COGNITIVE PSYCHOLOGY 24, 449–474 (1992)

Thinking through Uncertainty: Nonconsequential Reasoning and Choice

ELDAR SHAFIR

Princeton University

AND

AMOS TVERSKY

Stanford University

When thinking under uncertainty, people often do not consider appropriately each of the relevant branches of a decision tree, as required by consequentialism. As a result they sometimes violate Savage's sure-thing principle. In the Prisoner's Dilemma game, for example, many subjects compete when they know that the opponent has competed and when they know that the opponent has cooperated, but cooperate when they do not know the opponent's response. Newcomb's Problem and Wason's selection task are also interpreted as manifestations of nonconsequential decision making and reasoning. The causes and implications of such behavior, and the notion of quasi-magical thinking, are discussed.

Much of everyday thinking and decision making involves uncertainty about the objective state of the world and about our subjective moods and desires. We may be uncertain about the future state of the economy, our mood following an upcoming examination, or whether we will want to vacation in Hawaii during the holidays. Different states of the world, of course, often lead to different decisions. If we do well on the exam, we may feel that we deserve a break and want to go to Hawaii; if we do poorly, we may prefer to stay at home. When making decisions under uncertainty we need to consider the possible states of the world and their potential implications for our desires and actions. Uncertain situations may be thought of as disjunctions of possible states: either one state will

This research was supported by US Public Health Service Grant 1-R29-MH46885 from the National Institute of Mental Health to the first author, by Grant 89-0064 from the Air Force Office of Scientific Research to the second author, and by a grant from the Hewlett Foundation to the Stanford Center on Conflict and Negotiation. Correspondence and reprint requests should be addressed to Eldar Shafir, Department of Psychology, Princeton University, Princeton, NJ 08544.

0010-0285/92 $9.00

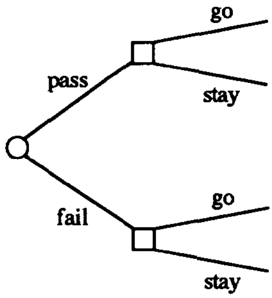

FIG. 1. A tree diagram for the Hawaiian vacation problem.

obtain, or another. A student who is uncertain about her performance on an exam, for instance, faces a disjunction of outcomes: passing the exam or failing the exam. In deciding whether or not to plan a vacation in Hawaii, the student needs to consider whether she would want to go to Hawaii if she were to pass the exam, and whether she would want to go if she were to fail, as diagrammed in Fig. 1. (As is customary, decision nodes are denoted by squares; chance nodes are denoted by circles.)

Most conceptions of decision making under uncertainty—both normative and descriptive—are *consequentialist* in the sense that decisions are determined by an assessment of the potential consequences and their perceived likelihood. According to this view, the student's decision to buy the Hawaiian vacation will depend on her subjective value of staying and going in the event that she passes the exam and in the event that she fails, and on her subjective probability of passing and failing.[1] Choices based on a consequentialist evaluation of anticipated outcomes are expected to satisfy a basic axiom of decision under uncertainty known as Savage's sure-thing principle (Savage, 1954, p. 21). The sure-thing principle (henceforth STP) says that if we prefer x to y given any possible state of the world, then we should prefer x to y even when the exact state of the world is not known. In the context of Fig. 1, it implies that if the student prefers going to staying both if she passes and if she fails the exam, then she should prefer going to staying even when the exam's outcome is not known. STP is an important implication of the consequentialist view. It captures a fundamental intuition of what it means for a decision to be determined by the anticipated consequences. It is a cornerstone of Expected Utility Theory, and it holds in other models which impose less stringent criteria of rationality.

[1] The notion of consequentialism appears in the philosophical and decision theoretic literature in a number of different senses. See, e.g., Hammond (1988), Levi (1991), and Bacharach and Hurley (1991) for technical discussion.

If, however, people do not always choose in a consequentialist manner, then STP may sometimes be violated. For example, we have shown elsewhere that many people who chose to purchase a vacation to Hawaii if they were to pass an exam *and* if they were to fail, decided to postpone buying the vacation in the disjunctive case, when the exam's outcome was not known (Tversky & Shafir, 1992). Having passed the exam, the vacation is presumably seen as a time of celebration following a successful semester; having failed the exam, the vacation becomes a consolation and time of recovery. Not knowing the outcome of the exam, we suggest, the decision maker lacks a clear reason for going and, as a result, may prefer to wait and learn the outcome before deciding to go, contrary to STP.

For another example of nonconsequential reasoning, imagine that you have agreed to bet on a toss of a coin in which you had equal chances to win $200 or lose $100. Suppose that the coin has been tossed, but that you do not know whether you have won or lost. Would you now want to play this gamble a second time? Alternatively, how would you feel about accepting the second gamble if you knew that you lost $100 on the first gamble? And finally, would you play the second gamble having discovered that you won $200 on the first gamble? We have shown that, contrary to STP, a majority of respondents accepted the second gamble both after having won as well as after having lost the first, but a majority rejected the second gamble when the outcome of the first was not known (Tversky & Shafir, 1992). This pattern—accept when win, accept when lose, but reject when do not know—was the single most frequent pattern of preferences exhibited by our subjects. We have suggested that people have a good reason for accepting the second gamble following a gain (namely, "I am up and no matter what happens I cannot lose"), and that they have a compelling albeit different reason for accepting the second gamble following a loss (namely, "I am down and this is my chance to get out of the red"). But when the outcome of the first gamble is unknown, people do not know whether they are ahead and cannot lose or whether they are behind and need to recover their losses. In this condition, we have argued, they may have no clear reason for accepting the additional gamble which, on its own, is not particularly attractive. We call the above pattern of preferences a *disjunction effect*. A disjunction effect occurs when people prefer x over y when they know that event A obtains, and they also prefer x over y when they know that event A does not obtain, but they prefer y over x when it is unknown whether or not A obtains. The disjunction effect amounts to a violation of STP, and hence of consequentialism.

In the present paper we explore nonconsequential behavior in several reasoning and decision making tasks. We suggest that various reasons and considerations are weighted differently in the presence of uncertainty than in its absence, giving rise to violations of STP. Our previous studies explored situations in which the reasons for a particular option (like going to Hawaii, or taking the gamble) were more compelling once the uncertainty was resolved than when the outcome was uncertain. The present studies focus on scenarios in which arguments that seem appealing while the outcome is uncertain lose much of their force once the uncertainty is resolved. It is proposed that the shift in perspective induced by the resolution of uncertainty may shed light on several puzzling manifestations of nonconsequential behavior. In the first part of the paper we explore one-shot Prisoner's Dilemmas, and a version of Newcomb's Problem played against a computer program. We then extend the analysis from decision making to reasoning. We suggest that nonconsequential reasoning plays an important role in Wason's selection task, and then describe a scenario in which the U.S. financial markets seem to exhibit nonconsequential behavior. Finally, we explore the implications of the present findings to the analysis of thinking in the face of uncertainty, and consider their relevance to the comparison between natural and artificial intelligence.

GAMES AND DECISIONS

Prisoner's Dilemma

The theory of games provides an analysis of the interaction among players who act according to specific rules. One particular two-person game which has received enormous attention is the Prisoner's Dilemma, or PD for short. (For extensive discussion, see Rapoport & Chammah, 1965; Rapoport, 1988). A typical PD is presented in Fig. 2. The cell entries

		OTHER cooperates	OTHER competes
YOU	cooperate	You: 75 Other: 75	You: 25 Other: 85
	compete	You: 85 Other: 25	You: 30 Other: 30

FIG. 2. A typical Prisoner's Dilemma. The cell entries indicate the number of points that you and the other player receive contingent on your choices.

indicate the payoffs (e.g., the number of points) received by each player. Thus, if both you and your opponent cooperate, each receives 75 points. On the other hand, if the other cooperates and you compete, you receive 85 points while the other receives 25, etc. What characterizes the PD is that regardless of the opponent's choice, each player fares better by competing than by cooperating; yet, if they both compete they do less well than if they had both cooperated. While many interesting strategies arise in the context of repeated games (see, e.g., Axelrod, 1984; Kreps & Wilson, 1982; Luce & Raiffa, 1957), the present discussion is confined to PD's that are played only once.

This is the simplest and sharpest form of a dilemma. Because the opponent is encountered only once, there is no opportunity for conveying strategic messages, inducing reciprocity, developing a reputation, or otherwise influencing the other player's choice of strategy. Because regardless of what the other does on this single encounter you will receive more points if you compete than if you cooperate, the dominant strategy is to compete. Nevertheless, some—presumably on ethical grounds—choose to cooperate. When Douglas Hofstadter (1983) presented a problem of this kind to a group of experts, roughly a third chose cooperation. Similar rates of cooperation were observed in a number of experimental studies (see, e.g., Rapoport, Guyer, & Gordon, 1976; Rapoport, 1988). The philosopher Dan Dennett captured the guiding ethical motivation when he remarked: "I'd rather be the person who bought the Brooklyn Bridge than the person who sold it. Similarly, I'd feel better spending $3 gained by cooperating than $10 gained by defecting." Evidently, some people are willing to forego some gains in order to make the cooperative, ethical decision.

Our previous discussion of nonconsequential reasoning suggests an alternative interpretation of the cooperation observed in one-shot PD games. Once the player knows that the other has chosen either to compete or to cooperate, it is clear that competition will be more advantageous to him than cooperation. But as long as the other has not made his decision, mutual cooperation looms as an attractive solution for both players. Although each player cannot affect the other's decision, he may be tempted to do his best (in this case, cooperate) to bring about the mutually desired state. This reasoning, of course, no longer applies once the outcome has occurred. Voting behavior is a case in point. We know that our individual vote is unlikely to affect the outcome of elections. Nevertheless, many of us who would not bother to vote once the outcome has been determined, are inclined to vote when the outcome of the elections is still pending. If this interpretation of cooperation in the PD game is correct, we expect a greater rate of cooperation in the disjunctive condition, when the other player's strategy is not known, than when the other player has chosen to compete or when the other has chosen to cooperate. This hypothesis is tested in the following study.

Method

Eighty Princeton undergraduates were presented with PD games displayed on a computer screen one at a time, in the format given in Fig. 2. On each trial, they chose whether to compete or cooperate by pressing the appropriate button. Subjects responded at their own pace, and once they chose their strategy, the screen cleared and the next game was presented. Each subject was presented with 40 games, of which only six were PD's. Other two-person games (with different payoff structures) were interspersed among the PD's in order to force subjects to consider each game anew, rather than adopt a "standard" strategy. Subjects were told that these games were being played with other students currently on the computer system, and that the outcomes would be determined by their choice and that of a new participant in each game. Their choices would not be made available to anyone playing with them. Thus, subjects were playing a series of one-shot games, each against a different opponent. In addition, subjects were told that they had been randomly assigned to a bonus group: this meant that, occasionally, they would be given information about the other player's already-chosen strategy before they had to choose their own. This information appeared on the screen along with the game, so that subjects could use it in making their decision. Subjects were to be paid according to the number of points that they accumulated throughout the session. They were paid $6.00 on average, and the entire session lasted approximately 40 min. The complete instructions appear in the Appendix.

We focus now on the six PD games that the subjects played. Each of these appeared three times throughout the session: once in the standard version where the other player's strategy was not known, a second time with the information that the other had competed, and a third time with the information that the other had cooperated. The standard version of each PD game appeared first, and the order of the other two was counterbalanced across subjects. The three versions of each game were separated by a number of other games in between. We refer to the three versions of each PD game as a PD "triad." The first 18 subjects were presented with four PD triads, and the remaining subjects played six PD triads each, yielding a total of 444 triads.

Results and Discussion

Subjects' responses to the PD triads are presented in Table 1. Table 1A summarizes subjects' chosen strategies, over all 444 games, when the other player competes and when the other player cooperates. Table 1B shows these same subjects' chosen strategy in the disjunctive case, when the other player's strategy is not known. When informed that the other has chosen to compete, the great majority of subjects reciprocate by competing. To cooperate would mean to turn the other cheek and forfeit points. Of the 444 games in which subjects were informed that the other had chosen to compete (Table 1A), only 3% resulted in cooperation. When informed that the other has chosen to cooperate, a larger percentage of subjects choose cooperation. This confirms the widespread senti-

TABLE 1
Prisoner's Dilemma

A. Other's strategy known[a]

		Other Player Competes		
		S Competes	*S* Cooperates	
Other Player Cooperates	*S* Competes	364	7	371 (84%)
	S Cooperates	66	7	73 (16%)
		430 (97%)	14 (3%)	444

B. Other's strategy not known[b]

		Other Player Competes	
		S Competes	*S* Cooperates
Other Player Cooperates	*S* Competes	113 Cooperate 251 Compete	3 Cooperate 4 Compete
	S Cooperates	43 Cooperate 23 Compete	5 Cooperate 2 Compete

[a] Joint distribution of subjects' (*S*) strategies when the other player competes and when the other player cooperates.
[b] Distribution of subjects' strategies when the other player's strategy is not known, broken down—as in A—according to subjects' choice of strategy when the other player competes and cooperates.

ment that there is an ethical inclination to reciprocate when the other cooperates. Of the 444 games in which subjects were told that the other player had cooperated, 16% resulted in cooperation. Now what should subjects do when the other's strategy is not known? Since 3% cooperate when the other competes and 16% cooperate when the other cooperates, we would expect an intermediate rate of cooperation when the other's strategy is not known. Instead, of the 444 games in which the other's strategy was unknown (Table 1B), a full 37% resulted in cooperation (the cooperation rates in the three versions are all significantly different, $p <$ .001 in all cases). The increased tendency to cooperate when uncertain about the other's chosen strategy cannot be attributed to a moral imperative of the type articulated by Dennett. Any account based on ethical considerations implies that the rate of cooperation should be highest when the other player is known to have cooperated, contrary to Table 1.

As expected, competition was the most popular strategy in all conditions. Consequently, the single most frequent choice pattern was to compete in all three versions. The next most frequent pattern, however, representing 25% of all response triads (113 out of 444 triads in Table 1B), was of the form: compete when the other competes, compete when the other cooperates, but cooperate when the other's strategy is not known. Sixty-five percent of the subjects exhibited such a disjunction effect on at least one of the six PD triads that they played. Of all triads yielding cooperation when the other player's strategy was unknown, 69% resulted in competition both when the other competed and when the other cooperated. This pattern is illustrated using the tree diagram of Fig. 3. The majority of subjects choose to compete at the upper branch (when the other cooperates) as well as at the lower branch (when the other competes). Contrary to STP, however, many cooperate when they do not know on which branch they are.

A behavioral pattern that violates a simple normative rule calls for both a positive analysis, which explains the specific factors that produce the observed response, and a negative analysis, which explains why the correct response is not made (Kahneman & Tversky, 1982). The conjunction fallacy (Tversky & Kahneman, 1983) is a case in point. The positive analysis of this phenomenon has invoked judgmental heuristics, such as availability and representativeness, whereas the negative analysis attributes conjunction errors to people's failure to detect the fact that one event is included in the other, or to their failure to appreciate the implication of this inclusion. Analogously, a negative analysis of the disjunction effect suggests that people do not evaluate appropriately all the relevant outcomes. This may occur because people sometimes fail to consider all the branches of the relevant decision tree, especially when the number of outcomes is large. Alternatively, people may consider all the relevant outcomes but, due to the presence of uncertainty, may not see their own

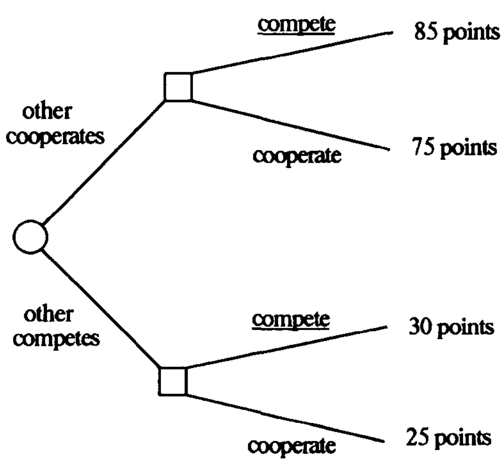

FIG. 3. A tree diagram illustrating the Prisoner's Dilemma presented in Figure 2. Decision nodes and chance nodes are denoted by squares and circles, respectively. Modal choices are underlined.

preferences very clearly. Consider the Hawaii scenario described earlier. A person who has just taken the exam but does not know the outcome may feel, without specifically considering the implications of success and failure, that this is not the time to choose to go to Hawaii. Alternatively, the person might contemplate the outcomes, but—uncertain about which outcome will occur—may feel unsure about her own preferences. For example, she may feel confident about wanting to go to Hawaii if she passes the exam, but unsure about whether she would want to go in case she failed. Only when she focuses exclusively on the possibility of failing the exam, does her preference for going to Hawaii become clear. A similar analysis applies to the present PD game. Not knowing the opponent's strategy, a player may realize that he wishes to compete if the other competes, but he may not be sure about his preference if the other were to cooperate. Having focused exclusively on the latter possibility, the player now sees more clearly that he wishes to compete in this case as well. The presence of uncertainty, we suggest, makes it difficult to focus sharply on any single branch; broadening the focus of attention results in a loss of acuity. The failure to appreciate the force of STP, therefore, is attributed to people's reluctance to consider all the outcomes, or to their reluctance to formulate a clear preference in the presence of uncertainty about those outcomes. This interpretation is consistent with the finding that, once people are made aware of their preferences given each possible outcome, STP is no longer violated (Tversky & Shafir, 1992).

Several factors may contribute to a positive analysis of the disjunction effect in the PD game. The game is characterized by the fact that an individually rational decision by each player results in an outcome that is not optimal collectively. Our subjects seem to exhibit a change of perspective that may be described as a shift from individual to collective rationality. Once the other's strategy is known, a player is "on her own." Only one column of the PD table is relevant (that which corresponds to the strategy chosen by the other), and the outcome of the game depends on her and her alone. The individually rational strategy, of course, is to compete. In the disjunctive condition, on the other hand, all four cells of the table are in play. The outcome of the game depends on the collective decision of both players, and the collectively optimal decision is for both to cooperate. Thus, the pattern of behavior observed in the PD may be explained, in part at least, by the greater tendency to adopt the collective perspective in the disjunctive version of the game.[2] Note, incidentally, that collective—albeit uncoordinated—action is quite viable. To the ex-

[2] A "collective action" interpretation of cooperative behavior in one-shot PD games is proposed by Hurley (1989, 1991). She interprets such behavior as "quite rational" since, according to her, it is motivated by "a concern to be part of, do one's part in, participate in

tent that our nonconsequentialist subjects play against one another, they stand to receive more points (for reaching the cooperate–cooperate cell) than would be awarded to consequentialist subjects (for their compete–compete cell). The potential benefits of cooperation in social dilemmas are discussed by Dawes and Orbell (1992).

A consequentialist subject who chooses to compete both when the other competes and when the other cooperates, should also compete when the other's decision—as is usually the case—is not known. Instead, uncertainty promotes a tendency to cooperate, which disappears once the other player's decision has been determined. It appears that many subjects did not appropriately evaluate each possible outcome and its implications. Rather, when the opponent's response was not known, many subjects preferred to cooperate, perhaps as a way of "inducing" cooperation from the other. Because subjects naturally assume that the other player—a fellow student—will approach the game in much the same way they do, whatever they decide to do, it seems, the other is likely to do the same. Along these lines, Messe and Sivacek (1979) have argued that people overestimate the likelihood that others will act the way they do in mixed-motive games. Such an attitude may lead subjects to cooperate in the hope of achieving joint cooperation and thereby obtaining the largest mutual benefit, rather than compete and risk joint competition. If they were able to coordinate a binding agreement, subjects would certainly agree on mutual cooperation. Being unable to secure a binding agreement in the PD game, subjects are nevertheless tempted to act in accord with the agreement that both players would have endorsed. Although they cannot actually affect the other's decision, subjects choose to "do their share" to bring about the mutually preferred state. This interpretation is consistent with the finding of Quattrone and Tversky (1984) that people often select actions that are diagnostic of favorable outcomes even though they do not cause those outcomes. A discussion of the relation between causal and diagnostic reasoning is resumed in the next section.

We have interpreted violations of STP in the above PD games as an indication that people do not evaluate the outcomes in a consequentialist manner. We now consider two alternative interpretations of the above findings. First, subjects might have cooperated in the disjunctive version of the game because they were afraid that their choices will be relayed to the other player before she had made her decision. This concern, of course, would not arise once the other's decision has already been made.

. . . a valuable form of collective agency" (1989, p. 150). As with the ethical arguments mentioned earlier, however, this interpretation entails that subjects should certainly be inclined to cooperate when the other has cooperated, contrary to the present findings.

Recall that subjects were specifically told that their choices would not be communicated to anyone playing with them. Nevertheless, they could have been suspicious. Post-experimental interviews, however, revealed that while a few subjects were suspicious about the actual, simultaneous presence of other players on the system, none were concerned that their choices would be surreptitiously divulged. It is unlikely that suspicion could account for subjects' strategies in the present experiment.

Second, it could be argued that the present results can be explained by the hypothesis that the tendency to compete increases as the experiment progresses. In order to observe subjects' untainted strategies in the standard prisoner's dilemmas, we presented these games before the known-outcome games. Hence, an increase in the tendency to compete as the experiment progressed could contribute to the observed pattern because the disjunctive problems—where cooperation was highest—generally occurred earlier in the experiment. However, no such temporal change was observed. The rate of cooperation in the first disjunction encountered (the sixth game played), the fourth disjunction (game number 15), and the last disjunction (game number 19), were 33%, 30%, and 40%, respectively. Similarly, the rate of cooperation when subjects were told that the other had cooperated averaged 13% for the first three occurrences and 21% for the last three. In effect, cooperation tended to increase as the experiment progressed, which would diminish the observed frequency of disjunction effects.

Recall that, like Hofstadter's experts, nearly 40% of our subjects chose to cooperate in a single-shot Prisoner's Dilemma. Once they discovered the other's strategy, however, nearly 70% of these cooperators chose to compete both when the other competed and when the other cooperated. These players followed a variant of Kant's Categorical Imperative: act in the way you wish others to act. They felt less compelled, however, to act in ways others have already acted. This pattern suggests that some of the cooperation observed in one-shot PD games may stem not from a moral imperative of the kind described by Dennett but, rather, from a combination of wishful thinking and nonconsequential evaluation. A similar analysis may apply to a related decision problem to which we turn next.

Newcomb's Problem

First published by Nozick (1969), Newcomb's Problem has since generated a lively philosophical debate that touches upon the nature of rational decision. The standard version of the Problem proceeds roughly as follows.

Suppose that you have two options: to take the contents of a closed box in front of you, or to take the contents of the closed box *plus* another, open box that you can see contains $1000 in cash. The closed box contains either one million dollars ($M) or nothing, depending on whether a certain being with miraculous powers of foresight, called the Predictor, has or has not placed $M there prior to the time at which you are to make your decision. You know that the Predictor will have placed the $M in the closed box if he has predicted that you will choose the closed box alone; he will have left the closed box empty if he has predicted that you will choose both boxes. You also know that almost everyone who has chosen both boxes found the closed box empty and received just $1000, while almost everyone who has chosen just the closed box has found $M in it. What is your choice?

A number of authors (e.g., Brams, 1975; Lewis, 1979; Sobel, 1991) have commented on the logical affinity of Newcomb's Problem with the Prisoner's Dilemma. In both cases, the outcome depends on the choice that you make and on that made by another being—the other player in the Prisoner's Dilemma, and the Predictor in Newcomb's Problem. In both cases, one option (competing or taking both boxes) dominates the other, yet the other option (cooperating or taking just one box) seems preferable if the being—the Predictor or the other player—knows what you will do, or will act like you.

The conflicting intuitions generated by Newcomb's problem proceed roughly as follows (see Nozick, 1969, for a more complete treatment).

Argument 1 (for one box). If I choose both boxes, the Predictor, almost certainly, will have predicted this and will not have put the $M in the closed box, and so I will get only $1000. If I take only the closed box, the Predictor, almost certainly, will have predicted this and will have put the $M in that box, and so I will get $M. Thus, if I take both boxes I, almost certainly, will get $1000, and if I take just the closed box I, almost certainly, will get $M. Therefore, I should choose just the closed box.

Argument 2 (for two boxes). The Predictor has already made his prediction and has already either put the $M in the closed box or has not. If the Predictor has already put the $M in the closed box, and I take both boxes I get $M + $1000, whereas if I take only the closed box, I get only $M. If the Predictor has not put the $M in the closed box, and I take both boxes I get $1000, whereas if I take only the closed box I get no money. Therefore, whether the $M is there or not, I get $1000 more by taking both boxes rather than only the closed box. So I should take both boxes.

When Martin Gardner (1973; 1974) published Newcomb's Problem in *Scientific American* and invited readers to send in their responses, roughly 70% of the readers who indicated a preference found Argument 1 more compelling and chose to take just the closed box, while 30% were driven by Argument 2 to take both boxes. Argument 2 relies on consequential reasoning reminiscent of STP, namely, whatever the state of the boxes, I will do better choosing both boxes than one only. Argument 1, on the other hand, is more problematic. While couched in terms of expected utility, it seems to suppose that what the Predictor will have predicted—although he has done so already—depends somehow on what I decide to do now. Excluding trickery, there are two interpretations of the Predictor's unusual powers. According to the first interpretation, the Predictor is simply an excellent judge of human character. Using some database (including, e.g., gender, background, and appearance), a predictor might be able to predict the decision maker's response with remarkable success. If this interpretation is correct, then you have no reason to take just one box: however insightful the Predictor's forecast, you will do better if you take both boxes rather than one box only. The second interpretation is that the Predictor has truly supernatural powers of insight. If you are unwilling to dismiss this possibility, then you may be justified in deferring to the mysterious powers of the Predictor and taking just one box (cf., Bar-Hillel & Margalit, 1972). This puzzle has captured the imagination of many people. An interesting collection of articles on Newcomb's Problem and its relation to the Prisoner's Dilemma is provided in Campbell & Sowden (1985).

Like Gardner's readers, many people presented with Newcomb's problem opt for one box only, contrary to the consequential logic of Argument 2. The choice of the single box may result from a belief in the Predictor's supernatural abilities. Alternatively, it may reflect a nonconsequential evaluation of the options in question. To distinguish between these interpretations, we created a credible version of Newcomb's problem that involves no supernatural elements. The role of the predictor is played by a fictitious computer program, whose predictions of subjects' choices are based on a previously established database. The experiment proceeded as follows. Upon completing the PD study described in the previous section, subjects ($N = 40$) were presented with the following scenario, displayed on the computer screen:

> You now have one more chance to collect additional points. A program developed recently at MIT was applied during this entire session to analyze the pattern of your preferences. Based on that analysis, the program has predicted your preference in this final problem.
>
> | 20 points | ? |
> |---|---|
> | Box A | Box B |
>
> Consider the two boxes above. Box A contains 20 points for sure. Box B may or may not contain 250 points. Your options are to:
>
> (1) Choose both boxes (and collect the points that are in both).
>
> (2) Choose Box B only (and collect only the points that are in Box B).
>
> If the program predicted, based on observation of your previous preferences, that you will take both boxes, then it left Box B empty. On the other hand, if it predicted

> that you will take only Box B, then it put 250 points in that box. (So far, the program has been remarkably successful: 92% of the participants who chose only Box B found 250 points in it, as opposed to 17% of those who chose both boxes.)
> To insure that the program does not alter its guess after you have indicated your preference, please indicate to the person in charge whether you prefer both boxes or Box B only. After you indicate your preference, press any key to discover the allocation of points.

This scenario provides a believable version of Newcomb's Problem. While the computer program is quite successful, it is by no means infallible.[3] Also, any suspicion of backward causality has been removed: assuming the experimenter does not cheat in some sophisticated fashion (and our post-experimental interviews indicated that no subject thought he would), it is clear that the program's prediction has been made, and can be observed at any point, without further feedback about the subject's decision. This problem has a clear "common cause" structure (see Eells, 1982): the subject's strategic tendencies in games of this kind, as observed in the preceding PD games, are supposedly predictive of both his preferred strategy in the next game and of the prediction made by the program. While the choice of a single box is diagnostic of the presence of 250 points in it, there can be no relevant causal influence between the two events. Under these conditions, there seems to be no defensible rationale for taking just one box. As Nozick (1969) points out, "if the actions or decisions . . . do not affect, help bring about, influence, and so on *which* state obtains, then whatever the conditional probabilities . . . , one should perform the dominant action," namely, take both boxes. In this situation, it would appear, people should choose both boxes since both boxes are better than one no matter what.

The results were as follows: 35% (14 of the 40 subjects) chose both boxes, while 65% preferred to take Box B only. The present scenario, which removed all supernatural elements from the original formulation of Newcomb's problem, yielded roughly the same proportions of choices for one and for both boxes as those obtained by Gardner from the readers of *Scientific American*. What can be said about the majority of subjects who preferred to take just one box? Had they known for certain that the 250 points were in that box (and could see that 20 were in the other), they surely would have taken both rather than just one. And certainly, if they knew that the 250 points were not in that box, they would have taken both rather than just the one that is empty. These subjects, in other words,

[3] In retrospect, the remarkably simple program, "Put 250 points in Box B if the subject has produced at least two disjunction effects in the PD experiment; otherwise, leave Box B empty," would have rewarded 70% of the one-boxers and only 29% of the two-boxers with 250 points in Box B. More sophisticated rules cold probably come closer to the alleged performance of the MIT program.

would have taken both boxes had they known that Box B is full, and they also would have taken both boxes had they known that Box B is empty. Consequentialist subjects should then choose both boxes even when it is not known whether Box B is full or empty. The majority, however, chose Box B alone when its contents were not known. Note that the hypothesis, discussed earlier, that attributes the disjunction effect to subjects' failure to predict their own preferences, cannot account for the present finding. No subject would have had any difficulty predicting his preference for more rather than fewer points, had he considered the possible states of the unknown box. Evidently, many subjects do not consider separately the consequences of the program's predictions, and as a result succumb to the temptation to choose the single box, which happens to be correlated with the higher prize.

Quasi-magical Thinking

Magical thinking refers to the erroneous belief that one can influence an outcome (e.g., the role of a die) by some symbolic or other indirect act (e.g., imagining a particular number) even though the act has no causal link to the outcome. We introduce the term quasi-magical thinking to describe cases in which people act as if they erroneously believe that their action influences the outcome, even though they do not really hold that belief. As in the Prisoner's Dilemma, the pattern of preferences observed in Newcomb's problem, may be described as quasi-magical thinking. When the program's prediction is known, the outcome depends entirely on the subject's decision and the obvious choice is to take both boxes. But as long as the program's prediction is not known and the eventual outcome depends on the behavior of both subject and program, there is a temptation to act as if one's decision could affect the program's prediction. As Gibbard and Harper (1978) suggest in an attempt to explain people's choice of a single box, "a person may . . . want to bring about an indication of a desired state of the world, even if it is known that the act that brings about the indication in no way brings about the desired state itself." Most people, of course, do not actually believe that they are able to alter the decision made by the program or the other player. Nevertheless, they feel compelled to "do their bit" in order to bring about the desired outcome. Another demonstration of such quasi-magical thinking was provided by Quattrone and Tversky (1984), whose subjects in effect "cheated" on a medical exam by selecting actions (e.g., holding their hand in very cold water for an extended period of time) that they believed were diagnostic of favorable outcomes (e.g., a strong heart) even though they must have known that their actions could not possibly produce the desired outcomes.

Quasi-magical thinking, we believe, underlies several phenomena related to self-deception and the illusion of control. Quattrone and Tversky (1984), for example, noted that Calvinists act as if their behavior will determine whether they will go to heaven or to hell, despite their belief in divine pre-determination, which entails that their fate has been determined prior to their birth. Several authors, notably Langer (1975), showed that people often behave as if they can exert control over chance events and, as a result, exhibit different attitudes and place larger bets when betting before rather than after a coin has been tossed (Rothbart & Snyder, 1970; Strickland, Lewicki, & Katz, 1966).[4] Most people, however, do not really believe that they can control the toss of a coin, nor that the choice of a single box in the Newcomb experiment can influence the program's already-made prediction. In these and other cases, people probably know that they cannot affect the outcome, but they act as if they could. It is told of Niels Bohr that, when asked by a journalist about a horseshoe (purported to bring good luck) hanging over his door, he explained that he of course does not believe in such nonsense, but heard that it helped even if one did not believe.

It is exceedingly difficult, of course, to ascertain what people really believe. The preceding discussion suggests that we cannot always infer belief from action. People may behave as if they could influence uncontrollable events even though they do not actually believe in being able to do so. For example, dice players who throw softly for low numbers and hard for high numbers (Henslin, 1967) may not necessarily believe that the nature of the throw influences the outcome. People who exhibit superstitious behaviors, such as wearing a good luck charm or avoiding crossing a black cat's path, may not actually believe that their actions can affect the future. There is a sense in which quasi-magical thinking appears more rational than magical thinking because it does not commit one to patently absurd beliefs. On the other hand, quasi-magical thinking appears even more puzzling because it undermines the link between belief and action. Whereas magical thinking involves indefensible beliefs, quasi-magical thinking yields inexplicable actions. The presence of uncertainty, we suggest, is a major contributor to quasi-magical thinking; few people act as if they can undo an already certain event by performing an action that is diagnostic of an alternative event. In this vein, subjects in Quattrone and Tversky's (1984) experiment would have been less willing to keep their hands in painfully cold water if they knew that they had strong

[4] One may distinguish between uncertainty about the outcome of a future event and uncertainty about the outcome of an event that has already occurred. While the present study does not systematically differentiate between the two, Greene and Yolles (1990) present data which give reason to expect more nonconsequential reasoning in the former than the latter.

or weak hearts than when their "diagnosis" was uncertain. And Calvinists would perhaps do fewer good deeds if they knew that they had already been assigned to heaven, or to hell, than while their fate remains a mystery.

GENERAL DISCUSSION

As demonstrated in the previous section, people often fail to consider the possible outcomes and consequences of uncertain events. The difficulties of thinking through uncertainty manifest themselves in a variety of situations: they encompass reasoning as well as decision making tasks, and they are observed both inside and outside the laboratory. In the present section, we extend the analysis of nonconsequential evaluation to deductive reasoning and economic forecast.

Wason's Selection Task

One of the best known tasks in research on human reasoning is the selection task, devised by Wason (1966). In a typical version of the task, subjects are presented with four cards, each of which has a letter on one side and a number on the other. Only one side of each card is displayed. For example:

E	D	4	7

Subjects' task is to indicate which cards must be turned over to test the rule: "If there is a vowel on one side of the card, then there is an even number on the other side of the card." The simplicity of the problem is deceptive—the great majority of subjects fail to solve it.[5] Most select only the *E*, or the *E* and the *4* cards, whereas the correct choices are the *E* and the *7* cards. The difficulty of the selection task is puzzling, especially because people generally have no trouble evaluating the relevance of the items that may be hidden on the other side of each card. Wason and Johnson-Laird (1970; see also Wason, 1969) have commented on the discrepancy between subjects' ability to evaluate the relevance of potential outcomes (i.e., to understand the truth conditions of the rule), and their inappropriate selection of the relevant cards. Subjects, for example, understand that neither a vowel nor a consonant on the other side of the *4* card contributes to the possible falsification of the rule, but they choose to turn the *4* card when its other side is not known. Similarly, subjects

[5] The success rate of initial choices in dozens of studies employing the basic form of the selection task (with "abstract" materials) typically ranges between 0 and somewhat over 20%. See Evans (1989) and Gilhooly (1988) for reviews.

understand that a consonant on the other side of the 7 card would not falsify the rule and that a vowel *would* falsify it, yet they neglect to turn the *7* card. The above pattern, which resembles a disjunction effect, arises when subjects who are easily able to evaluate the relevance of a specific outcome, fail to apply this knowledge when facing a disjunction of outcomes. As Evans (1984, p. 458) notes, "this strongly confirms the view that card selections are not based upon any analysis of the consequences of turning the cards." Like the people who postpone the trip to Hawaii when the exam's outcome is not known, and those who cooperate in the disjunctive version of the Prisoner's Dilemma, subjects performing the selection task fail to consider the consequences of each of the events. Instead of considering the consequences of each particular kind of symbol on the other side of the card, they appear to remain behind a veil of uncertainty when the card's other side is not known.

Numerous studies have explored the elusive thought process that underlies subjects' performance on the selection task. Indeed, a complex pattern of content effects has emerged from a number of variations on the original task (see, e.g., Johnson-Laird, Legrenzi, & Legrenzi, 1972; Griggs & Cox, 1982; Wason, 1983; Evans, 1989, for a review; although see also Manktelow & Evans, 1979, for conflicting reports). To explain these findings, researchers have suggested verification biases (Johnson-Laird & Wason, 1970), matching biases (Evans & Lynch, 1973; Evans, 1984), memories of domain-specific experiences (Griggs & Cox, 1982; Manktelow & Evans, 1979), pragmatic reasoning schemas (Cheng & Holyoak, 1985, 1989), and an innate propensity to look out for cheaters (Cosmides, 1989). What these explanations have in common is an account of performance on the selection task that fails to refer to formal reasoning. Instead, people are assumed to focus on items that have been explicitly mentioned, to apply pre-stored knowledge structures, or to remember relevant past experiences. "The inferential processes that occur in these cases," concludes Wason (1983, p. 69), "are not . . . instances of 'logical' reasoning." Thus, people find it relatively easy to reason logically about each isolated outcome, but a disjunction of outcomes leads them to suspend logical reasoning. This is reminiscent of the eight-year-olds studied by Osherson and Markman (1974–75) who when asked about a concealed, single-color poker chip, whether it is true that "Either the chip in my hand is yellow or it is not yellow?", responded "I don't know" because they could not see it. While most adults find the poker chip disjunction trivial, subtler disjunctions can lead to a temporary suspension of judgment.

The Disjunction Effect in Financial Markets

One result of nonconsequential decision making is that people will sometimes seek information that has no impact on their decision. In the

Hawaii problem described earlier, for example, subjects were willing in effect to pay for information that was not going to change their choice but—as we have interpreted it—was merely going to clarify their reasons for choosing. In a variation on the earlier PD experiment, we presented a new group of subjects with the same PD games, but this time, instead of being told the other's decision, subjects were offered, for a very small fee, the opportunity to learn the other's decision before making their own choice. The great majority of subjects chose to compete regardless of whether the opponent had decided to compete or to cooperate, but on 81% of the trials subjects first chose to pay to discover the opponent's decision. Although this behavior can be attributed to curiosity, we conjecture that people's willingness to pay for the information would have diminished had they realized that it would not affect their decision. Searching for information that has no impact on decision may be quite frequent in situations of uncertainty. For example, we may call to find out whether a beach hotel has a pool before making a reservation, despite the fact we will end up going whether it has a pool or not. One intriguing case of a nonconsequential evaluation of information is provided by the following account regarding the U.S. financial markets.

In the weeks preceding the 1988 U.S. Presidential election, the financial markets in the U.S. remained relatively inactive and stable "because of caution before the Presidential election" (*The New York Times*, Nov. 5, 1988). "Investors were reluctant to make major moves early in a week full of economic uncertainty and 7 days away from the Presidential election" (*The Wall Street Journal*, Nov. 2). The market, reported the *Wall Street Journal*, was "killing time." "There is literally nothing going on, and there probably won't be at least until Wednesday" observed the head of a trading desk at Shearson Lehman Hutton, referring to the day following the election (*WSJ*, Nov. 8). "Once the uncertainty of the election is removed, investors could begin to develop a better feeling about the outlook for the economy, inflation and interest rates," remarked the president of an investment firm (*NYT*, Nov. 2). "The outcome of the election had cast a decided cloud over the market in recent days. Its true direction is likely to surface rapidly in coming days," explained a portfolio strategist (*NYT*, Nov. 9). And, in fact, immediately following the election, a clear direction surfaced. The dollar plunged sharply to its lowest level in 10 months, and stock and bond prices declined. During the week following Bush's victory the DOW Jones industrial average fell a total of 78 points.[6] "The post-election reality is setting in," explained the co-chairman of an investment committee at Goldman, Sachs & Co. (*WSJ*,

[6] Some believed that the central banks were actually involved in preventing the dollar from plummeting just before the U.S. presidential election (see, e.g., *WSJ*, 11/2–4).

Nov. 21). The dollar's decline, explained the analysts, "reflected continued worry about the U.S. trade and budget deficits," "the excitement of the election is over, the honeymoon is over, and economic reality has set back in" (*WSJ*, Nov. 10). The financial markets, said the front page of the *NYT* on November 12, "had generally favored the election of Mr. Bush and had expected his victory, but in the three days since the election they have registered their concern about where he goes from here." Of course, the financial markets would likely have registered at least as much concern had Mr. Dukakis been elected. Most traders agree, wrote the *WSJ* on election day, "the stock market would drop significantly if Democratic candidate Michael Dukakis stages a come-from-behind victory." In fact, the market reacted to Bush's victory just as it would have reacted to Dukakis's. "When I walked in and looked at the screen," explained one trader after the election, "I thought Dukakis had won" (*NYT*, Nov. 10).

After long days of inactivity preceding the election, the market declined immediately following Bush's victory, and certainly would have declined at least as much had Dukakis been the victor. Of course, a thorough analysis of the financial markets' behavior is likely to reveal numerous complications. There is, for example, the possibility that an unexpected margin of victory, a surprising last-minute outcome, could have contributed to the paradoxical effect. As it happens, however, "newspapers and television networks came about as close as polling specialists believe is possible to forecasting the results of [the] election" (*NYT*, Nov. 10, 1988). In the week preceding the election, while some thought a Dukakis "surge" still possible, polls conducted by Gallup, ABC News/*Washington Post*, NBC/*Wall Street Journal*, and *New York Times*/CBS News predicted a Bush victory by an average margin of 9 percentage points, 1 point off the eventual 8-point margin. Similarly, the Democrats were expected to retain control of both the Congress (where they were predicted to pick up one or two seats) and the House of Representatives, which is precisely what occurred. The election results do not appear to have been a surprise. At least on the surface, this incident has all the makings of a disjunction effect: the markets were going to decline if Bush was elected, they were going to decline if Dukakis was elected, but they resisted any change until they knew which of the two had been elected. Being at the node of such a momentous disjunction seems to have stopped Wall Street from seriously addressing the consequences of the election. While either elected official would have led the financial markets to "register their concern about where he goes from here," the interim situation of uncertainty highlighted the need for "caution before the election." After all, how can we worry about "where he goes from here" before we know who is doing the going?

CONCLUDING COMMENTS

Patterns of decision and reasoning that violate STP were observed in simple contexts involving uncertainty. These patterns, we suggest, reflect a failure on the part of people to detect and apply this principle rather than a lack of appreciation for its normative appeal. When we first asked subjects to indicate their preferred course of action under each outcome and only then to make a decision in the disjunctive condition, the majority of subjects who opted for the same option under every outcome chose that option also when the precise outcome was not known (Tversky & Shafir, 1992). The frequency of disjunction effects, in other words, substantially diminishes when the logic of STP is made salient. Like other normative principles of decision making, STP is generally satisfied when its application is transparent, but is sometimes violated when it is not (Tversky & Kahneman, 1986).

A number of factors may contribute to the reluctance to think consequentially. Thinking through an event tree requires people to assume momentarily as true something that may in fact be false. People may be reluctant to make this assumption, especially when another plausible alternative (another branch of the tree) is readily available. It is apparently difficult to devote full attention to each of several branches of an event tree (see also Slovic & Fischhoff, 1977). As a result, people may be reluctant to entertain the various hypothetical branches. Furthermore, they may lack the motivation to traverse the tree simply because they presume, as is often the case, that the problem will not be resolved by separately evaluating the branches. We usually tend to formulate problems in ways that have sifted through the irrelevant disjunctions: those that are left are normally assumed to involve genuine conflict.

The disjunctive scenarios investigated in this paper were relatively simple, involving just two possible outcomes. Disjunctions of multiple outcomes are more difficult to think through and, as a result, are more likely to give rise to nonconsequential reasoning. This is particularly true for economic, social, or political decisions, where the gravity and complexity of situations may conceal the fact that all possible outcomes are eventually—perhaps for different reasons—likely to lead to a similar decision. Critics of U.S. nuclear first-strike strategies, for example, have maintained that while every plausible array of Russian missiles argues against the viability of an American first-strike, American strategists have insisted on retaining that option while the exact array of Russian arsenals is not known. Of course, the strategies involved in such scenarios are exceedingly complex, but it is conceivable that a first-strike option appears attractive partly *because* the adversary's precise arsenals are not known.

Shortcomings in reasoning have typically been attributed to quantitative limitations of human beings as processors of information. "Hard problems" are typically characterized by reference to the "amount of knowledge required," the "memory load," or the "size of the search space" (cf. Kotovsky, Hayes, & Simon, 1985; Kotovsky & Simon, 1990). These limitations play a critical role in many problems. They explain why we cannot remember all the cards that have previously come up in a poker game, or why we are severely limited in the number of steps that we can plan ahead in a game of chess. Such limitations, however, are not sufficient to account for all that is difficult about thinking. In contrast to many complicated tasks that people perform with relative ease, the problems investigated in this paper are computationally very simple, involving a single disjunction of two well-defined states. The present studies highlight the discrepancy between logical complexity on the one hand and psychological difficulty on the other. In contrast with the "frame problem" (McCarthy & Hayes, 1969; Hayes, 1973), for example, which is trivial for people but exceedingly difficult for AI, the task of thinking through disjunctions is trivial for AI (which routinely implements "tree search" and "path finding" algorithms) but very difficult for people. The failure to reason consequentially may constitute a fundamental difference between natural and artificial intelligence.

APPENDIX

Instructions Given to Subjects in Prisoner's Dilemma Game

Welcome to the Intercollegiate Computer Game. The game will be conducted on an IBM PC. In this game you will be presented with situations involving you and one other player. Each situation will require that you make a strategic decision: to cooperate or to compete with the other player. The other player will have to make a similar decision.

Each situation will present a payoff-matrix that will determine how many points each of you earns depending on whether you compete or cooperate. One such matrix looks like the following.

	Other Cooperates	Other Competes
You Cooperate	You: 20 Other: 20	You: 5 Other: 25
You Compete	You: 25 Other: 5	You: 10 Other: 10

According to this matrix, if you both cooperate you will both earn a considerable number of points (20 points each). If you cooperate and the other competes, the other will earn 25 points and you will earn only 5 points. Similarly, if you compete and the other cooperates, you will earn 25 points and the other will earn only 5 points. Finally, if you both choose to compete, you will earn only 10 points each.

You will be presented with numerous matrices of the kind shown above. In each case, you will be asked to indicate whether you choose to compete or to cooperate. As in the matrix above, you will frequently do rather well if you both cooperate, you will do worse if you both compete, and one will often do better than the other if one competes and the other cooperates.

You will be playing with other students who are currently on the computer system. For each new matrix you will be matched with a different person. Thus, you will never play against the same person more than once.

You have been arbitrarily assigned to the bonus group. A random bonus program will occasionally inform you of the strategy that the other player has already chosen. Thus, for example, upon being presented with a new matrix, you may be told that the other player has chosen to compete. You are free to use the bonus information to help you choose your own strategy. (Your strategy will not be revealed to anyone who is playing with you.)

At the end of the game, the points that you accumulate will be converted (via a pre-determined algorithm) to actual money that will be paid to you. The more points you accumulate, the more money you will earn.

Of course, there are no "correct" choices. People typically find certain situations more conducive to cooperation and others to competition. The matrices differ significantly, and their outcomes depend both on your choice and on that of a different player at each turn. Please observe each matrix carefully and decide separately on your preferred strategy in each particular case. Also, be sure to note those cases where the bonus program informs you of the other player's choice. If you have any questions, please ask the person in charge. Otherwise, turn to the terminal and begin.

REFERENCES

Axelrod, R. (1984). *The evolution of cooperation.* New York: Basic Books.

Bacharach, M., & Hurley, S. (1991). Issues and advances in the foundations of decision theory. In M. Bacharach & S. Hurley (Eds.), *Foundations of decision theory: Issues and advances* (pp. 1–38). Oxford: Basil Blackwell.

Bar-Hillel, M., & Margalit, A. (1972). Newcomb's paradox revisited. *British Journal for the Philosophy of Science,* 23, 295–304.

Brams, S. J. (1975). Newcomb's Problem and Prisoner's Dilemma. *Journal of Conflict Resolution,* 19(4), 596–612.

Campbell, R., & Sowden, L. (Eds.). (1985). *Paradoxes of rationality and cooperation: Pris-*

oner's Dilemma and Newcomb's Problem. Vancouver: The University of British Columbia Press.

Cheng, P. W., & Holyoak, K. J. (1985). Pragmatic reasoning schemas. *Cognitive Psychology*, **17**, 391–416.

Cheng, P. W., & Holyoak, K. J. (1989). On the natural selection of reasoning theories. *Cognition*, **33**, 285–313.

Cosmides, L. (1989). The logic of social exchange: has natural selection shaped how humans reason? *Cognition*, **31**, 187–276.

Dawes, R. M., & Orbell, J. M. (1992). The potential benefit of optional play in a one-shot prisoner's dilemma game. To appear in K. Arrow et al. (Eds.), *Barriers to conflict resolution*, New York: Norton.

Eells, E. (1982). *Rational decision and causality*. Cambridge: Cambridge University Press.

Evans, J.St B. T. (1984). Heuristic and analytic processes in reasoning. *British Journal of Psychology*, **75**, 451–468.

Evans, J.St B. T. (1989). *Bias in human reasoning: Causes and consequences*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Evans, J.St B. T., & Lynch, J. S. (1973). Matching bias in the selection task. *British Journal of Psychology*, **64**, 391–397.

Gardner, M. (1973, July). Free will revisited, with a mind-bending prediction paradox by William Newcomb. *Scientific American*, Vol. 229, No. 1, pp. 104–108.

Gardner, M. (1974, March). Reflections on Newcomb's problem: a prediction and free-will dilemma. *Scientific American*, Vol. 230, No. 3, pp. 102–109.

Gibbard, A., & Harper, W. L. (1978). Counterfactuals and two kinds of expected utility. In C. A. Hooker, J. J. Leach, & E. F. McClennen, (Eds.), *Foundations and applications of decision theory* (Vol. 1, pp. 125–162). Dordrecht: Reidel.

Gilhooly, K. J. (1988). *Thinking: Directed, undirected, and creative* (2nd ed). San Diego, CA: Academic Press.

Greene, S. B., & Yolles, D. J. (1990). *Perceived determinacy of unknown outcomes*. Unpublished manuscript, Princeton University.

Griggs, R. A., & Cox, J. R. (1982). The elusive thematic-materials effect in Wason's selection task. *British Journal of Psychology*, **73**, 407–420.

Hammond, P. (1988). Consequentialist foundations for expected utility. *Theory and Decision*, **25**, 25–78.

Hayes, P. (1973). The frame problem and related problems in artificial intelligence. In A. Elithorn & D. Jones (Eds.), *Artificial and human thinking*. San Francisco: Jossey-Bass.

Henslin, J. M. (1967). Craps and magic. *American Journal of Sociology*, **73**, 316–330.

Hofstadter, D. R. (1983, June). Dilemmas for superrational thinkers, leading up to a luring lottery. *Scientific American*. Reprinted in Hofstadter, D. R. (1985). *Metamagical themas: Questing for the essence of mind and pattern*. New York: Basic Books.

Hurley, S. L. (1989). *Natural reasons: Personality and Polity*. New York: Oxford University Press.

Hurley, S. L. (1991). Newcomb's Problem, Prisoner's Dilemma, and collective action. *Synthese*, **86**, 173–196.

Johnson-Laird, P. N., Legrenzi, P., & Legrenzi, S. M. (1972). Reasoning and a sense of reality. *British Journal of Psychology*, **63**, 395–400.

Johnson-Laird, P. N., & Wason, P. C. (1970). A theoretical analysis of insight into a reasoning task. *Cognitive Psychology*, **1**, 134–148.

Kahneman, D., & Tversky, A. (1982). On the study of statistical intuitions. *Cognition*, **11**, 123–141.

Kotovsky, K., Hayes, J. R., & Simon, H. A. (1985). Why are some problems hard?: Evidence from Tower of Hanoi. *Cognitive Psychology*, **17**, 248–294.

Kotovsky, K., & Simon, H. A. (1990). What makes some problems really hard: Explorations in the problem space of difficulty. *Cognitive Psychology*, **22**, 143–183.

Kreps, D., & Wilson, R. (1982). Reputations and imperfect information. *Journal of Economic Theory*, **27**, 253–279.

Langer, E. J. (1975). The illusion of control. *Journal of Personality and Social Psychology*, **32**, 311–328.

Levi, I. (1991). Consequentialism and sequential choice. In M. Bacharach & S. Hurley (Eds.), *Foundations of decision theory: Issues and advances* (pp. 92–122). Oxford: Basil Blackwell.

Lewis, D. (1979). Prisoner's Dilemma is a Newcomb Problem. *Philosophy and Public Affairs*, **8**, 235–240.

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

Manktelow, K. I., & Evans, J.St B. T. (1979). Facilitation of reasoning by realism: Effect or non-effect? *British Journal of Psychology*, **70**, 477–488.

McCarthy, J., & Hayes, P. (1969). Some philosophical problems from the standpoint of Artificial Intelligence. In B. Meltzer & D. Michie (Eds.), *Machine intelligence*. New York: American Elsevier.

Messe, L. A., & Sivacek, J. M. (1979). Predictions of others' responses in a mixed-motive game: Self-justification or false consensus? *Journal of Personality and Social Psychology*, **37**(4), 602–607.

Nozick, R. (1969). Newcomb's problem and two principles of choice. In Nicholas Rescher (Ed.), *Essays in honor of Carl G. Hempel*. Dordrecht: Reidel.

Osherson, D. N., & Markman, E. (1974–75). Language and the ability to evaluate contradictions and tautologies. *Cognition*, **3**(3), 213–226.

Quattrone, G. A., & Tversky, A. (1984). Causal versus diagnostic contingencies: On self-deception and on the voter's illusion. *Journal of Personality and Social Psychology*, **46**(2), 237–248.

Rapoport, A. (1988). Experiments with n-person social traps I: Prisoner's Dilemma, weak Prisoner's Dilemma, Volunteer's Dilemma, and Largest Number. *Journal of Conflict Resolution*, **32**(3), 457–472.

Rapoport, A., & Chammah, A. (1965). *Prisoner's Dilemma*. Ann Arbor: University of Michigan Press.

Rapoport, A., Guyer, M. J., & Gordon, D. G. (1976). *The 2 × 2 game*. Ann Arbor: University of Michigan Press.

Rothbart, M., & Snyder, M. (1970). Confidence in the prediction and postdiction of an uncertain event. *Canadian Journal of Behavioral Science*, **2**, 38–43.

Savage, L. J. (1954). *The foundations of statistics*. New York: Wiley & Sons.

Slovic, P., & Fischhoff, B. (1977). On the psychology of experimental surprises. *Journal of Experimental Psychology: Human Perception and Performance*, **3**, 544–551.

Sobel, J. H. (1991). Some versions of Newcomb's Problem are Prisoner's Dilemmas. *Synthese*, **86**, 197–208.

Strickland, L. H., Lewicki, R. J., & Katz, A. M. (1966). Temporal orientation and perceived control as determinants of risk-taking. *Journal of Experimental Social Psychology*, **2**, 143–51.

Tversky, A., & Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review*, **90**, 293–315.

Tversky, A., & Kahneman, D. (1986). Rational choice and the framing of decisions. *Journal of Business*, **59**(4, 2), 251–278.

Tversky, A., & Shafir, E. (1992). The disjunction effect in choice under uncertainty. *Psychological Science*, in press.

Wason, P. C. (1966). Reasoning. In B. M. Foss (Ed.), *New horizons in psychology* (Vol. 1). Harmandsworth: Penguin.

Wason, P. C. (1969). Structural simplicity and psychological complexity: Some thoughts on a novel problem. *Bulletin of the British Psychological Society*, **22**, 281–284.

Wason, P. C. (1983). Realism and rationality in the selection task. In J.St B. T. Evans (Ed.), *Thinking and reasoning: Psychological approaches*. London: Routledge & Kegan Paul.

Wason, P. C., & Johnson-Laird, P. N. (1970). A conflict between selecting and evaluating information in an inferential task. *British Journal of Psychology*, **61**, 509–515.

(Accepted April 7, 1992)