Naive Strategic Thinking

Eugenia Steingold (genya@wjh.harvard.edu)
Department of Psychology
Harvard University
33 Kirkland Street
Cambridge, MA 02138

P. N. Johnson-Laird (phil@princeton.edu)
Department of Psychology
Princeton University
Green Hall
Princeton, NJ 08540

Abstract

The mental model theory postulates that reasoners build mental models of the situations described in premises, and that each model represents a possibility. The theory's main principle — the principle of *truth*— is that these representations are incomplete, because mental models represent only what is true, and not what is false. The present paper defends an analogous principle of *self-interest*: when individuals have to think strategically, they tend to represent only their own options and payoffs, not those of their opponents. Four experiments have corroborated this principle.

Introduction

Strategic thinking underlies those decisions that depend on what other individuals are likely to decide. Such decisions are ubiquitous in business, politics, and daily life. To outwit an adversary, or to maximize the benefits of co-operation with a partner, you often need to think about what this other individual is likely to choose to do. The theory of games provides a normative theory of strategic thinking, especially since John Nash's theorem determining those strategies that are optimal for all players — the so-called Nash equilibrium (see, e.g., Dixit and Nalebuff, 1991). However, as Reinhard Selten, co-winner of the Nobel prize with Nash in 1994, has remarked: game theory is rational theology (personal communication). That is, na ve individuals in daily life are unlikely to abide by its canons of rationality. The term na ve here refers to individuals who have not mastered game theory; it does not impugn their intelligence. Thinking about what other people may be thinking is indeed likely to be difficult, because it calls for a representation, not only of your own state of mind, but also of another person's preferences, and for inferences about their likely strategies. If you fail to think through their optimal strategy, then you are unlikely to select a rational strategy for your self.

Consider as an example a simple game in which you have two options (A and B), your opponent has two options (A and B), you both make your choices simultaneously, and you both receive payoffs as a function of your choices. We show your payoff followed by your opponent's payoff in each cell:

		Your opponent's options	
		A	B
Your options	A	**$4/$4**	**$1/$3**
	B	**$2/$4**	**$6/$1**

If your opponent chooses A then you should choose A and receive $4. But, if your opponent chooses B, you should choose B and receive $6. Hence, to make the right choice you should think about what your partner is likely to choose. If she is rational, she should think: If I choose A, then regardless of what my opponent chooses, I get $4; but if I choose B, then regardless of what my opponent chooses, I will get less than $4. Therefore, I should choose A (because B is *dominated* by this strategy). Hence, if you are rational and you know that your opponent is rational, then you should choose A. This sort of thinking exemplifies game theory, which formalizes optimal strategies on the assumption that players are rational.

In contrast, the present paper argues that the task of inferring rational predictions about other players choices is too difficult for naive human players. The paper accordingly proposes an alternative account of how na ve individuals make strategic decisions.

Mental Models and Na ve Strategic Thinking

The mental model theory postulates that

reasoning depends on understanding the meaning of premises and using this meaning and general knowledge to envisage the states of affairs that are possible given the truth of the premises (Johnson-Laird and Byrne, 1991). Each mental model represents a possibility. A conclusion is necessary if it holds in all the models of premises. It is possible if it holds in at least one model of premises (Bell and Johnson-Laird, 1998). And its probability — its likelihood of being true — depends on the proportion of models of the premises in which it is true (Johnson-Laird, Legrenzi, Girotto, Legrenzi, and Caverni, 1999). The theory has also been extended to account for meaning and reasoning with causal relations (Goldvarg and Johnson-Laird, 2001). A fundamental assumption of the theory is the principle of *truth*: in order to minimize the load on working memory, mental models normally represent only what is true. The failure to represent what is false gives rise to illusory inferences of various sorts, i.e. inferences that nearly everyone makes, but that are wrong (see, e.g., Goldvarg and Johnson-Laird, 2000).

The model theory extends naturally to human strategic thinking. It postulates that individuals construct mental models of the options and payoffs. In order to minimize the load on working memory, however, the theory is based on an assumption analogous to the principle of truth. According to the principle of *self-interest*, mental models normally represent only individuals own options and payoffs. The failure to represent other players payoffs should give rise to systematic errors in strategic thinking. The aim of our experiments was to test the principle of self-interest.

Experiment 1: Games as payoff matrices

Our first experiment examined whether or not individuals spontaneously represent their partner s payoffs in games in which their optimal choice depends on their partner s choice. The participants played 25 games, each presented in the form of a payoff matrix. Five of the games were *independent*, that is, the participants optimal choice did not depend on their partners choice. Here is an example of an independent game in which only the participant's payoffs are shown:

		Your partner's options	
		A	B
Your options	A	**$5**	**$6**
	B	**$4**	**$5**

In the absence of information about their partner's payoffs, there are three possible strategies that the participants could adopt:

1) They could maximize their mean payoffs. In this game it is option A, which leads to a mean payoff of $5.5.
2) They could maximize their minimum gain. Option A yields the larger minimum gain (of $5).
3) They could maximize their maximum gain. Option A yields the maximum gain (of $6).

Hence, in general, with independent games all three strategies lead to the same choice.

The remaining 20 games were *dependen*t, that is, to make the optimal choice, the participants needed to know their partner's choice, e.g.:

		Your partner's options	
		A	B
Your options	A	**$5**	**$7**
	B	**$6**	**$4**

Here, if the partner chooses A, then the participant should choose B, whereas if the partner chooses B, the participant should choose A. Granted the principle of self-interest, however, the participants should not notice the difference between independent and dependent games. They should be prepared to play both sorts of game without knowledge of their partner's payoffs. In the present game, the three preceding strategies all lead to the choice of option A. The experiment included three other sorts of dependent game that allowed us to identify the participants' strategies.

We tested 20 Princeton undergraduates, and in this and the other experiments, we checked that they were na ve about game theory. The instructions explained that the participants would be presented with simple games, and that their task was to decide which option they would choose in each game. Before the participants responded to each game, they were asked: Do you have all the necessary information to play the game? If *No*, what else do you need to know to make the decision? The participants then made a choice.

Results

The majority of the participants responded that they had the all the necessary information to play the game: only five participants requested their partners' payoffs on more than half the trials. Hence, the bias to play the games without knowing these payoffs was reliable (Sign test, $n = 20$, $p < .025$). There was no reliable difference between dependent games and independent games in the tendency for participants to play in ignorance of their partners' payoffs (15 participants in both cases played on more than half the trials; $F < 1$, $p > .25$). The preferred strategy was to maximize the maximum possible gain (14 out of 20 participants on more than half the trials, $p < .05$).

Experiment 2: Conditional descriptions of games

The previous experiment presented games in the form of payoff matrices. Skeptics could argue that this format is not familiar enough to na ve participants to enable them to do justice to their ability. We therefore carried out a similar experiment, but each game was presented in a set of conditional assertions, which are easy to understand. For example, an independent game was described in the following way:

> If you choose A and your partner chooses A, you will receive $5.
> If you choose A and your partner chooses B, you will receive $6.
> If you choose B and your partner chooses A, you will receive $4.
> If you choose B and your partner chooses B, you will receive $5.

One group of participants received such descriptions, which are incomplete because they do not say anything about the partner's payoffs. A second group of participants received only partial descriptions of their own payoffs, e.g:

> If you choose A and your partner chooses A, then you will receive $5.
> If you choose B and your partner chooses B, then you will receive $6.

These descriptions were only for cases in which both players chose the same options (AA and BB). A third group of participants received descriptions of only their partner's payoffs but not their own, e.g.:

> If you choose A and your partner chooses A, your partner will receive $5.
> If you choose A and your partner chooses B, your partner will receive $6.
> If you choose B and your partner chooses A, your partner will receive $4.
> If you choose B and your partner chooses B, your partner will receive $5.

According to the principle of self-interest, the participants should be more likely to ask for their own payoffs than for the payoffs of their partners. The participants in each group were asked three questions: Would you play the game? Do you have all the necessary information to make a choice? If *No*, what else do you need to know to make your choice? The procedure and the 25 games were the same as those in Experiment 1. We tested ten Princeton undergraduates in each group.

Results

In the group that knew only their own payoffs, 100% of responses were decisions to play both dependent and independent games. In the group that knew only their own partial payoffs, 100% of responses were decisions to play both sorts of games. But, as predicted, in the group that knew only their partners' payoffs only 40% of the responses (4 participants on more than half the trials) were decisions to play the games. The difference between the three groups is significant, Kruskal-Wallis ($\chi2 = 14.5$, $p < .001$). Hence, the majority of participants who had only their partners payoffs refused to play the games without knowing their own payoffs. The participants' preferred strategy was to maximize their average payoff (80% of strategies, Binomial test, $p < .005$).

Experiment 3: Games with real opponents

The participants in the previous experiments might have played games without knowing their partners' payoffs because the games seemed artificial and unreal. We therefore carried out a replication of Experiment 2 in a way that was closer to real games. The participants played against real opponents and they could win real money. Each experimental session tested two participants at a time, who had not previously met. They were told that they would be playing a set of games against each other. And they then went to different rooms, and each participant received the instructions and carried out the games. The design and procedure were otherwise identical to Experiment 2. We tested 10 Princeton undergraduates in each group. The participants were told that the person who won the most nominal money in the games would enter a lottery with a possibility to win $20.

Results

The step towards verisimilitude slightly enhanced performance, but the results otherwise replicated the previous experiment. Two participants who knew only their own payoffs requested additional information on more than half the trials. Five participants who knew only their opponents' payoffs requested additional information on more than half the trials. Nine participants who knew only their own partial payoffs requested additional information (but only about the rest of their own payoffs). The difference between the three groups was significant (Kruskal-Wallis $\chi 2 = 11.21$, $p < .005$). A planned comparison between those who knew their own payoffs and those who knew their opponents' payoffs was also significant ($z = 3.2$, $p < .03$). The participants' preferred strategy was again to maximize their average payoff (70% of participants, $p < .005$). Hence, on the whole, individuals who knew their own payoffs were happy to play with real partners and for real money.

Experiment 4: The effect of losses.

In the previous experiments, none of the games threatened the participants with a loss of money depending on the outcome of a game. People are known to think differently about losses than about gains. They are risk seeking when they think about losses but risk averse when they think about gains (Tversky and Kahneman, 1981). Gains and losses are represented asymmetrically. The negative effect of a loss of a certain amount is greater than the positive effect of a gain of the same amount. Hence, people may represent games with losses differently from the way they represent games with only gains. Likewise, losses may also affect individuals' strategic thinking. In particular, they might think more carefully and require more information about losses than they do when they think about gains. This greater care may, in turn, elicit a need for information about their partner's options and payoffs. A different possibility, however, is that potential losses would make individuals anxious and confused. Hence, they might represent less information than usual, and make even fewer requests for their partner's payoffs. To examine these possibilities, Experiment 4 used a set of games similar to those in the previous experiments, but it included payoffs that were losses. One group of participants played games with gains only in the first block of the experiment and games with gains and losses in the second block. Another group received the two blocks in the opposite order. The procedure was similar to Experiment 3, using both real partners and a real monetary reward.

Results

The participants who received only gains in the first block were more likely to play in that block (97% of responses) than the group who received gains and losses in the first block (80% of responses in the first block, Kruskal-Wallis, $p < .005$). Overall, there was no difference between the first and second blocks in the propensity to play. But, the effect of loss in inhibiting participants from playing was greater in the second block than in the first block (Kruskal-Wallis $\chi 2 = 10.4$, $p < .05$).

Both groups were equally likely to request additional information (36% and 35% of responses). There was no significant effect of block. However, there was again a significant interaction between the group and the block, Kruskal-Wallis ($\chi 2 = 10.5$, $p < .05$). Only one participant who received only gains in the first block requested more information on more than half the trials; whereas six of the participants who received gains and losses in the first block requested more information on more than half the trials (Kruskal-Wallis $\chi 2= 11.2076$, $p < .005$). Thus, the participants did not notice that something was missing when they played games with gains only. But, when they moved on to games with gains and losses, they tended to notice that the games were incomplete. In contrast, those who received games with gains and losses showed no change in their requests for additional information when they moved on to games with only gains (36% requests in the first block, and 36% requests in the second block). In sum, losses had their largest effects when they appeared after the games with gains only. The participants preferred strategy was again to maximize average payoffs (85% of strategies, $p < .005$).

General Discussion

Our everyday experiences often call for representing what other people think, believe and desire. Strategic decision making is one of many situations that call for individuals to think about their partners preferences and choices. Moreover, they often have to take into account their opponents considerations of their own options, and so on and on. This problem may easily become intractable, and, at the very least, highly complex. The present paper has defended the principle of *self-interest*: when individuals have to think strategically, they tend to focus on their own options and payoffs. The experiments corroborated this principle. Experiment 1 demonstrated the effect with payoff matrices. Experiment 2 replicated it with conditional descriptions of games. Both Experiment 2 and 3 showed that individuals notice when their own payoffs are missing, and that they then request them. When their own payoffs are described only in part, they request information about the rest of their payoffs. Otherwise, they are prepared to play in ignorance of their partner's payoffs both for dependent and independent games. Experiment 4 showed that when participants can lose money depending on their choices, the majority of them were still prepared to play without asking for additional information. It is rational to play independent games in ignorance of your partner's payoffs or likely choice of option; but it is not rational to play dependent games in these circumstances.

Why don't individuals realize that they should ask for their partner's payoffs? The simplest game that elicits the need for strategic thinking is one in which two players each have two options. Such a game, however, calls for reasoners to hold in mind four separate possibilities in order to work out an optimal strategy. Four simple possibilities are known to be at the limit of typical adult competence in reasoning (Johnson-Laird and Byrne, 1991), and so it is natural for reasoners to focus on their own payoffs. This tendency, as we have shown, is influenced by experimental manipulations. We believe that if na ve individuals were given immediate payoffs after each game, and ran the risk of loss with sub-optimal choices, then they would soon realize the need to infer their opponents' likely choices. They should then be able to eliminate clearly dominated strategies in simple games. But, it is unlikely that they would be able to compute an equilibrium for more complex games — the number of possibilities would overwhelm the processing capacity of working memory. As our results imply, na ve individuals do not normally envisage the payoffs of their opponents, and that is why they do not ask for this information. As the adage says, it is hard to put oneself into others' shoes. This difficulty can, in turn, lead to an erroneous choice of option in those cases in which the optimal choice depends on the choices of others.

Acknowledgements

This research was supported in part by a grant from the National Science Foundation (Grant BCS 0076287) to the second author to study strategies in reasoning. We thank the following colleagues for their advice: Victoria Bell, Zachary Estes, Mary Newsome, Vladimir Sloutsky, Jean-Baptiste van der Henst, and Yingrui Yang.

References

Bell,V., and Johnson-Laird, P.N. (1998). A model theory of modal reasoning. Cognitive Science, 22, 25-51.

Dixit, A.K., and Nalebuff, B.J. (1991) Thinking Strategically: The Competitive Edge in Business, Politics, and Everyday Life. New York: Norton.

Goldvarg, E., and Johnson-Laird, P.N. (2001). Na ve causality: a mental model theory of causal meaning and reasoning. Cognitive Science, 25, 565-610.

Goldvarg, Y., and Johnson-Laird, P.N. (2000). Illusions in modal reasoning. Memory & Cognition, 28(2), 282-294.

Johnson-Laird, P.N., and Byrne, R.M.J. (1991). Deduction. Hillsdale, NJ: Lawrence Erlbaum Associates.

Johnson-Laird, P.N., Legrenzi, P., Girotto, V., Legrenzi, M.S., and Caverni, J-P. (1999). Na ve probability: A mental model theory of extensional reasoning. Psychological Review, 106, 62-88.

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