

Hawkish Biases*

Daniel Kahneman
Eugene Higgins Professor of Psychology, Emeritus
Professor of Psychology and Public Affairs, Emeritus
Woodrow Wilson School
Princeton University
kahneman@princeton.edu

Jonathan Renshon
Weatherhead Center for International Affairs
Department of Government
Harvard University
jrenshon@fas.harvard.edu

Daniel Kahneman is a Senior Scholar at the Woodrow Wilson School of Public and International Affairs and a fellow of the Center for Rationality at the Hebrew University in Jerusalem. Daniel Kahneman won the Nobel Prize in Economic Sciences in 2002 for his work on prospect theory and decision-making under uncertainty.

Jonathan Renshon is a PhD candidate in the Department of Government at Harvard University. Renshon studies national security decision-making, foreign policy analysis and political psychology. He is the author of Why Leaders Choose War: The Psychology of Prevention (Greenwood, 2006), and his work has appeared in the *Journal of Strategic Studies*, *Political Psychology*, *Foreign Policy* and the *Journal of Conflict Resolution*.

* This essay appears in a slightly shorter and modified form in Trevor Thrall and Jane Cramer (eds.) American Foreign Policy and the Politics of Fear: Threat Inflation Since 9/11 (New York: Routledge Press, 2009), 79-96.

In a short article in *Foreign Policy* in 2007, we put forward a hypothesis of “directionality” in cognitive biases that generated a fair amount of debate and some controversy (Kahneman and Renshon 2007). This chapter is an attempt to expand upon and clarify our original argument.

In the last few decades, cognitive and social psychologists have described many cognitive biases –predictable errors in the ways that individuals interpret information and make decisions.¹ An unexpected and significant pattern emerges when these biases are viewed as a set: we find, almost without exception, that the biases recently uncovered by psychological research favor hawkish decisions in conflict situations. We use the term “hawkish” to describe a propensity for suspicion, hostility and aggression in the conduct of conflict, and for less cooperation and trust when the resolution of a conflict is on the agenda. Actors who are susceptible to hawkish biases are not only more likely to see threats as more dire than an objective observer would perceive, but are also likely to act in a way that will produce unnecessary conflict. We do not contend, of course, that all decisions in the international political context will be hostile or aggressive as a result of biases of cognition and preference, only that more decisions will be so more often than they would be in the absence of bias. Nor do we contend that all suspicions are produced by biased thinking, and therefore unjustified. Our point is only that the collective effect of the biases that psychology has identified is to increase the probability that agents will act more hawkishly than an objective observer would deem appropriate. For a more concrete image, suppose that a national leader is exposed to conflicting advice from a hawk and a dove. Our contention is that cognitive biases will tend to make the hawk’s arguments more persuasive than they deserve to be.

The biases discussed in this chapter pose a difficult methodological problem for political scientists, since we cannot “prove” the bias to have been at fault in any given decision. Instead, we define the bias and invite readers to consider its consequences in conflict situations by invoking a hypothetical objective observer. We can only hope that

¹ More specifically, a bias exists when an error in estimating or assessing a value is more likely in one direction than another. For example, in estimating our own skills as drivers, a bias exists because we are far more likely to *over*-estimate our abilities relative to others than to *under*-estimate them. In the absence of an objective criterion, we take the opinions of knowledgeable but uninvolved observer as a definition of an unbiased judgment. The perspective of history provides another “unbiased” view of the situation that decision makers faced.

this perspective allows dispassionate analysts to take that role. And we can only hope that the retrospective judgments are not overly tainted by the familiar hindsight bias, which makes it all too easy to explain past disasters by finding flaws in the decisions of actors who cannot defend themselves. The test of our approach is whether it offers historians and political scientists a useful heuristic. The biases that we list are to be viewed as hypotheses, to be confirmed by examining evidence of what the decision makers believed, desired or feared at the critical time. For example, students of conflict should expect to find evidence of unreasonably negative interpretations of the opponent's intentions, and overly optimistic assessments of the situation by both sides, because these biases have been established by prior research.

In our original analysis, we compiled a list of known biases, and proceeded to trace the implications of those biases for international conflict. This chapter proceeds in much the same way. For each bias, we first present empirical evidence illustrating how and when it is likely to affect judgment and decision-making. Note that the evidence on which we rely was documented in experimental situations that did *not* involve conflict – the biases are considered to be general features of cognition and preference. We then proceed to examine the potential behavioral implications of each bias in situations in international conflict. The biases and their main effects are listed in Table 1.

Table 1. Biases examined in this chapter

Bias	Primary Effect in Conflict Situations
POSITIVE ILLUSIONS	Biased overconfidence raises the probability of violent conflict occurring and of deadlock in negotiations (when the parties overestimate their bargaining position or ability)
FAE	Perceive hostile actions of adversaries as due to unchanging, dispositional factors and discount the role of contextual factors; neglect the effects of one's own hostility on the behavior of adversaries
ILLUSION OF TRANSPARENCY	Ignore how one's actions are likely to be perceived by others, resulting in behavior that is likely to provoke aggression or hostility
ENDOWMENT EFFECT/LOSS AVERSION	Induces an aversion to making concessions and a reluctance to accept objectively "fair" exchanges.
RISK SEEKING IN LOSSES	Reluctance to settle, prolongation of conflict
PSEUDO-CERTAINTY	Lowers probability of concessions if there is a potential that those concessions might advantage an opponent in a possible future conflict and concurrently <i>raises</i> the probability of conflict occurring by adopting a worst-case scenario of the other's intentions.
REACTIVE DEVALUATION	Unconscious devaluation of offers, concessions or plans suggested by rivals or adversaries makes it difficult to reach agreement.

Positive Illusions

One of the most robust findings in cognitive and social psychology is that individuals often fall victim to “positive illusions.”² Among these positive illusions are: unrealistically positive views of one’s abilities and character, the illusion of control, and unrealistic optimism (Taylor and Brown 1988: 195-6).

Unrealistically positive views of the self have been documented in many domains. Among other things, most people believe themselves to be better than average drivers, decision-makers and negotiators (Svenson 1981: 143; Bazerman 1998: 69). A survey of university professors found that 94 percent believed themselves to be better teachers than the average at their institution, which of course is a statistical impossibility (Cross 1977). Because individuals resist information that conflicts with positive self-assessments, these unrealistically positive views of oneself are generally robust over time (Crary 1966: 246; Marks 1984: 203).

The “illusion of control” is an exaggerated perception of the extent to which outcomes depend on one’s actions. When people were given a button and instructed to cause a particular color to appear on the screen, they erroneously believed that they had substantial control over events, even when the outcomes were actually determined by a computer (Martin, Abramson et al. 1984). Experiments have shown that people act as if they can control the outcome of rolling a die, and are more willing to bet when they do the rolling (Silverman 1964: 114; Langer 1975: 312,324; Campbell 1986: 290). It has also been demonstrated that stress (common in conflict or crisis decision-making) increases the preference for strategies that engender a feeling of control, even if it is illusory, and even if it leads to worse outcomes (Friedland, Keinan et al. 1992: 923). In a competitive situation, the illusion of control causes each side to believe that the outcome of the competition depends mostly on its own actions and abilities, even when it depends equally on the achievements of competitors.

The third positive illusion is “unrealistic optimism.” The evidence for “illusory,” or biased, optimism comes from the comparisons of individuals’ judgments of themselves

² There is an important caveat to the findings described in this section. The first is that there may well be cultural variation in unrealistic optimism. Some studies have found that individuals from societies that do not place much emphasis on the individual are less likely to evince “self-enhancing” biases such as unrealistic optimism. As an example, one study found that Japanese were much less likely than Canadians to demonstrate unrealistic optimism. See (Heine and Lehman 1995).

and of others. People generally believe that the probability of positive outcomes (such as having a gifted child, or enjoying their first job) is higher for themselves than for their peers, and judge the probability of negative events (such as being the victim of a crime or being in a car accident) as less likely for themselves than for others (Robertson 1977: 136; Weinstein 1980: 806; Perloff and Fetzer 1986: 502). In addition, experimental evidence suggests that people's predictions of what *will* occur correspond closely to what they would *like* to happen, rather than what is objectively likely to occur (Sherman 1980: 211). A study of entrepreneurs who had started small businesses revealed a striking discrepancy between their expectations of success (typically .80 or more) and the actual probability of success for a small business, which is about 1/3 (Cooper, Woo et al. 1988). One important cause of this bias seems to be "reference group neglect," in which individuals discount the abilities or skills of the peer group against which they are competing (Camerer and Lovallo 1999: 307). Experts are not immune to these positive illusions. One recent experiment, for example, found that professional financial analysts, making judgments and predictions about their areas of expertise, were just as overconfident as the base group of students (Glaser, Langer et al. 2005).

Within political science, many scholars have found evidence supportive of the notion that leaders' positive illusions have led to more wars than would have occurred in the absence of that bias. Stephen Van Evera argued, for instance, that leaders often have unrealistically positive views of the balance of military power, overestimate their "will" relative to their adversary, overestimate the loyalty and abilities of key allies, and underestimate the cost of potential wars (Van Evera 1999).

A group of researchers has recently documented the link between overconfidence and war in a simulated conflict situation (Johnson, McDermott et al. 2006). Johnson et al. conducted an experiment in which participants (drawn from the Cambridge, MA area, but not exclusively composed of students) played an experimental wargame. Subjects gave ranked assessments of themselves relative to the other players prior to the game, and in each of the six rounds of the game chose between negotiation, surrender, fight, threaten or do nothing; they also allocated the fictional wealth of their "country" to either military, infrastructure or cash reserves. Players were paid to participate and told to expect bonuses if they "won the game" (there was no dominant strategy and players

could “win” using a variety of strategies). Players were generally overly optimistic, and those who made unprovoked attacks were especially likely to be overconfident (Johnson, McDermott et al. 2006: 2516).³

The consequences of positive illusions in conflict and international politics are overwhelmingly harmful. Except for relatively rare instances of armed conflicts in which one side knows that it will lose but fights anyway for the sake of honor or ideology, wars generally occur when each side believes it is likely to win -- or at least when rivals’ estimates of their respective chances of winning a war sum to more than 100 percent (Johnson 2004: 4). Fewer wars would occur if leaders and their advisors held realistic assessments of their probability of success; that is, if they were less optimistically overconfident.

Positive illusions also have damaging implications for negotiations. Neale and Bazerman found that overconfident negotiators (those with biased expectations of their likelihood of success) exhibited less concessionary behavior, and experienced more impasses, than realistically confident individuals (Neale and Bazerman 1985: 34; Bazerman 2001: 222).

Positive illusions generally favor hawkish, aggressive behavior when conflict exists or when a side already contemplates hostile actions. The implications of optimistic biases are less clear for tense situations that may lead either to conflict or to negotiated settlement. Actors can be overly optimistic about the prospects of negotiating a peaceful settlement. Of course, optimism in negotiations does not necessarily yield good outcomes.

Fundamental Attribution Error (FAE)

The Fundamental Attribution Error (FAE) is a systematic bias in the explanation of the behavior of others, and is perhaps the most studied bias in social cognition. Because the mental states of others are not directly observable, people inevitably rely on inferences to explain the behaviors they observe as due to personal dispositions or to situational pressures. The robust finding is that these causal attributions are biased,

³ This effect was statistically significant only for men.

exaggerating the role of the other's dispositions and intentions and discounting the role of the situation as the other perceives it.⁴

There is a vast literature on the FAE, but one of the earliest examples is still among the most evocative. In a famous early experiment, Jones and Harris asked participants to read short essays about Cuba that were either pro- or anti-Castro in content. In the "choice" condition participants were told that the writers had freely chosen the position for which they argued in their essay. In the "no choice" condition participants were told that the writers had been assigned the position for which they were to argue.

The participants then estimated the writers' actual attitude toward Castro. The surprising result of the experiment was that these estimates were strongly influenced by the position of the essay, *even when the position was not chosen by the writer*. Indeed, there was only a minor difference between the judgments made of the same essays in the "Choice" and "No choice" condition. (Jones and Harris 1967: 6). The authors of this classic experiment concluded that people are prone to attribute behaviors they observe to personal dispositions, and prone to neglect the influence of situational pressures – even the overwhelming pressure of a specific instruction to adopt a particular position in an essay (Jones and Harris 1967: 22). Some years later, the tendency to underestimate the role of the situation in explaining the behaviors of others was called the Fundamental Attribution Error (Ross 1977).

This attribution error is remarkably robust, and people who are informed about the bias are not immune to it. Students who had learned about the attribution error continued to over-emphasize dispositional factors and to neglect the importance of situational context or constraints (Darley and Batson 1973: 100; Pietromonaco and

⁴ While most social psychologists agree on the existence of the FAE, there is a still ongoing debate about *why* the attribution error occurs. One theory states that it is a result of the anchoring and adjustment heuristic. In this theory, a behavior (such as statement by an opposing leader threatening military action) provides an initial "anchor," leading the observer to perceive that the speaker has a correspondent attitude (being aggressive/belligerent). A correction is then made to take into account the amount of choice that the opposing leader had. However, because such adjustments are almost always insufficient, there results a pattern of biased attributions of behavior. Another explanation is that "behavior engulfs the field," implying that the behavior of the actor is almost always perceived as the most salient information, rather than the colorless background of the situation. Whatever the explanation for why the bias exists, the most relevant fact for our purpose is that it does exist, and biases our judgments of others' behavior. See (Heider 1958; Tversky and Kahneman 1974: 1128; Quattrone 1982: 596).

Nisbett 1982: 1). More recent research suggests that “rumination” (i.e., spending more time thinking about something) actually increases attribution errors. Subjects who were asked to take several minutes to imagine the motives, intentions and strategies of the other players were more likely to be suspicious of their partners in a computer game (Kramer 1994: 218-9). In line with previous research on the subject (Wilson and Kraft 1993: 409), rumination also increased the subjects’ confidence in the accuracy of their erroneous judgments.

Explanations of another person’s behavior reflect prior beliefs and attitudes toward that person: actions that are consistent with expectations are attributed to internal, or dispositional factors, while actions that appear inconsistent with prior beliefs are attributed to situational factors (Regan, Straus et al. 1974). Thus, subjects attributed to stable dispositions the good actions of people they liked and the bad actions of people they did not like, and attributed behaviors that violated expectations to fleeting, situational variables. Thus, the Fundamental Attribution Error –the tendency to over-attribute behavior to disposition - effectively reverses when disliked or distrusted actors commit positive actions.

Field evidence from the Middle East supports the same conclusion. Heradstveit conducted interviews with political activists in Egypt, Israel, Jordan, Lebanon and Syria and found strong support for the predictions of the fundamental attribution error: actors tended to over-attribute the hostile behavior of adversaries to dispositions and correspondingly disregarded contextual factors. However, this effect was reversed for positive behaviors (Heradstveit 1981: 4). Beliefs in the hostile intentions of adversaries tend to be self-perpetuating – and of course they also tend to be self-fulfilling.

When another country acts in an aggressive, belligerent or deceptive manner, the explanation of its behavior is of paramount importance. Were its actions driven by domestic political necessity -- or perhaps provoked by our own actions? Or do their hostile actions reflect the true goals and character of the other side? The alternative attributions lead to different policy choices. If, for instance, the behavior is a response to one’s own aggressive behavior, then attempts to restore trust may be appropriate. If, however, the same behavior reflects a deeper hostility, friendly gestures are likely to be futile.

The Fundamental Attribution Error strongly favors hawkish arguments in conflict situations. When hostility and suspicion already exist, actors will tend to attribute moderate behavior of antagonists to situational constraints (“they *had* to do that”) while attributing more hostile actions to dispositions. Bad behavior by adversaries will reinforce prior beliefs, while good behavior will be disregarded as “forced”. The hawkish position is justified both when the opponents yield and when they do not. Of course, antagonists do often have hostile dispositions, but that does not disprove our argument, which is that leaders will make these inferences to an excessive degree – beyond the level of suspicion that an objective observer would consider appropriate. Furthermore, the same bias will hinder efforts toward conciliation by causing leaders and their advisors to disregard the positive actions taken by adversaries.

Of course, the Fundamental Attribution Error will not cause hawkish arguments to prevail every time. And in some cases, history might judge an argument that was biased to have been correct in retrospect. One need only think of Europe in the 1930’s to wish that perhaps more politicians had adopted a dispositional attribution of Hitler’s actions, instead of excusing them as a reaction to the position of Germany following the Treaty of Versailles.

Illusion of Transparency

We have thus far made the case for hawkish arguments being advantaged by cognitive biases in how actors explain behavior. But how do people believe *their own* behavior will be explained? Individuals realize, of course, that they are not as transparent to others as they are to themselves. However, they typically do not make sufficient allowances for this difference in perspective. As a result, people tend to overestimate the extent to which their own feelings, thoughts or motivations “leak out” and are apparent to observers (Gilovich and Savitsky 1999: 167).

In recent demonstrations of this bias, participants in a “truth-telling game” overestimated the extent to which their lies were readily apparent to others, witnesses to a staged emergency believed their concern was obvious even when it was not, and negotiators overestimated the degree to which the other side understood their preferences (even in the condition in which there were incentives to maintain secrecy) (Gilovich,

Savitsky et al. 1998; Van Boven, Gilovich et al. 2003: 117). The common theme is that people generally exaggerate the degree to which their internal states are apparent to observers.

The transparency bias has pernicious implications for international politics. When the actor's intentions are hostile, the bias favors redoubled efforts at deception. When the actor's intentions are *not* hostile, the bias increases the risk of dangerous misunderstandings. Because they believe their benign intentions are readily apparent to others, actors underestimate the need to reassure the other side. Their opponents – even if their own intentions are equally benign -- are correspondingly more likely to perceive more hostility than exists and to react in kind, in a cycle of escalation. The transparency bias thus favors hawkish outcomes through the mediating variable of misperception.

The memoirs of U.S. Secretary of State Dean Acheson provide an illustration. Commenting on the American decision to drive to the Yalu River during the Korean War, he wrote that “no possible shred of evidence could have existed in the minds of the Chinese Communists about the non-threatening intentions of the forces of the United Nations (Jervis 1980: 583).” Though the U.S./U.N. forces did not have any intention of attempting to directly invade China, it should have been clear to Acheson and other U.S. decision-makers that their march toward the Yalu River would be perceived as threatening by Chinese leaders. Indeed, the People's Republic of China had already issued warnings that they would intervene militarily if any non-South Korean forces crossed the 38th parallel separating North and South Korea. Ignoring those warnings and being unable to see their actions as China would perceive them cost the U.S. dearly: China intervened with almost 800,000 troops and at one point pushed U.S./South Korean forces to a line well south of Seoul.

Loss Aversion

The assertion that “losses loom larger than gains” was the most important claim of Prospect Theory (Kahneman and Tversky 1979: 279). It implied an abrupt change in the slope of the value function at the point that separates gains from losses, as seen in Figure 1. The difference in the slopes of the value function in the positive and negative domains is labeled loss aversion.

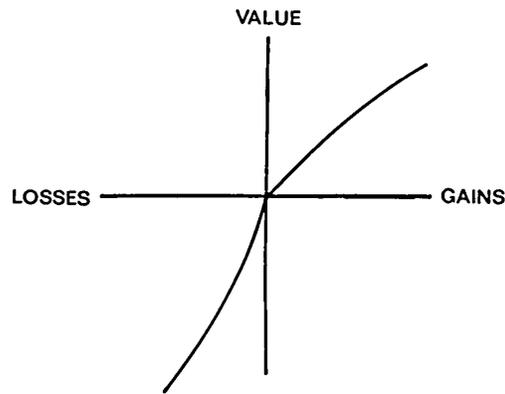


Figure 1: Hypothetical Value Function

The main evidence for loss aversion in prospect theory was the extreme reluctance of people to accept gambles with equal probabilities to win and lose. A majority of respondents will typically refuse to play a gamble in which they face equal probabilities to lose x or to win $2x$ (for example, 50% chance to lose \$100 and 50% chance to win \$200). Soon after the publication of prospect theory, Richard Thaler (1980) noted that loss aversion could explain the observation that he labeled the endowment effect: the fact that the monetary value that people assign to a good depends on whether or not it is already part of their endowment. This conceptual advance extended prospect theory from a theory of choice between gambles to a more general model of decision making, and provided the foundation of behavioral economics.

Some years later, Kahneman, Knetsch and Thaler (1991: 195) reported a demonstration of the endowment effect that has become standard. The critical experiment was conducted in a classroom. Half of the participants were given an attractive coffee mug. They were told the mug was theirs to keep, but were given an opportunity to sell it and were asked to record their cash equivalent (minimal selling price) for the mug. The other participants were told that they could choose between receiving a mug and receiving an amount of money, and were asked to indicate their cash equivalent for the mug. The options that faced the two groups were effectively identical: they could go home with a mug or with a sum of money. But the cash equivalents were quite different:

the participants who owned a mug valued it at \$7.12 while the participants who did not own a mug valued it at \$3.50.⁵

The point of this demonstration – and by extension, of the endowment effect – is that people do not put value on the states of “having a mug” or “not having a mug”. Depending on their endowment, they value the *changes* of “getting a mug” and “giving up a mug.” Furthermore, the psychological aversion to giving up a mug is more intense than the attraction of receiving one, by loss aversion. Recent work has suggested that one determinant of the endowment effect is emotional attachment; and that loss aversion occurs in proportion to the emotion experienced when an individual considers giving up a good (Ariely, Huber et al. 2005: 134; Novemsky and Kahneman 2005: 139-40).

Loss aversion also contributes to the explanation of an important phenomenon of choice known as the *status-quo bias*. Samuelson and Zeckhauser (1988: 5), who coined that term, illustrated the bias by the choices that Harvard University employees made when new options were added to health-care coverage. New employees chose a plan from the set of options, and continuing employees were free to switch to the new plan. Because new and continuing employees faced the same options, the distribution of their choices should have been the same, but it was not. The continuing employees were very likely to retain their current plan, regardless of the advantages of the new options.

It is easy to see that loss aversion contributes to the status quo bias. An individual who considers switching from her current plan A to an alternative plan B will naturally evaluate the features of the alternative plan from the reference point of her current plan. Plan B is viewed as a bundle of advantages and disadvantages relative to Plan A – and the disadvantages are given more weight than the advantages, a manifestation of loss aversion. If the individual owned Plan B, of course, she would think of the same choice in terms of the advantages and disadvantages of Plan A – and the disadvantages of that plan would loom larger than its advantages.

There are important exceptions to the endowment effect. In particular, the effect does not apply to “exchange goods” – goods that are specifically held to be traded, not to be used (Kahneman, Knetsch et al. 1990: 1344). Furthermore, highly experienced

⁵ For more recent demonstrations of this effect, see (Sen and Johnson 1997: 105; Carmon and Ariely 2000: 360).

traders exhibit little loss aversion even for goods they have purchased as personal possessions (List 2003) Individuals with much experience trading sports memorabilia were less likely to exhibit an endowment effect for assorted sports collectibles.

Loss aversion has an unfortunate effect on negotiations, because actors treat their own concessions as “losses,” but evaluate the concessions they receive as gains. This “concession aversion” can be illustrated by negotiations over mutual disarmament (Kahneman and Tversky 1995: 56) Because our own losses carry more psychological weight, offers that are objectively fair or equal will not seem so to either side: a country will feel the *loss* of a 10% reduction of its arsenal more than it will experience the *gain* of an equal reduction in the arsenal of its adversary. The asymmetric evaluation of gains and losses poses a serious obstacle to agreement (Levy 1997: 105).

Concession aversion is exacerbated by two additional factors: anticipated regret and the “sucker’s payoff.” Individuals faced with choice problems anticipate the regret they may experience if they do not achieve the best outcome. And the outcome of being betrayed by an opponent one has trusted –known as the “sucker’s payoff” –constitutes an especially severe form of regret. Actors are reluctant to expose themselves to this outcome by making concessions that could be exploited by their opponents. Anticipated regret favors both an aversion to concessions and the avoidance of risky cooperation.

The second implication of loss aversion for world politics concerns the relative speed at which people adapt to (or “integrate”) gains and losses. The endowment effect appears to be largely instantaneous. The mug that was received a minute ago is immediately absorbed as part of one’s endowment. In contrast, it appears that people are slower to adjust their reference point following losses (Kahneman, Knetsch et al. 1990: 1342).

Consider a scenario in which Country A has lost a valuable strategic territory to Country B during the course of a war. Because of the quick adjustment to its new endowment, B is likely to adopt the new territory into its “status quo portfolio”, with a corresponding increase in its willingness to expend blood and treasure to defend it.⁶

⁶ The exception to this is if the territory is taken and held as a bargaining chip (or “exchange good”), in which case one would not expect loss aversion to factor in the decision-making. Of course, part of the art of negotiation is to convince your opponent of your aversion to losing things that in fact you consider to be mere bargaining chips.

However, if actors do not adapt to losses as quickly as gains, Country A is likely to view the new situation as a loss relative to the ‘normal’ reference situation in which it held the territory. In this scenario, both countries will be operating in the domain of losses: B because of the instant endowment effect and Country A because of the slow adaptation to losses. As we see in the next section, actors are likely to take substantial risks in order to avoid accepting a loss -- when both sides face losses, the likelihood of conflict increases.

Risk Seeking in Losses

Formal models of rational choice do not prescribe a particular risk attitude. Rational decision makers can be as risk-averse or risk-seeking as they wish – the only requirement is that they must be internally consistent. Furthermore, models of rational choice do not explicitly specify the nature of the outcomes to which utilities are attached. However, the relevant outcomes and the appropriate risk attitude are both specified when the rational choice model (aka utility theory) is applied to a particular situation. In most applications, the outcomes are assumed to be final states – e.g., a complete description of the state of affairs when uncertainty is resolved. In the context of financial decisions, for example, rational agents are said to evaluate the outcomes of decisions as states of wealth, not as gains and losses. Furthermore, a moderately risk averse attitude is widely assumed to be characteristic of reasonable (rational) agents.

Prospect Theory questioned both these aspects of rational choice models. The idea that the carriers of value are gains and losses – not final states -- helps explain robust observations that are difficult to reconcile with rationality, including the endowment effect, the status quo bias and the susceptibility to many framing manipulations. A further departure from the rational model is the observation that risk-seeking preferences are prevalent when people face difficult choices, especially when one of the options involves accepting a sure loss.

For an example, consider the following problem:

Problem 1: Choose between:

- A. a sure loss of \$900
- B. a 90% chance to lose \$1000 and 10% chance to lose nothing

A large majority of people who have been asked this question choose the gamble over the sure loss – a risk-seeking preference. According to Prospect Theory, this preference is explained by two separate factors. The first is the shape of the value function (see Figure 1). Because of diminishing marginal disutility, the difference between a loss of \$900 and a loss of \$1,000 is relatively small. The second (and probably more important) cause of risk seeking in difficult situations is known as the *certainty effect*.

The certainty effect refers to the overweighting of outcomes that are certain, relative to outcomes that are merely probable (Tversky and Kahneman 1981: 455). Kahneman and Tversky illustrated this effect by the game of Russian roulette. Most people share the strong intuition that one should be willing to pay more to reduce the number of bullets from 1 to 0 than from 4 to 3 (Kahneman and Tversky 1979: 265). The reduction of the probability of dying is the same in the two cases, but removing the single bullet achieves the certainty of a good outcome, while reducing the number of bullets from 4 to 3 does not eliminate the uncertainty. Utility theory, in contrast, assumes that the rational response to probability is linear and does not distinguish between the two cases.

In the domain of gains, the certainty effect and the shape of the value function both favor risk aversion. For example, a large majority of people will prefer a sure gain of \$850 over a 90% chance to win \$1,000. The certainty effect contributes to this preference because the “decision weight” that is associated with a probability of .90 is significantly *less* than 90% of the decision weight of a sure thing (Kahneman and Tversky 1979: 280-81).⁷ The overweighting of the sure gain relative to a gain that is merely probable favors the risk-averse choice in this example. When an actor faces a choice between a sure loss and a possible loss, however, the certainty effect contributes to risk-seeking preferences by exactly the same logic. Certainty enhances the aversion to a sure loss just as it enhances the attractiveness of a sure gain.

⁷ The certainty effect applies only to moderate or high probabilities. Prospect Theory asserts that low probabilities are overweighted, contributing to the attractiveness of both gambles and insurance policies.

In policy debates, arguments that draw on the certainty effect often appear more compelling and persuasive than they should be. Experimental evidence dovetails with the common-sense observation that the invocation of certainties has a rhetorical power that mere probabilities lack.

Evidence of the special aversiveness of sure losses and the attendant increase in risk acceptance can be found in a variety of domains. For instance, Fiegenbaum and Thomas argue that aversion to certain losses accounts for the negative correlation between risk and return in investment decisions. Their large-n analysis of companies' investment portfolios indicated that most firms are risk-acceptant when they are suffering losses or are below targeted "aspiration levels" and risk-avoidant when they are above those levels (1988: 97).

In another domain, Rachlinski examined choices in the context of a hypothetical litigation case. He showed that decisions concerning whether to pursue litigation or settle vary with the domain of the individual. He also showed that respondents who were in the weaker position (low stated probability of prevailing in court) *and* were in the domain of losses were far more likely than their counterparts (weak position but in the domain of gains) to pursue a costly and risky litigation with a small chance of success (Rachlinski 1996).

The prototypical risk-seeking behavior occurs in desperate situations: faced with a choice between a significant loss and the prospect of a bigger disaster – but with a chance to escape – actors are prone to accept gambles that appear unreasonable to an objective observer. This analysis applies to the losing side in a conflict. Faced with a choice between surrendering or fighting on with a slim chance of escaping the worst, the leaders of the losing side will be tempted to persevere. If there is one thing that is worse than a sure loss it is a loss that is both sure and immediate.

The risk seeking behavior of losers is closely related to the phenomenon known as escalation of commitment, which is observed when a project that has consumed considerable resources appears to be failing. The rational prescription is to "ignore sunk

costs” and evaluate the costs of new efforts against the value of their expected outcome, but decision makers often persevere even with poor odds of success.⁸

An agency problem often compounds the psychological difficulties of giving up a lost cause. The leaders of the losing entity will not personally do well by giving up now, and the threat they face may not be worse if they go on fighting. But the slim chance of victory is often enough even for followers to support a lost cause, although they are likely to suffer for it. Indeed, the tendency to escalate commitment seems to be exacerbated when decisions are considered by groups, and when the actor who made the original decision is still in charge, circumstances that are often observed in political decision-making (Staw 1976: 27; Whyte 1993: 430; Schaubroeck and Davis 1994: 59).

Pseudo-Certainty

The pseudo-certainty effect refers to a decision-making bias in the response to multi-stage decisions. Specifically, it describes the tendency for individuals contemplating multiple-stage problems/scenarios to assign a high decision weight to an outcome that is certain if the second stage is reached, neglecting its contingent nature. Pseudo-certainty was illustrated by Kahneman and Tversky using the following two problems:⁹

Problem 2: Assume a two-stage game. In the first stage, there is a 75% chance to end the game with nothing and a 25% chance to move into the second stage. If you reach the second stage, your choice is between:

- A. a sure win of \$30
- B. 80% chance to win \$45

Your choice must be made before the game starts.

In this problem, a substantial majority (74%) chose **A**, which has the appeal of a seemingly certain gain. Now consider the next problem:

⁸ There are alternative explanations for the escalation of commitment, chief among them self-justification theory. However, a large amount of laboratory and field evidence has been amassed which point to prospect theory as at least a partial cause of the tendency to escalate commitment and take greater risks in doing so. For reviews, see (Brockner 1992; Bazerman 1998).

⁹ Results are from Tversky and Kahneman (1981: 453).

Problem 3: Which of the following options would you prefer?

C. 25% chance to win \$30

D. 20% chance to win \$45

A and C, and B and D, are respectively identical in terms of outcomes and probabilities. And yet, the single-stage version and the sequential version are treated differently.¹⁰ In Problem 3, the majority (58%) prefers D, because the difference between the probabilities appears smaller than in Problem 2. When contemplating multiple-stage problems in which the first stage is probabilistic and the second-stage contains a certain outcome, individuals tend to ignore the first-stage altogether. The preference for the option that yields a sure gain in the second stage is an example of the certainty effect that was discussed in the preceding section. In this case, however, the certainty that actors find attractive is only illusory: choice A is no more certain than C, it only seems to be so.

This effect has significant implications for international politics, in which decision-makers often encounter situations that involve multiple stages or interactions. In particular, pseudo-certainty is likely to exacerbate the reluctance to make strategically significant concessions. The dilemma that confronted Israel in the peace negotiations with Egypt in the 1970's illustrates the problem. At the conclusion the 1967 Six-Day War, Israel had gained control over the Sinai Peninsula, a strategically valuable region between the Mediterranean and Red Seas. If war broke out, the territory would be immensely valuable by providing strategic depth to prevent a large massing of Arab troops near Israel's borders (McPeak 1976: 429-30; Middletown 1991). Yigal Allon, writing when he was Deputy Prime Minister and Minister of Foreign Affairs for Israel, declared the Sinai to be "critical to Israeli defense (Allon 1976)."

Though Israel eventually ceded control over the Sinai in the Camp David Accords of 1978, Israeli hawks fought stubbornly against such a concession, and came close to carrying the day.¹¹ The intuitive appeal of their argument is easy to explain as an

¹⁰ For further experimental support for this phenomenon, see (Quattrone and Tversky 1988: 731; Kahneman and Tversky 1995: 52-3).

¹¹ For contemporaneous accounts of the Israeli domestic political arguments for and against the concession, see "Israeli Cabinet Vote Backs Compromises for Mideast Peace," *The Globe and Mail* (March 15, 1979): 1; H.D.S. Greenway, "Debate Over Concessions Concerns Many Israelis," *Washington Post* (December 9, 1977): A1; "And Now to Meet Those Deadlines," *The Economist* (March 17, 1979): 61.

instance of pseudo-certainty. The decision whether to give up the Sinai involved a two-stage dilemma. In the first stage, the question was whether another war would occur between Israel and Egypt. The second stage was the outcome of a war if one broke out. There was apparent certainty in the second stage: the Sinai was definitely an advantage if that stage was reached. In truth, however, the Sinai was only contingently valuable: holding it was important only *if* war broke out, but pseudo-certainty tends to mask such truths. The hypothesis of pseudo-certainty is that the probabilistic element in the first stage of the problem – the possibility that war may not break out – will be ignored or neglected, while the certain element of the second stage is over-emphasized. The actual probability that the war will be lost is the product of $P(\text{War Occuring}) \times P(\text{Losing War})$, and is influenced by the possible effect of an agreement in reducing the probability of war, as well as by the increased risk in case of war. However, pseudo-certainty leads actors to frame the problem in their minds as $P(\text{Losing War} \mid \text{War Occuring})$ rather than the more accurate $P(\text{Losing War} \mid \text{War Occuring}) \times P(\text{War Occuring})$.

By ignoring the first stage of such games, actors under-emphasize the effect of their own actions. The focus of Israeli hawks on the strategic advantage conferred by the Sinai *if war occurs* caused a correspondent neglect of the effect that giving up the Sinai might have on diminishing the probability of war in the first place. Decades of peace between Israel and Egypt strongly suggest that the hawks were wrong.

As this case illustrates, focusing attention on considerations that will become relevant only if war occurs leads decision-makers to plan for the worst-case scenario – often a recipe for self-fulfilling prophecies. Dovish arguments, such as “we should not assume that war will surely occur” or “doing this now increases the chances of war later” sound feeble. The rhetorical force of a “sure thing” is a powerful advantage for hawks, even if the certainty they invoke is illusory.

Reactive Devaluation

As already noted, the “bad behavior” of rivals is likely to be attributed to long-term hostile intentions, even when these inferences are not logically or factually supported. Similar distortions are observed in the context of negotiations, where actors must assess the offers presented by their adversaries. Here, the evidence suggests that

individuals assign different values to proposals, ideas and plans of action based on their authorship. This bias, known as “reactive devaluation,” is likely to be significant stumbling block in negotiations between adversaries.

In one recent experiment, Israeli Jews evaluated an actual Israeli-authored peace plan less favorably when it was attributed to the Palestinians than when it was attributed to their own government, and Pro-Israeli Americans saw a hypothetical peace proposal as biased in favor of Palestinians when authorship was attributed to Palestinians, but as “evenhanded” when they were told it was authored by Israelis. In fact, the phenomenon is visible even within groups. In the same experiment, Jewish “hawks” perceived a proposal attributed to the dovish Rabin government as bad for Israel, and good for the Palestinians, while Israeli Arabs believe the opposite (Maoz, Ward et al. 2002).

The effect of reactive devaluation in international politics is that leaders and their advisors are likely to significantly underestimate the value of proposals made by adversaries. We recognize, of course, that leaders have a duty to be wary of seemingly beneficial proposals or plans advanced by potentially untrustworthy sources. Our point is only that the evidence of reactive devaluation suggests that politicians are likely to be more wary and more suspicious than an unbiased observer would consider appropriate.

Conclusion

In our view, neither psychology nor decision science can provide a theory of interstate conflict (Kahneman and Tversky 2000: xi). It is simply unreasonable to expect a theory of choices between gambles with specified probabilities and monetary outcomes to predict or explain the decisions of national leaders as they wage war and negotiate peace (Kahneman and Tversky 2000: xi). It is similarly impossible to derive confident predictions about the future judgments and choices of national actors from notions such as the fundamental attribution error or loss aversion. There is too much else going on. By the same token, post hoc explanations of the judgments and decisions of national leaders in terms of the psychological mechanisms that we have illustrated are at best tentative and incomplete. A legitimate question can be asked – and often has been asked: if the psychology of judgment and decision making supports neither confident

predictions nor complete explanations of the actions of national leaders, how useful can it be?

A possible answer is that the concepts of psychology – including the seven biases that we discussed in this chapter – provide templates for patterns that can be recognized in complex situations. The training of histologists provides a useful analogy. Histologists are not born with an ability to differentiate different types of cancer cells, and cancer cells of a particular type do not look exactly alike and are hard to distinguish from other cells. The trained histologist has seen multiple examples of cells that share the same label, and has learned to identify them in microscope slides. The slides look different to the trained eye, they make more sense. We offer the observations discussed in this chapter as templates that may help make sense of past events and provide expectations about the future. Sense making and expectation are much weaker than explanations and predictions, but still potentially useful. The notion of the endowment effect, for example, prepares us to distinguish truly painful concessions that are treated as losses from other concessions in which bargaining chips are exchanged. And the Fundamental Attribution Error prepares us to find that participants in a conflict underestimate their role in provoking the opponent's hostility. We believe these intellectual tools may be useful, but repeat that “No warranty is implied... The scholars who use the tools to explain complex behaviors do so at their own risk” (Kahneman and Tversky 2000: xi).

The theme of this chapter was a chance observation, made while drawing a list of the biases of judgment and choice uncovered during the last few decades: it appeared that these biases were themselves biased in a hawkish direction. We started out by imagining a decision maker exposed to competing advisors, a hawk and a dove. If the decision maker is susceptible to the usual list of biases, she will tend to find the hawk's arguments more believable and the hawk's recommendations more attractive than they ought to be. We have found no deep reason for this observation, and we do not present it as a general and exceptionless claim. The most we can say for it is that it is intriguing.

SOURCES

- Allon, Y. (1976). "Israel: The Case for Defensible Borders." Foreign Affairs **55**: 38-53.
- Ariely, D., J. Huber, et al. (2005). "When Do Losses Loom Larger Than Gains?" Journal of Marketing Research **42**(2): 134-138.
- Bazerman, M. H. (1998). Judgment in Managerial Decision Making. New York, John Wiley & Sons.
- Bazerman, M. H. (2001). Why Negotiations Go Wrong. The Negotiation Sourcebook. I. G. Asherman, S.V. Asherman. MA, Human Resources Development Press.
- Brockner, J. (1992). "The Escalation of Commitment to a Failing Course of Action: Toward Theoretical Progress." Academy of Management Review **17**(1): 39-61.
- Camerer, C. and D. Lovallo (1999). "Overconfidence and Excess Entry: An Experimental Approach." American Economic Review **89**(1): 306-318.
- Campbell, J. D. (1986). "Similarity and Uniqueness: The Effects of Attribute Type, Relevance, and Individual Differences in Self-Esteem and Depression." Journal of Personality and Social Psychology **50**(2): 281-94.
- Carmon, Z. and D. Ariely (2000). "Focusing on the Forgone: How Value Can Appear So Different to Buyers and Sellers." Journal of Consumer Research **27**(3): 360-370.
- Cooper, A., C. Woo, et al. (1988). "Entrepreneurs' perceived chances for success." Journal of Business Venturing **3**(2): 97-108.
- Crary, W. G. (1966). "Reactions to Incongruent Self-Experiences." Journal of Consulting Psychology **30**(3): 246-52.
- Cross, K. P. (1977). "Not Can, But Will College Teaching Be Improved?" New Directions for Higher Education **17**.
- Darley, J. M. and D. Batson (1973). "From Jerusalem to Jericho": A study of Situational and Dispositional Variables in Helping Behavior." Journal of Personality and Social Psychology **27**: 100-108.
- Fiegenbaum, A. and H. Thomas (1988). "Attitudes Toward Risk and the Risk-Return Paradox: Prospect Theory Explanations." Academy of Management Journal **31**(1): 85-106.
- Friedland, N., G. Keinan, et al. (1992). "Controlling the Uncontrollable: Effects of Stress on Illusory Perceptions of Controllability." Journal of Personality and Social Psychology **63**(6): 923-31.
- Gilovich, T. and K. Savitsky (1999). "The Spotlight Effect and the Illusion of Transparency: Egocentric Assessments of How We Are Seen by Others." Current Directions in Psychological Science **8**(6): 165-168.
- Gilovich, T., K. Savitsky, et al. (1998). "The Illusion of Transparency: Biased Assessments of Others' Ability to Read One's Emotional States." Journal of Personality and Social Psychology **75**(2): 332-346.
- Glaser, M., T. Langer, et al. (2005). "Overconfidence of Professionals and Lay Men: Individual Differences Within and Between Tasks?"
- Heider, F. (1958). The Psychology of Interpersonal Relations, Lawrence Erlbaum Associates.
- Heine, S. J. and D. R. Lehman (1995). "Cultural Variation in Unrealistic Optimism: Does the West Feel More Vulnerable Than The East?" Journal of Personality and Social Psychology **68**(4): 595-607.

- Heradstveit, D. (1981). The Arab-Israeli Conflict: Psychological Obstacles to Peace. New York, Oxford Univ Press.
- Jervis, R. (1980). "The Impact of the Korean War on the Cold War." Journal of Conflict Resolution **24**(4): 563-92.
- Johnson, D. D. P. (2004). Overconfidence and War: The Havoc and Glory of Positive Illusions. Cambridge, MA, Harvard University Press.
- Johnson, D. D. P., R. McDermott, et al. (2006). "Overconfidence in Wargames: Experimental Evidence on Expectations, Aggression, Gender and Testosterone." Proceedings of the Royal Society of London **273**(1600): 2513-2520.
- Jones, E. E. and V. A. Harris (1967). "The Attribution of Attitudes." Journal of Experimental Social Psychology **3**(1): 1-24.
- Kahneman, D., J. L. Knetsch, et al. (1990). "Experimental Tests of the Endowment Effect and the Coase Theorem." Journal of Political Economy **98**(6): 1325.
- Kahneman, D., J. L. Knetsch, et al. (1991). "Anomalies: The Endowment Effect, Loss Aversion, and Status Quo Bias." Journal of Economic Perspectives **5**(1): 193-206.
- Kahneman, D. and J. Renshon (2007). "Why Hawks Win." Foreign Policy(158): 34-38.
- Kahneman, D. and A. Tversky (1979). "Prospect Theory: An Analysis of Decision under Risk." Econometrica **47**(2): 263-292.
- Kahneman, D. and A. Tversky (1995). Conflict Resolution: A Cognitive Perspective. Barriers to Conflict Resolution. K. Arrow, Mnookin, R.H., Ross, L., Tversky, A. and R. Wilson. New York, W.W. Norton: 44-62.
- Kahneman, D. and A. Tversky (2000). Choices, Values, and Frames. New York, Cambridge University Press.
- Kramer, R. M. (1994). "The Sinister Attribution Error: Paranoid Cognition and Collective Distrust in Organizations." Motivation and Emotion **18**(2): 199-230.
- Langer, E. J. (1975). "The Illusion of Control." Journal of Personality and Social Psychology **32**(2): 311-328.
- Levy, J. S. (1997). "Prospect Theory, Rational Choice, and International Relations." International Studies Quarterly **41**(1): 87-112.
- List, J. A. (2003). "Does Market Experience Eliminate Market Anomalies?" Quarterly Journal of Economics **118**(1): 41-71.
- Maoz, I., A. Ward, et al. (2002). "Reactive Devaluation of an " Israeli" vs." Palestinian" Peace Proposal." Journal of Conflict Resolution **46**(4): 515.
- Marks, G. (1984). "Thinking One's Abilities Are Unique and One's Opinions Are Common." Personality and Social Psychology Bulletin **10**(2): 203.
- Martin, D. J., L. Y. Abramson, et al. (1984). "Illusion of control for self and others in depressed and nondepressed college students." Journal of Personality and Social Psychology **46**(1): 125-136.
- McPeak, M. A. (1976). "Israel: Borders and Security." Foreign Affairs **54**: 426-443.
- Middletown, D. (1991). Sinai Pullout: Israel to Lose in Air Power. New York Times. New York: 9.
- Neale, M. A. and M. H. Bazerman (1985). "The Effects of Framing and Negotiator Overconfidence on Bargaining Behaviors and Outcomes." Academy of Management Journal **28**(1): 34-49.
- Novemsky, N. and D. Kahneman (2005). "How Do Intentions Affect Loss Aversion?" Journal of Marketing Research **42**(2): 139-140.

- Perloff, L. S. and B. K. Fetzer (1986). "Self-other Judgments and Perceived Vulnerability to Victimization." Journal of Personality and Social Psychology **50**(3): 502-510.
- Pietromonaco, P. R. and R. Nisbett (1982). "Swimming Upstream Against the Fundamental Attribution Error: Subjects' Weak Generalizations from the Darley and Batson Study." Social Behavior and Personality **10**(1): 1-4.
- Quattrone, G. A. (1982). "Overattribution and Unit Formation: When Behavior Engulfs the Field." Journal of Personality and Social Psychology **42**: 593-607.
- Quattrone, G. A. and A. Tversky (1988). "Contrasting Rational and Psychological Analyses of Political Choice." American Political Science Review **82**(3): 719-736.
- Rachlinski, J. (1996). "Gains, Losses and the Psychology of Litigation." Southern California Law Review **70**(113): 113-185.
- Regan, D. T., E. Straus, et al. (1974). "Liking and the Attribution Process." Journal of Experimental Social Psychology **10**(4): 385-97.
- Robertson, L. S. (1977). "Car Crashes: Perceived Vulnerability and Willingness to Pay for Crash Protection." Journal of Community Health **3**(2): 136-41.
- Ross, L. (1977). The intuitive psychologist and his shortcomings: Distortions in the attribution process. Advances in experimental social psychology. L. Berkowitz. New York, Academic Press. **10**: 173-220.
- Samuelson, W. and R. Zeckhauser (1988). "Status Quo Bias in Decision Making." Journal of Risk and Uncertainty **1**(1): 7-59.
- Schaubroeck, J. and E. Davis (1994). "Prospect Theory Predictions When Escalation Is Not The Only Chance to Recover Sunk Costs." Organizational Behavior and Human Decision Processes **57**(1): 59-82.
- Sen, S. and E. J. Johnson (1997). "Mere-Possession Effects without Possession in Consumer Choice." Journal of Consumer Research **24**(1): 105-117.
- Sherman, S. J. (1980). "On the Self-Erasing Nature of Errors of Prediction." Journal of Personality and Social Psychology **39**(2): 211-221.
- Silverman, I. (1964). "Self-Esteem and Differential Responsiveness to Success and Failure." Journal of Abnormal Psychology **69**: 115-9.
- Staw, B. M. (1976). "Knee-deep in the Big Muddy: A Study of Escalating Commitment to a Chosen Course of Action." Organizational Behavior and Human Performance **16**(1): 27-44.
- Svenson, O. (1981). "Are We All Less Risky and More Skillful Than Our Fellow Drivers?" Acta Psychologica **47**(2): 143-148.
- Taylor, S. E. and J. D. Brown (1988). "Illusion and Well-Being: A Social Psychological Perspective on Mental Health." Psychological Bulletin **103**(2): 193-210.
- Thaler, R. (1980). "Toward a Positive Theory of Consumer Choice." Journal of Economic Behavior and Organization **1**(1): 39-60.
- Tversky, A. and D. Kahneman (1974). "Judgment under Uncertainty: Heuristics and Biases." Science **185**(4157): 1124-1131.
- Tversky, A. and D. Kahneman (1981). "The framing of Decisions and the Psychology of Choice." Science **211**(4481): 453-458.
- Van Boven, L., T. Gilovich, et al. (2003). "The Illusion of Transparency in Negotiations." Negotiation Journal **19**(2): 117-131.

- Van Evera, S. (1999). Causes of War: Power and the Roots of Conflict. New York, Cornell University Press.
- Weinstein, N. D. (1980). "Unrealistic Optimism About Future Life Events." Journal of Personality and Social Psychology **39**(5): 806-820.
- Whyte, G. (1993). "Escalating Commitment in Individual and Group Decision Making: A prospect theory Approach." Organizational Behavior and Human Decision Processes **54**(3): 430-455.
- Wilson, T. D. and D. Kraft (1993). "Why Do I Love Thee?: Effects of Repeated Introspections about a Dating Relationship on Attitudes toward the Relationship." Personality and Social Psychology Bulletin **19**(4): 409.