2 - 2p + 2)^2]) - 1$ and set it to 0.

Adding 1 to both sides of the equation, we obtain $[(2p+1)(2p^2 - 2p + 2) - (4p-2)(p^2 + p)]/[(2p^2 - 2p + 2)^2] = 1$.

As the denominator is never zero for $1/2 < p < 1$, we multiply both sides by it to get $(2p+1)(2p^2 - 2p + 2)) - (4p-2)(p^2 + p) = (2p^2 - 2p + 2)^2$.

Expanding this out gives $4p^3 - 2p^2 + 2p + 2 - 4p^3 - 2p^2 + 2p = (2p^2 - 2p + 2)^2$,

And combining like terms yields $-4p^2 + 4p + 2 = (2p^2 - 2p + 2)^2$.

Expanding out the second side then yields $-4p^2 + 4p + 2 = 4p^4 - 8p^3 + 12p^2 - 8p + 4$.

Adding $4p^2 - 4p - 2$ to both sides yields $4p^4 - 8p^3 + 16p^2 - 12p + 2 = 0$.

Solving this polynomial with a TI-86 calculator, we find that $p = 0.76995$. The values of p that make this polynomial 0 are the values of p that make the derivative 0; that is, they are the critical points. Since $q - p > 0$ for all p between .5 and 1, and $q = p$ at $p = .5$ and $p = 1$, we can conclude that this critical point gives us a global maximum on $.5 < p < 1$, which is what we wanted to find.

REFERENCES

Ainslee, G. (2001). *Breakdown of will*. New York: Cambridge University Press.

Arendt, H. (1963). *Eichmann in Jerusalem: A report on the banality of evil*. New York: Viking Press.

Ariely, D. (2008). *Predictably irrational*. New York: HarperCollins.

Ariely, D., & Levav, J. (2000). Sequential choice in group settings: Taking the road less traveled and less enjoyed. *Journal of Consumer Research, 27*, 279-290.

Arkes, H. R., & Blumer, C. (1984). The psychology of sunk cost. *Organizational Behavior and Human Performance, 35*, 124-140.

Arkes, H. R., & Mellers, B. A. (2002). Do juries meet our expectations? *Law and Human Behavior, 26*, 625-639.

Aronson, E. (2003). *The social animal* (9th ed.). New York: Worth Publishers.

Asch, S. E. (1952). Effects of group pressure on the modification and distortion of judgments. In G. E. Swanson, T. M. Newcomb, & E. L. Hartley (Eds.), *Readings in*

social psychology (2nd ed., pp. 2-11). New York: Holt.

Asch, S. E. (1956). Studies of independence and conformity: I. A minority of one against a unanimous majority. *Psychological Monograph, 70*(9, whole number 416).

Banerjee, V. (1992). A simple model of herd behavior. *Quarterly Journal of Economics, 107*, 797-817.

Bargh, J. A. (2007). *Social psychology and the unconscious*. New York: Psychology Press.

Bargh, J. A., & Morsella, E. (2008). The unconscious mind. *Perspectives on Psychological Science, 3*, 73-79.

Becker, G. S. (1991). A note on restaurant pricing and other examples of social influences on price. *Journal of Political Economy, 99*, 1109-1116.

Bernheim, B. D. (1994). A theory of conformity. *Journal of Political Economy, 102*, 841-877.

Bikhchandani, S., Hirschleifer, D., & Welch, I. (1992). A theory of fads, fashion, custom, and cultural change as informational cascades. *Journal of Political Economy, 100*, 992-1026.

Bond, R., & Smith, P. B. (1996). Culture and conformity: A meta-analysis of studies using Asch's (1952b, 1956) line judgment task. *Psychological Bulletin, 119*, 111-137.

Boyd, R., & Richerson, P. J. (2005). *The origin and evolution of culture*. New York: Oxford University Press.

Brams, S. J. (1994). *Theory of moves*. New York: Cambridge University Press.

Brams, S. J. (2003). *Biblical games* (2nd ed.). Cambridge, MA: MIT Press.

Burger, J. M. (2009). Replicating Milgram: Would people still obey today? *American Psychologist, 64*, 1-11.

Campbell, D. T. (1990). Asch's moral epistemology for socially shared knowledge. In I. Rock (Ed.), *The legacy of Solomon Asch: Essays in cognition and social psychology* (pp. 39-52). Hillsdale, NJ: Lawrence Erlbaum and Associates.

Cesarani, D. (2006). *Becoming Eichmann*. Cambridge, MA: Perseus Books.

Cesario, J., Plaks, J. E., & Higgins, E. T. (2006). Automatic social behavior as motivated preparation to interact. *Journal of Personality and Social Psychology, 90*, 893-910.

Cialdini, R. B. (2001). *Influence: Science and practice* (4th ed.). Boston, MA: Allyn and Bacon.

Crutchfield, R. S. (1955). Conformity and character. *American Psychologist, 10*, 191-198.

Darley, J. M., & Batson, C. D. (1973). 'From Jerusalem to Jericho': A study of situational and dispositional helping behavior. *Journal of Personality and Social Psychology, 27*, 100-108.

Darley, J. M., & Latané, B. (1968). Bystander intervention in emergencies: Diffusion of responsibility. *Journal of Personality and Social Psychology, 8*, 377-383.

Dawes, R. M. (1988). Plato vs. Russell: Hoess and the relevance of cognitive psychology. *Religious Humanism, 22*, 20-26.

Dawes, R. M. (1989). Statistical criterion for establishing a truly false consensus effect. *Journal of Experimental Social Psychology, 25*, 1-17.

Dawes, R. M. (1998). Behavioral decision making and judgment. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *The handbook of social psychology* (4th ed., Vol. 1, pp. 497-548). Boston, MA: McGraw-Hill.

Dershowitz, A. M. (2000). *The genesis of justice*. New York: Warner Books.

Deutsch, M., & Gerard, H. B. (1955). A study of normative and informational social influences upon individual judgment. *Journal of Abnormal and Social Psychology, 51*, 629-636.

Diekmann, A. (1985). Volunteer's dilemma. *Journal of Conflict Resolution, 29*, 605-610.

Dugatkin, L. A. (2000). *The imitation factor: Evolution beyond the gene*. New York: Free Press.

Epley, N., & Gilovich, T. (1999). Just going along: Nonconscious priming and conformity to social pressure. *Journal of Experimental Social Psychology, 35*, 578-589.

Franzen, A. (1999). The volunteer's dilemma: Theoretical models and empirical evidence. In M. Foddy, M. Smithson, S. Schneider, & M. Hogg (eds.), *Resolving social dilemmas: Dynamic, structural, and intergroup aspects* (pp. 135-148). New York: Psychology Press.

Freedman, J. L., & Fraser, S. C. (1966). Compliance without pressure: The foot-in-the-door technique. *Journal of Personality and Social Psychology, 4*, 195-203.

Garcia, S. M., Weaver, K., & Moskowitz, G. B. (2002). Crowded minds: The implicit bystander effect. *Journal of Personality and Social Psychology, 83*, 843-853.

Gigerenzer, G. (2008). Moral intuition = fast and frugal heuristics? In W. Sinnott-Armstrong (Ed.), *Moral psychology: Vol. 2. The cognitive science of morality: Intuition and diversity* (pp. 1-46). Cambridge, MA: MIT Press.

Graziano, W. G., Jensen-Campbell, L. A., Shebilske, L. J., & Lundgren S. R. (1993). Social influence, sex differences, and judgments of beauty: Putting the interpersonal back in interpersonal attraction. *Journal of Personality and Social Psychology, 65*, 522-531.

Griskevicius, V., Goldstein, N. J., Mortensen, C. R., Cialdini, R. B., & Kenrick, D. (2006). Going along versus going alone: When fundamental motives facilitate strategic (non)conformity. *Journal of Personality and Social Psychology, 91*, 181-294.

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

Hodges, B. H., & Geyer, A. L. (2006). A nonconformist account of the Asch experiments: Values, pragmatics, and moral dilemmas. *Personality and Social Psychology Review, 10*, 2-19.

Huck, S., & Oechssler, J. (2000). Informational cascades in the laboratory: Do they occur for the right reasons? *Journal of Economic Psychology, 21*, 661-671.

Hung, A. A., & Plott, C. R. (2001). Information cascades: Replication and an extension to majority rule and conformity-rewarding institutions. *The American Economic Review, 91*, 1508-1520.

Judd, C. M., James-Hawkins, L., Yzerbyt, V., & Kashima, Y. (2005). Fundamental dimensions of social judgment: Understanding the relations between judgments of competence and warmth. *Journal of Personality and Social Psychology, 89*, 899-913.

Kahneman, D. (2003). A perspective on judgment and choice: Mapping bounded rationality. *American Psychologist, 58*, 697-720.

Kameda, T., & Tindale, S. (2006). Groups as adaptive devices: Human docility and group aggregation mechanisms in evolutionary context. In M. Schaller, J. A. Simpson, & D. T. Kenrick (Eds.), *Evolution and social psychology* (pp. 317-341). New York: Psychology Press.

Kelley, H. H. (1972). Causal schemata and the attribution process. In E. E. Jones, D. E. Kanouse, H. H. Kelley, R. S. Nisbett, S. Valins, & B. Weiner (Eds.), *Attribution: Perceiving the causes of behavior* (pp. 151-174). Morristown, NJ: General Learning Press.

Ketelaar, T. (2006). The role of moral sentiments in economic decision making. In D. De Cremer, M. Zeelenberg, & J. K. Murnighan (Eds.), *Social psychology and economics* (pp. 97-116). Mahwah, NJ: Erlbaum.

Keynes, J. M. (1936). *The general theory of employment, interest and money*. London, UK: Macmillan.

Kihlstrom, J. F., & Cantor, N. (2000). Social intelligence. In R. J. Sternberg (Ed.), *Handbook of intelligence* (pp. 359-379). New York: Cambridge University Press.

Krueger, J. I. (2002). On the reduction of self-other asymmetries: Benefits, pitfalls, and other correlates of social projection. *Psychologica Belgica, 42*, 23-41.

Krueger, J. I. (2003). Return of the ego—self-referent information as a filter for social prediction: Comment on Karniol (2003). *Psychological Review, 110*, 585-590.

Krueger, J. I. (2007). From social projection to social behaviour. *European Review of Social Psychology, 18*, 1-35.

Krueger, J. I. (2009). A componential model of situation effects, person effects and situation-by-person interaction effects on social behavior. *Journal of Research in Personality, 43*, 127-136.

Krueger, J. I., & Acevedo, M. (2007). Perceptions of self and other in the prisoner's dilemma: Outcome bias and evidential reasoning. *American Journal of Psychology, 120*, 593-618.

Krueger, J. I., Massey, A. L., & DiDonato, T. E. (2008). A matter of trust: From social preferences to the strategic adherence of social norms. *Negotiation & Conflict Management Research,1*, 31-52.

Kruglanski, A. W., Raviv, Bar-Tal, D., Raviv, Al., Sharvitt, K., Ellis, S., Bar, R., Pierro, A., & Mannetti, L. (2005). Says who: Epistemic authority effects in social judgment. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 37, pp. 345-392). New York: Academic Press.

Kuhn, T. S. (1962). *The structure of scientific revolutions*. Chicago, IL: University of Chicago Press.

Latané, B., & Darley, J. M. (1968). Group inhibition of bystander intervention in emergencies. *Journal of Personality and Social Psychology, 10*, 215-221.

Latané, B., & Nida, S. (1981). Ten years of research on group size and helping. *Psychological Bulletin, 89*, 308-324.

Latané, B., & Rodin, J. (1969). A lady in distress: Inhibiting effects of friends and strangers on bystander intervention. *Journal of Personality and Social Psychology, 5*, 189-202.

Levy, P., Lundgren, D., Ansel, M., Fell, D., Fink, B., & McGrath, J. E. (1972). Bystander effect in a demand-without-threat situation. *Journal of Personality and Social Psychology, 24*, 166-171.

Lichtenstein, S., & Slovic, P. (1971). Reversals of preferences between bids and choices in gambling decisions. *Journal of Experimental Psychology, 89*, 46–55.

Luce, R. D., & Raiffa, H. (1957). *Games and decisions*. New York: Wiley.

Manning, R., Levine, M., & Collins, A. (2007). The Kitty Genovese murder and the social psychology of helping: The parable of 38 witnesses. *American Psychologist, 62*, 555-562.

Milgram, S. (1961, December). Nationality and conformity. *Scientific American*, 45-51.

Milgram, S. (1963). Behavioral study of obedience. *Journal of Abnormal and Social Psychology, 67*, 371-378.

Milgram, S. (1965). Some conditions of obedience and disobedience to authority. *Human Relations, 18*, 57-76.

Milgram, S. (1974). *Obedience to authority*. Princeton, NJ: Princeton University Press.

Milgram, S. (1979). Interpreting obedience: Error and evidence—A reply to Orne and Holland. In A. G. Miller (Ed.), *The social psychology of psychological research* (pp. 138-154). New York: The Free Press.

Milgram, S., & Hollander, P. (1964, June 15). The murder they heard. *Nation*, 602-604.

Milgram, S., & Sabini, J. (1979). Candid camera. *Culture and Society, 16*, 72-75.

Monin, B., Sawyer, P. J., & Marquez, M. J. (2008). The rejection of moral rebels: Resenting those who do the right thing. *Journal of Personality and Social Psychology, 95*, 76-93.

Moscovici, S. (1985). Social influence and conformity. In G. Lindzey, & E. Aronson (Eds.), *Handbook of social psychology* (Vol. 2, pp. 347-412). New York: Random House.

Nichols, S., & Knobe, J. (2007). Moral responsibility and determinism: The cognitive science of folk intuitions. *Nous, 41*, 663-665.

Orne, M. T. (1962). On the social psychology of the psychological experiment: With particular reference to *demand characteristics* and their implications. *American Psychologist, 17*, 776-783.

Ottaviani, M., & Sørensen, P. (2000). Hard behavior and investment: Comment. *The American Economic Review, 90*, 695-704.

Packer, D. J. (2008). Identifying systematic disobedience in Milgram's obedience experiments: A meta-analytic review. *Perspectives on Psychological Science, 3*, 301-304.

Penner, L. A., Dovidio, J. F., Piliavin, J. A., & Schroeder, D. A. (2005). Prosocial behavior: Multilevel perspectives. *Annual Review of Psychology, 56*, 365-392.

Piliavin, I. M., & Piliavin, J. A. (1975). Costs, diffusion, and the stigmatized victim. *Journal of Personality and Social Psychology, 32*, 429-438.

Piliavin, I. M., Rodin, J., & Piliavin, J. A. (1969). Good Samaritanism: An underground phenomenon? *Journal of Personality and Social Psychology, 13*, 289-299.

Rapoport, A. (1988). Experiments with N-person social traps I: Prisoner's dilemma, weak prisoner's dilemma, volunteer's dilemma, and larges number. *Journal of Conflict Resolution, 32*, 457-472.

Roberts, S., & Pashler, H. (2000). How persuasive is a good fit? A comment on theory testing. *Psychological Review, 107*, 358-367.

Rook, L. (2006). An economic psychological approach to herd behavior. *Journal of Economic Issues, 40*, 75-95.

Rosenthal, R. W. (1981). Games of perfect information, predator pricing and the chain-store paradox. *Journal of Economic Theory, 25*, 92-100.

Rozin, P. (2007). Exploring the landscape of modern academic psychology: Finding and filling the holes. *American Psychologist, 62*, 754-766.

Salganik, M. J., Dodds, P. S., & Watts, D. J. (2006). Experimental study of inequality and unpredictability in an artificial cultural market. *Science, 311*, 854-856.

Schachter, S. (1951). Deviation, rejection, and communication. *Journal of Abnormal and Social Psychology, 46*, 190-208.

Schelling, T. C. (1960). *Strategy of conflict*. New York: Oxford University Press.

Sheridan, C. L., & King, R.G. (1972) Obedience to authority with an authentic victim. *Proceedings of the 80th Annual Convention of the American Psychological Association* (pp. 160–165). Washington, DC: American Psychological Association.

Skinner, B. F. (1955-56). Freedom and the control of men. *American Scholar, 25*, 47-65.

Spiegel, S. (1967). *The last trial*. Philadelphia, PA: The Jewish Publication Society of America.

Sunstein, C. (2003). *Why societies need dissent*. Cambridge, MA: Harvard University Press.

Tanford, S., & Penrod, S. (1984). Social influence model: A formal integration of research on majority and minority influence processes. *Psychological Bulletin, 95*, 189-225.

Tarde, G. de (1895). *Les lois de l'imitation; étude sociologique*. Paris: Alcan.

Trotter, W. (1916). *Instincts of the herd in peace and war*. London, England: Oxford University Press.

Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science, 211*, 453-458.

Van Lange, P. A. M. (1999). The pursuit of joint outcomes and equality in outcomes: An integrative model of social value orientation. *Journal of Personality and Social Psychology, 77*, 337-349.

Weesie, J., & Franzen, A. (1998). Cost sharing in a volunteer's dilemma. *Journal of Conflict Resolution, 42*, 600-618.

Wegner, D. M., & Bargh, J. A. (1998). Control and automaticity in social life. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *The handbook of social psychology*, Vol. 1 (4th ed., pp. 446-496). New York: McGraw-Hill.

Williams, K. D. (2007). Ostracism. *Annual Review of Psychology, 58*, 425-452.

Zajonc, R. B. (1999). One hundred years of rationality assumptions in social psychology. In A. Rodrigues & R. V. Levine (Eds.). *Reflections on 100 years of experimental social psychology* (pp. 200-214). New York: Basic Books.