Journal of Behavioral Decision Making, *J. Behav. Dec. Making*, **25**: 95–108 (2012)
DOI: 10.1002/bdm.717

Levels of Theory-of-Mind Reasoning in Competitive Games

ADAM S. GOODIE[1]*, PRASHANT DOSHI[2] and DIANA L. YOUNG[3]

[1] *Department of Psychology, University of Georgia, USA*

[2] *Department of Computer Science, University of Georgia, USA*

[3] *Department of Psychological Science, Georgia College and State University, Georgia, USA*

ABSTRACT

The literature on recursive theory of mind (TOM) reasoning in interactive decision making (reasoning of the type "I think that you think that I think...") has been pessimistic, suggesting that adults attribute to others levels of reasoning that are low and slow to increase with learning. In four experiments with college-age adults playing sequential games, we examined whether choices and predictions were consistent with believing that others pursue their immediate self-interest, or with believing that others reason through their own decision making, with fixed-sum games that were simpler and more competitive. This manipulation led to higher-level default TOM reasoning; indeed, reasoning against a lower-level opponent was frequently consistent with assuming the opponent's reasoning to be higher-level, leading to sub-optimal choices. We conclude that TOM reasoning is not of a low level in all game settings; rather, individuals may display effective TOM reasoning, reflecting realistic assumptions about their opponents, in competitive and relatively simple games. Copyright © 2010 John Wiley & Sons, Ltd.

KEY WORDS decision making; games; learning; recursive reasoning; theory of mind

INTRODUCTION

In the developmental psychology literature, children as young as 4 years appear to appreciate the beliefs, desires, and emotions of others, which is generally referred to as "theory of mind" (TOM; Wellman & Gelman, 1998; Wellman, Cross, & Watson, 2001). Children as young as 2 expect others to feel happy if their own desires are met and unhappy if they are not (Wellman & Banerjee, 1991), regardless of their own feelings (Flavell, Mumme, Green, & Flavell, 1992). There is a gap in young children's understanding of thought processes (Flavell, Green, & Flavell, 1998), though this is largely closed by the age of 8 (Flavell, Green, & Flavell, 2000). Overall, the developmental literature documents timely progress toward nuanced and accurate understanding of others' cognitive processes.

It may be surprising, then, that the adult decision making literature on reasoning levels in recursive reasoning (reasoning of the type "I think that you think that I think...") is pessimistic, suggesting that individuals assume others have low levels of reasoning, and learn only slowly and partially to respond optimally to others who demonstrate higher levels of reasoning (e.g., Hedden & Zhang, 2002). This suggests that individuals either lack high levels of recursive reasoning, or have low opinions of their fellow human beings' levels of reasoning.

Perner and Wimmer (1985) developed a hierarchical system of classifying TOM in developmental studies, which was refined in interpreting studies of adult decision making (Hedden & Zhang, 2002). This system mirrors the model of "level-*k* reasoning" in the economic literature (Camerer, Ho, & Chong, 2004; Stahl & Wilson, 1994, 1995), which has been applied to beauty contests (Ho, Camerer, & Weigelt, 1998), auctions (Crawford & Iriberri, 2007), and strategic lying (Crawford, 2003). Within this system, in *0th-level reasoning*, an individual considers only his or her immediate desires and beliefs, and treats others as inert. In *1st-level reasoning*, an individual expects others to act with regard to their own immediate desires and beliefs and to consider others (including the reasoner) as inert. Note that 1st-level reasoning amounts to attributing 0th-level reasoning to others. In *2nd-level reasoning*, a reasoner expects others to take the reasoner's own desires and beliefs into account, or in other words attributes 1st-level reasoning to others. In general, *n*th-level reasoning consists of attributing (*n-1*)th-level reasoning to others.

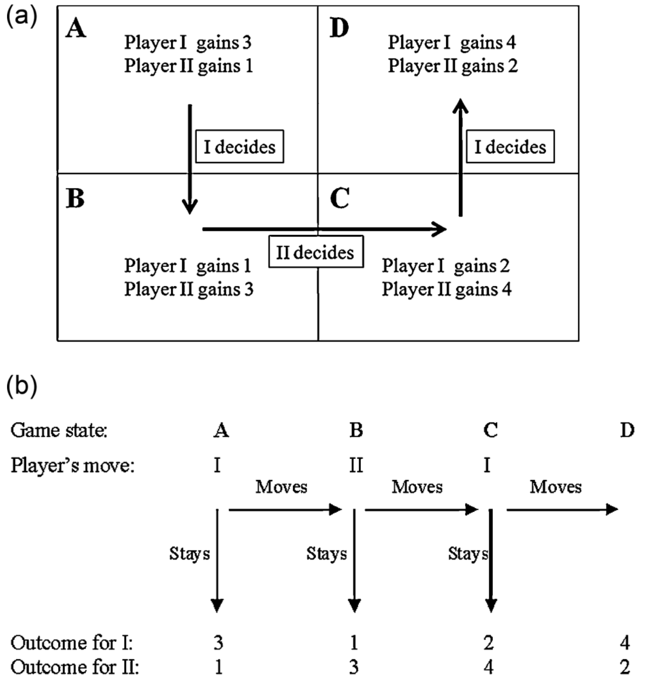
An example of a game that permits examination of levels of reasoning is depicted in Figure 1a and b in both matrix and tree representations. Two players begin at cell A and perform actions alternately. Player I decides whether to end the game at state A or advance the game state from A to B; if the game advances to B, then Player II decides whether to end there or move to C; and if the game advances to C, then Player I decides whether to end there or advance to D. Each player obtains the outcome indicated for him at the state where the game ends. Each player prefers higher numbers to lower numbers and is indifferent to the outcome obtained by the other.

The mutually-rational solution to this game is as follows: If the players find themselves at C, then Player I prefers 4 to 2 and would move from C to D. Thus Player II chooses at B between the outcomes at states B and D. Because Player II prefers 3 to 2, she would choose to stay at B rather than move to C. Thus Player I chooses at A between the outcomes at A and B. Preferring 3 to 1, Player I would stay at A.

However, a 1st-level TOM reasoner would not stay but move at A. A 0th-level Player II would move from B to C, seeking to improve from 3 to 4 as an outcome and not

---

* Correspondence to: Adam S. Goodie, Department of Psychology, University of Georgia, Athens, GA 30602-3013, USA. E-mail: goodie@uga.edu



(c)

Level of Reasoner	What the Reasoner Thinks as Player I at A
0th	I can have 3 at A, or 1 at B. I prefer 3 to 1; therefore I'll stay at A.
1st	If I move to B, then Player II will think: *I can have 3 at B, or 4 at C.* *I prefer 4 to 3; therefore I'll move from B to C.* When Player II moves from B to C, I can move to D and get 4. Because I prefer 4 at D, rather than 3 at A, I'll move from A to B.
2nd	If I move from A to B, then Player II will think: *If I move from B to C, then Player I will think:* ***I can have 2 at C, or 4 at D.*** ***I prefer 4 to 2; therefore I'll move from C to D.*** *Because Player I will move from C to D, I'm really choosing between 3 at B and 2 at D.* *I prefer 3 to 2; therefore I'll stay at B.* Because Player II will stay at B, I can have 3 at A or 1 at B. I prefer 3 to 1; therefore I'll stay at A.
	Note: *Italics* indicate what Player I thinks that Player II thinks. ***Bold italics*** indicate what Player I thinks that Player II thinks that Player I thinks.

Figure 1. A three-stage general-sum sequential game, adapted from Hedden and Zhang (2002), in (a) matrix, (b) tree format, and (c) typical reasoning that is expected in 0th, 1st and 2nd level reasoning.

contemplating that Player I would move from C to D. Hence, Player I reasoning at the first level would move from A to B. Figure 1c provides examples of how 0th-, 1st- and 2nd-level reasoners would approach the game depicted in Figure 1.

Hedden and Zhang (2002) conducted two experiments in which participants played 64 different games in the role of Player I, with the matrix structure depicted in Figure 1a, differing in the various outcomes in each cell. The programmed partner was either a 1st-level or a 0th-level reasoner, which were dubbed respectively as "*predictive*" and "*myopic*" opponents. At first, participants generally responded as if expecting the partner to be myopic. Those for whom this assumption was correct continued to perform well. Those with a predictive partner learned slowly and incompletely to respond optimally. Similarly, Stahl and Wilson (1995) reported performance over 12 games in which most participants failed to attribute strategic reasoning to their co-players. Likewise, Camerer, Ho, and Chong (2004) concluded that participants reach an average of 1.5 steps in many contexts, attributing 0th or 1st level reasoning to their partners.

The present research

We sought theoretically motivated demonstrations of higher level reasoning than has previously been shown in the adult decision making literature through two primary manipulations of the game: competitiveness and realism. Research has shown that individuals attend more to competitive games than to equivalent non-competitive games (Bornstein, Gneezy, & Nagel, 2002; Lieberman, 1997; Nickell, 1996; Rapoport & Budescu, 1992; Rindova, Becerra, & Contardo, 2004; see Johnson, Maruyama, Johnson, Nelson, & Skon, 1981 for a contrasting perspective). At a cognitive level, evolutionary pressures may have led to the development of modules that are adapted to respond optimally to socially relevant situations (Cosmides, 1989; Gigerenzer & Hug, 1992). In order to achieve a competitive environment, we transformed the 2x2 matrix games from general-sum games, in which the outcomes for players are mutually independent, to fixed-sum games, in which any increase in gain to one player implies an equivalent loss to the other. Fixed-sum games are inherently competitive. In addition to making a more competitive environment, a fixed-sum structure also makes the game simpler, as there are four rather than eight outcome values for players to attend to.

There is also evidence that individuals perform better in concrete, realistic settings than in abstract, vague settings (e.g., Griggs & Cox, 1982). To manipulate realism, one group (Abstract) saw the formats presented in Figure 1. The other group (Realistic) was additionally provided with a military cover story and accompanying graphical representations of a soldier (Player I) moving to various locations in an attempt to obtain information, while an adversarial aircraft (Player II) patrolled the area trying to prevent such activity.

We designed the games in terms of probability of gain, rather than magnitude of gain as has been utilized in previous research. There were several reasons for this: First, it formed an informative extension on previously used methods. Second, many competitive games consist of vying for an indivisible good, such as winning a game, a job, or space in a selective scholarly journal. Hence any action taken by a player does not affect the magnitude of a possible gain but rather its likelihood. Third, although gain probability has some properties that distinguish it from gain magnitude, it also has critical properties in common with gain magnitude,

DOI: 10.1002/bdm

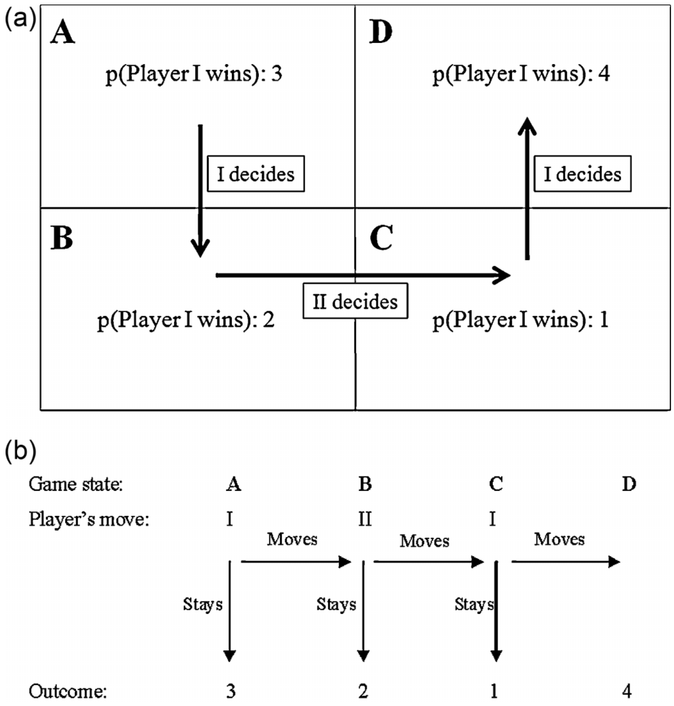


Figure 2. A three-stage fixed-sum sequential game in (a) matrix and (b) tree format.

most important of which in the present context is a presumption of stable preferences. Just as one can be presumed to prefer \$4 to \$1, one can be presumed to prefer a 40% chance of winning \$1000 to a 10% chance of winning \$1000.

EXPERIMENT 1: THE FIXED-SUM PROBABILISTIC GAME

Because our game had a fixed sum of outcomes, each cell's outcome can be characterized by a single number. We use greater numbers to reflect greater likelihoods of Player I winning; hence, Player I aims to end at the cell with the greatest possible number, and Player II aims to end at the cell with the least possible number.

An example is presented in Figure 2. Player I would prefer 4 (at D) to 1 (at C) and thus would move from C. Player II thus chooses at B between 2 (at B) and 4 (at D). Preferring lower numbers, Player II would stay at B. Thus Player I chooses at A between 3 (at A) and 2 (at B) and stays at A. If Player II considers only immediate payoffs and does not reason about Player I's desires, then Player II would move from B to C. Player I should then move from A to B, relying on Player II moving from B to C. Thus, in the game presented in Figure 2, a 1st-level reasoner moves at A, whereas a 2nd-level reasoner stays at A. This sequence, 3-2-1-4, is the only one out of 24 possible orderings of 1-4 that distinguishes behaviorally between attributing myopic and predictive reasoning in this way.

For the critical trials, we constructed quadruplets of probabilities in the 3-2-1-4 ordering, using all probabilities in [.1,.9] in increments of .05. For any game, the difference between any probability and the next highest probability was 0.15, 0.2 or 0.25. Using these rules we devised 40 test trials, grouped into four blocks of 10 trials for analyses.

Methods

Participants

We recruited 136 (70 female) participants who met basic criteria of learning the rules of the game from the Research Pool of the Psychology Department at the University of Georgia in exchange for partial psychology course credit and performance-based monetary incentive. Twenty-six individuals failed to meet the basic learning criteria.

Up to six participants were studied at a time at individual computer workstations in two separate rooms, each with three workstations. Participants were given verbal instructions together, which indicated that they would play against one of the players in the other room. All players played as Player I, against a computer opponent.[1]

Trials

The first 25 trials comprised a training phase that introduced participants to the task but which did not allow the participants to learn whether their opponent was a 0th- or 1st-level reasoner. The training phase consisted of 10 trials of the 1-4-2-3 type, 4 trials of the 2-3-4-1 type, 3 trials of the 3-4-2-1 type, and 8 trials of the 4-2-3-1 type. These trial types

---

[1] The design used deception, as participants thought they were playing against other humans but in fact were playing against a computer program. This was necessary because the opponent, Player II, not only needed to be perceived as human, but also needed to utilize consistent 0th- or 1st-level reasoning. Groups were constrained to comprise even numbers of participants divided equally between the rooms.

DOI: 10.1002/bdm

were selected so as not to reveal the opponent's strategic type. Notice that in all of these game types, the rational action of the opponent remains the same whether the opponent is 0th- or 1st- level. Trials of the various types were randomly interspersed. During trials 11-25 in the training phase, participants were required to display five consecutive trials without an error in order to move on to the test phase, although this requirement was not revealed to them. Once the criterion was met, the game advanced to the test phase. If the participant did not reach criterion after 25 trials, then she did not advance to the test phase.

The test phase consisted of 80 trials, with the 40 test trials of the 3-2-1-4 type interspersed randomly with 20 trials of the 2-3-1-4 type and 20 trials of the 3-2-4-1 type. Following each game, the outcome was determined according to the probability in effect at the conclusion of the game. (For the first five trials, the player with the higher probability of winning was determined to be the winner to avoid unrepresentative early experience.)

In order to enhance the appearance that the opponent was human, Player II took 45 seconds to respond in the first trial, 30 seconds in the second trial, 20 seconds in the third trial, 15 seconds in the fourth trial, 10 seconds in the fifth trial, and 2–6 seconds (selected randomly from a rectangular distribution) for the remaining trials.

Participants were paid \$0.50 for each game they won.

Results and discussion

Data were analyzed using a 2x2 between-subjects design to examine the effects of opponent type and realism on achievement scores and learning.

We defined an *achievement score* as the proportion of games in which the highest possible probability of success was attained, given the opponent's strategy (myopic or predictive). Mean achievement scores are depicted in Figure 3.

Performance in a competitive, fixed-sum game was markedly different from what had been previously observed in comparable general-sum games. Default responding on the test trials was more consistent with 2nd- or higher-level reasoning, whereas in previous studies it had been more consistent with 1st-level reasoning. Second-level reasoning results in high achievement scores against a predictive opponent and low achievement scores against a myopic opponent. Average achievement in the first block was .877 among participants with a predictive opponent, well above .5, and .300 among those with a myopic opponent, well below .5. Overall achievement scores in the Predictive group ($M = 0.953$) were also significantly greater than the Myopic group ($M = 0.443$; $F(1,132) = 125.6$, $p < .001$; partial $\eta^2 = .488$).

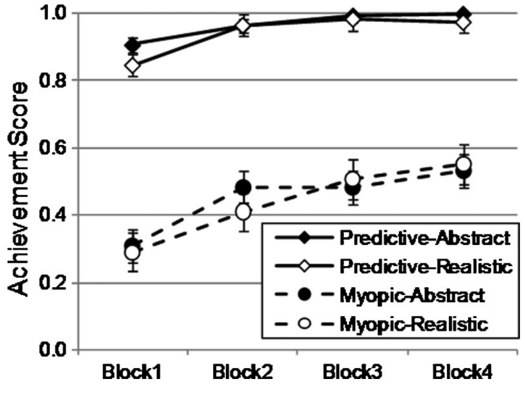


Figure 3. Achievement scores in Experiment 1, reflecting proportion of trials that optimized outcomes for all groups.

We analyzed patterns of learning with regard to achievement scores. We define a metric L as the trial after which performance over the most recent 10 test trials (or over all trials during trials 5-9) *never* failed to achieve statistical significance (cumulative binomial probability < .05). For participants who never permanently achieved statistical significance, L was assigned a value of 40. L is thus bounded by [5,40]. In the Myopic groups, 40 of the 67 participants never permanently established above-random performance. Only four participants achieved L scores of 5-10. In contrast, among the Predictive groups, 47 out of 69 participants achieved scores of 5-10, and only one participant was assigned a value of 40. Overall, participants learned more slowly against a Myopic opponent ($M = 31.68$, $SE = 2.06$) than against a Predictive opponent ($M = 10.08$, $SE = 1.07$; $F(1,132) = 176.1$, $p < .001$; partial $\eta^2 = .572$).

Note that participants could learn about the opponent's strategy only when they chose to move at A, which was the optimal response for those in the Myopic condition but the non-optimal response for those in the Predictive condition. Those in the Predictive condition thus achieved near-perfect performance with only one or two opportunities on average to observe the opponent's response. Those in the Myopic condition, as they chose to move more often, obtained learning opportunities at an accelerating rate, which however did not culminate in optimal performance.

Did participants believe they were playing against other humans, rather than against a computer? We conducted a post-experimental survey that included the questions:

(1) *At the beginning of the study, how strongly did you believe that you were playing the games against another person?*
(2) *By the end of the study, how strongly did you believe that you were playing the games against another person?*

Both questions were answered on a 7-point Likert scale anchored at 1 (''Completely sure I WASN'T playing another person''), 4 (''No idea at all''), and 7 (''Completely sure I WAS playing another person''). Average responses showed substantial believability at the beginning of the study ($M = 5.06$, $SD = 1.73$), which declined by the end of the study ($M = 2.69$, $SD = 1.95$). Between-group differences were non-significant.

All main effects and interactions involving the format of the game, abstract versus realistic, were not significant.

In Experiment 1 we demonstrated that in a competitive, fixed-sum game, default performance was consistent with participants' using 2nd- or higher-level reasoning, and that adaptive learning was faster and more complete in

DOI: 10.1002/bdm

competition against a 1st-level reasoner than against a 0th-level reasoner. We observed no evidence to support the role of realism in levels of reasoning. This could be because our realism manipulation was weak, and does not necessarily imply that realistic settings have no impact on TOM reasoning.

EXPERIMENT 2: FIXED- AND GENERAL-SUM GAMES

In Experiment 2 we directly investigated the impact of game type (fixed-sum versus general-sum) on levels of reasoning, replicating Experiment 1 for the fixed-sum condition and extending prior results for the general-sum condition. Also, participants were asked to predict whether their opponent would stay at B or move to C, with wording appropriate for the Abstract and Realistic conditions.

Methods

Participants

Participants were recruited from the same population as Experiment 1 and divided randomly into general- and fixed-sum games. For the fixed-sum game we recruited 118 participants (60 female), and for the general-sum game we recruited 114 participants (65 female) who met basic criteria of learning the rules of the game. Twenty-two failed the learning criterion in the fixed-sum game, and 31 in the general-sum game.

General-sum game

We used a similar general-sum game to Hedden and Zhang (2002). The first 24 trials comprised a training phase to introduce participants to the task in general without providing trials from which participants could learn their opponent's level of reasoning. The testing phase consisted of two blocks of 20 trials, with each block comprising 16 test trials and four distractor trials. As in the fixed-sum game, the outcomes referred to likelihood rather than magnitude of gains for Player I (the participant), with outcomes assigned as 20, 40, 60, and 80%.

Participants in the Abstract condition were presented with the 2x2 matrix and tree representations of the game. Those in the Realistic condition read an illustrated military cover story.

Game outcomes were determined in the same manner as in Experiment 1; however, no monetary incentive was used (Case, Fantino, & Goodie, 1999; Goodie & Fantino, 1995, 1996).

Finally, we recorded reaction times, both for participants to form and express their expectations (*Prediction RT)* and then to make and express their decisions at A (*Decision RT*). Because 2nd-level reasoning is more complex than 1st-level reasoning, reaction times were expected to be longer when 2nd-level reasoning was engaged.

Results and discussion

As in Experiment 1, there was an absence of significant main effects or interactions involving the manipulation of realism. Consequently we report analyses that collapse across game format.

We first present separate analyses of the two game types and then analyze them in comparison with each other.

Fixed-sum game

Mean achievement scores for the fixed-sum game are depicted in Figure 4. Performance in the fixed-sum game again showed default responding that was consistent with 2nd- or higher-level reasoning. Default achievement (that in Block 1) was .887 among participants with a Predictive opponent, and only .410 among those with a Myopic opponent. Overall achievement scores for those facing a Predictive opponent (.957, SE = .007) were significantly greater than those facing a Myopic opponent (.576, SE = .042; $F(1,116) = 84.373$, $p < .001$; partial $\eta^2 = .421$).

In the Myopic groups, 29 of the 58 participants never permanently established above-random performance, and only six participants achieved L scores of 5-10. Among the Predictive groups, 40 out of 60 participants achieved L scores of 5-10, and only two participants were assigned a value of 40. Overall, participants learned more slowly against a Myopic opponent (M = 29.9, SE = 1.38) than against a Predictive opponent ($M = 10.7$, SE = 1.35; $F(1,114) = 99.5$, $p < .001$; partial $\eta^2 = .462$).

Mean *prediction scores* were calculated, reflecting the proportion of predictions that were consistent with 2nd-level TOM reasoning in each block. Low scores suggest more 1st-level reasoning; high scores suggest more 2nd-level reasoning. Prediction score results are depicted in Figure 5 and support the conclusion that participants who faced a predictive opponent made predictions consistent with 2nd-level reasoning, with an overall mean prediction score of .787 (SE = .036) for the Predictive condition compared with a mean prediction score of .216 (SE = .036) for the Myopic condition ($t(116) = 11.368$, $p < .001$; partial $\eta^2 = .527$). A mixed-model test was then conducted using the four 10-trial blocks as the within-subjects factor and opponent type as the

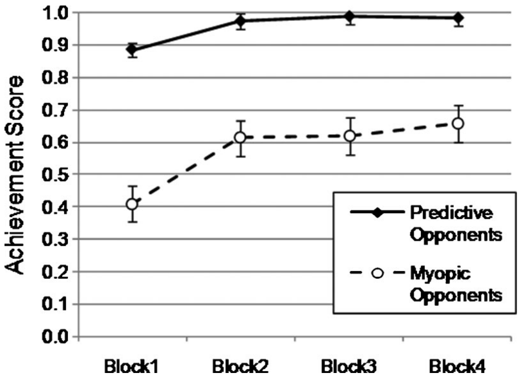


Figure 4. Achievement scores in the fixed-sum game in Experiment 2, reflecting proportion of trials that optimized outcomes.

DOI: 10.1002/bdm

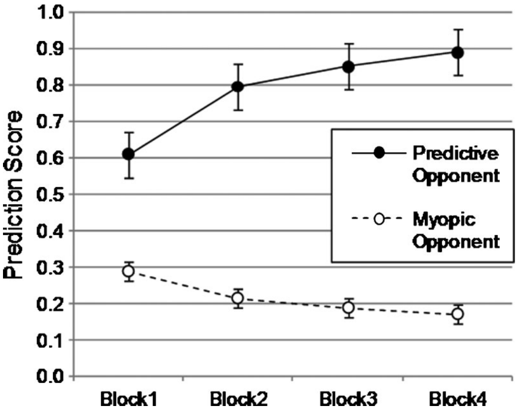


Figure 5. Prediction scores in the fixed-sum game in Experiment 2, reflecting the proportion of time participants predicted opponent would act in a predictive manner.

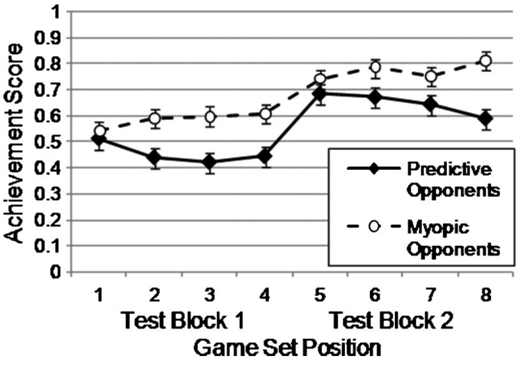


Figure 6. Achievement scores in the general-sum game in Experiment 2, reflecting proportion of trials that optimized outcomes.

between-subjects factor. Multivariate tests for the main effect of block narrowly failed to reach statistical significance (Wilks $\Lambda = .935$; $F(3,114) = 2.62, p = .054$); however, the interaction between opponent type and block suggests that the difference in block-by-block changes in mean prediction scores between Predictive and Myopic conditions is statistically significant (Wilks $\Lambda = .716$, $F(3,114) = 15.1$, $p < .001$; partial $\eta^2 = .284$). The Serlin-adjusted effect size was .265. These results suggest that changes in prediction scores were driven by the Predictive group participants' learning to use 2nd-level reasoning.

Regarding the extent to which participants believed their opponent was human, the same questions were asked post-experimentally as had been asked in Experiment 1, and the results were similar. Average responses showed substantial believability at the beginning of the study ($M = 4.95$, $SD = 1.73$), which declined by the end of the study ($M = 2.75$, $SD = 2.02$). Between-group differences were not significant.

Our interpretation of the fixed-sum results is consistent with that from Experiment 1: Individuals displayed default 2nd-level reasoning and, when playing against a Predictive opponent, quickly achieved and sustained near-total responding consistent with 2nd-level reasoning. Participants who played against a myopic opponent learned slowly and incompletely to respond optimally, reminiscent of prior results (Hedden & Zhang, 2002) that had been observed with a *predictive* opponent.

General-sum game

We computed achievement scores in the same manner as in the fixed-sum game, and the results are depicted in Figure 6. Learning took place, as the main effect of block was significant (Wilks $\Lambda = .402$, $F(1,112) = 166$, $p < .001$), with both groups showing increasing achievement. Also, individuals playing against a myopic opponent had higher overall achievement scores than those playing against a predictive opponent (.679 versus .549) ($F(1,112) = 27.564$, $p < .001$; partial $\eta^2 = .198$).

Prediction score data for the general-sum game are presented in Figure 7. They show that participants had a default expectation that their opponent would act in a manner consistent with 0th-level reasoning, and the participants thus engaged in 1st-level reasoning. While the mean prediction scores over all critical trials for participants facing predictive opponents (.367, SE = .030) were significantly greater than for those facing myopic opponents (.235, SE = .029; $F(1,112) = 10.215$, $p < .01$; partial $\eta^2 = .084$), the scores are generally low. Learning took place that was responsive to the opponent, as those with a Predictive opponent showed an increase in prediction scores, whereas those with a Myopic opponent showed a slight decrease across time.

We hypothesized that prediction reaction times (RT) should be longer when engaging in 2nd-level reasoning than when engaging in 1st-level reasoning. Group-level data are shown in Figure 8 and show that participants spent more time making 2nd-level predictions than 1st-level predictions ($t(102) = 5.02$, $p < .001$), although they did not take more time to make choices consistent with 2nd-level reasoning ($t(102) = 0.606$, $p = .546$). Figure 9 depicts the RT effect in the prediction phase at an individual level. Each participant is represented by a data point, with average prediction RT when engaging in 2nd-level reasoning on the x-axis, and average prediction RT when engaging in 1st-level reasoning on the y-axis. There were 57 participants out of 105 who spent more than 1 second longer on 2nd-level reasoning. Only 30 participants who engaged in both levels of reasoning had

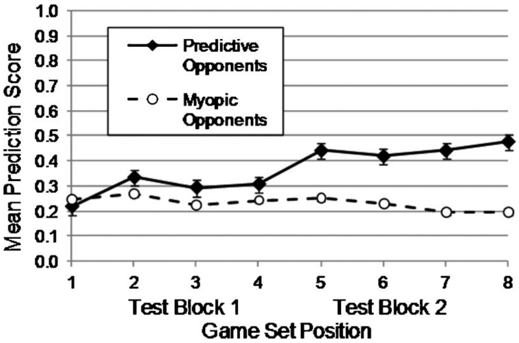


Figure 7. Prediction scores in the general-sum game in Experiment 2, reflecting the proportion of time participants predicted opponent would act in a predictive manner.

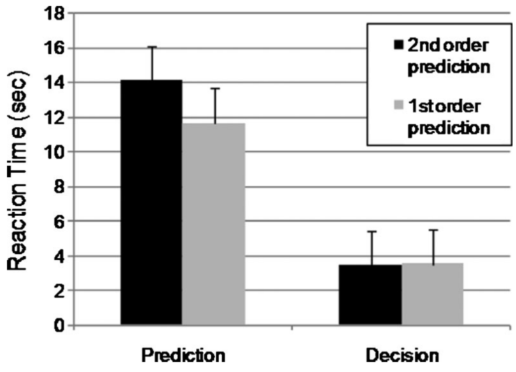


Figure 8. Group-level reaction times (RTs) in the general-sum game in Experiment 2.

an average prediction RT that was higher when engaging in 1st-level reasoning, of which 14 spent more than 1 second longer on 1st-level reasoning (for both comparisons, $p < .001$ by a binomial test).[2] Interestingly, several participants in the Myopic condition took considerably longer than others to make 2nd-level decisions, which has no parallel among participants in the Predictive condition. We speculate that this results to a large extent from the conflict between the observed default of 2nd-level reasoning and the reinforced behavior consistent with 1st-level reasoning.

For believability questions, average responses again showed substantial believability at the beginning of the study ($M = 5.22$, $SD = 1.93$), which declined by the end of the study ($M = 3.55$, $SD = 2.20$). Between-group differences were not significant.

Average levels of rationality errors were .351 in Block 1 and .278 in Block 2. These values are similar to those observed previously, including an absence of overall difference in error rates between groups ($t(112) = 1.52$, $p = .131$). Mixed model analyses for rationality error rates indicate that the interaction between opponent type and set position was not significant (Wilks $\Lambda = .904$, $F(7,106) = 1.61$, $p = .141$).

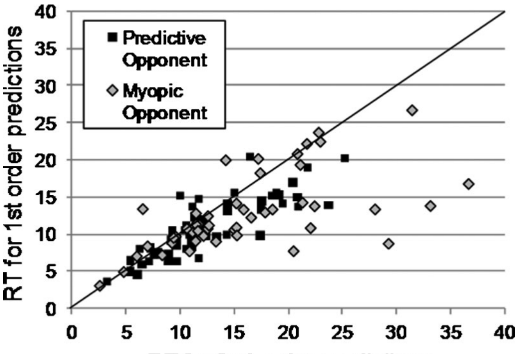


Figure 9. Individual-level reaction times (RTs) in the general-sum game in Experiment 2.

In all regards, the results we observed with the general-sum game are consistent with those of Hedden and Zhang (2002). In light of the relatively brief reaction times at the decision phase, as well as the absence of significant differences at that phase, we speculate that some of the cognitive processing related to decision making, including that which would be more complex for 2nd-level reasoning than for 1st-level reasoning, may have taken place as part of a unified process that led to both predictions of the opponent's action and a decision about the participant's own action.

Comparing fixed- and general-sum game performance

Results comparing fixed- with general-sum game achievement scores are shown in Figure 10. In the first five trials, these reflect significantly better performance in the fixed sum game when playing against a predictive opponent, but *worse* performance against a myopic opponent. This effect is reflected in a significant interaction ($F(1,228) = 36.9$, $p < .001$; partial $\eta^2 = .139$). Overall achievement scores are depicted in Figure 10b. The fixed-sum game succeeded in yielding largely correct predictions of a 1st-level opponent's responding, with .958 optimal performance. It appears that performance was better against a predictive opponent in the fixed-sum game and better against a myopic opponent in the general-sum game, and this interaction between opponent type and game type is significant ($F(1,228) = 110.7$, $p < .001$; partial $\eta^2 = .327$).

Results comparing fixed- with general-sum game prediction scores are shown in Figure 11. In the earliest trials, these reflect significantly higher default scores, defined as

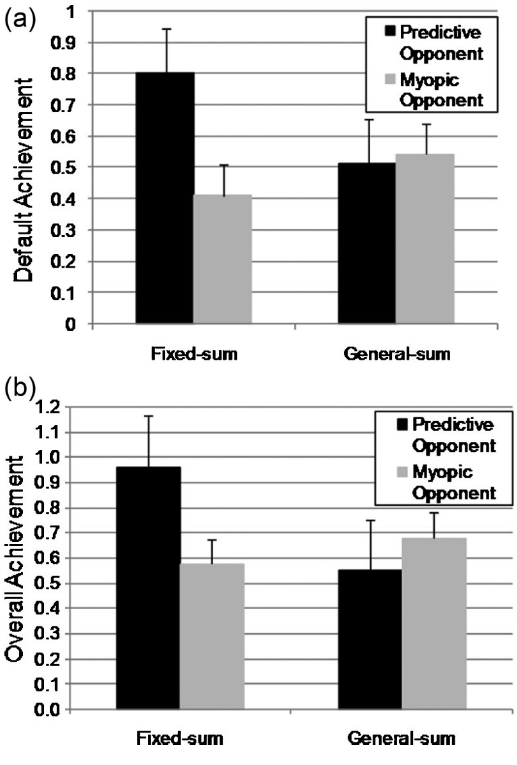


Figure 10. Comparisons between fixed- and general-sum games in Experiment 2. Achievement scores, reflecting proportion of trials that optimized outcomes: (a) default scores and (b) overall scores.

---

[2] Participants who engaged in only one level of reasoning are excluded from these analyses. This was typically the case in the fixed-sum game with a Predictive opponent, and because of this, reaction time analyses are not presented for the fixed-sum game.

DOI: 10.1002/bdm

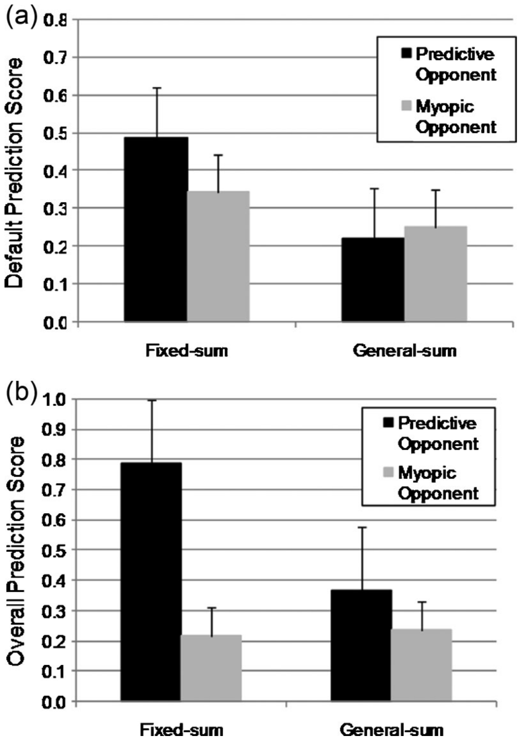


Figure 11. Comparisons between fixed- and general-sum games in Experiment 2. Prediction scores, reflecting the proportion of time participants predicted opponent would act in a predictive manner: (a) default scores and (b) overall scores.

performance in the first four trials, in the fixed-sum game (.417, SE = .035) than the general-sum game (.235, SE = .025; $t(230) = 4.249$, $p < .001$). Overall, the fixed-sum game succeeded in yielding accurate predictions of a 1st-level opponent's responding (shown in Figure 11b), with fewer errors (21%) than were observed against a myopic opponent in the general-sum game (24%). Those playing against a predictive opponent ended with prediction scores that reflected primarily 2nd-level reasoning. This is reflected in a significant interaction between game and opponent ($F(1,228) = 45.2$, $p < .001$; partial $\eta^2 = .166$).

The possible rote use of backward induction

Midway through data collection, we became concerned about the possibility that participants might rotely apply learned rules such as minimax or backward induction rather than reason through what Player II would think. Consequently, we began administering a post-experimental questionnaire which the last 175 participants answered. The questions included:

(1) *Did you use backward induction or a minimax strategy?*
(2) *Do you know what backward induction is? If yes, please describe it briefly.*
(3) *Do you know what a minimax strategy is? If yes, please describe it briefly.*

Two independent raters assessed responses to these questions. For the first question, responses were grouped into three categories: Backward induction, minimax, or neither. For the second and third questions, raters formed binary assessments of whether the responses constituted claims of knowledge or not. The raters achieved high inter-rater reliability with $K = .90$. Out of the 175 responses, 130 (74.3%) responded in the negative to all three questions in the judgment of both raters. An additional 13 participants responded to the first question with either "backward induction" or "minimax" and subsequently indicated that they did not know what their endorsed strategy was. (It is possible they interpreted the first question as requiring a response of *backward induction* or *minimax*, and not permitting a response of *no*.) Thus 143 out of 175 polled participants (81.7%) either gave completely negative responses or indicated use of one strategy without being able to explain what that strategy was.

When performance was optimal, the participant may have engaged in backward induction or its equivalent, but it is instructive to consider how she arrived at such a strategy. If a participant has been formally trained in game theory and its methods, then the use of backward induction may reflect rote reinforcement learning rather than high levels of TOM reasoning. Rote reinforcement learning would most likely be accompanied by knowing the formal name of the strategy.

On the other hand, it is possible that participants might devise backward induction spontaneously, without knowing its formal name. Participants have been observed to engage spontaneously in backward induction (e.g., Erev & Rapoport, 1990). The assumption of mutual knowledge of rationality that is required to devise backward induction spontaneously involves reasoning at high levels, at least to the level that is required to solve a particular problem. Thus, if a participant is found to engage in backward induction or its equivalent, unless she is repeating a reinforced behavior from formal training, then she is exhibiting at least the level of TOM reasoning that is required to solve the problem.

EXPERIMENT 3: THIRD-LEVEL REASONING

Because 2nd-level reasoning was observed so pervasively in simple, competitive games in Experiments 1 and 2, both by default and by means of rapid learning, in Experiment 3 we sought to discover whether we could also observe 3rd-level reasoning. Would participants act, either by default or through learning, as if they expected their opponent to utilize 2nd-level reasoning? The extended game is shown in Figure 12, in which, compared to the game depicted in Figure 1, there are four stages rather than three. We continue to refer to an opponent with 1st-level reasoning as "predictive" and with 0th-level reasoning as "myopic," and now add the term "superpredictive" to refer to 2nd-level reasoning in the opponent. The game depicted in Figure 12b, which has the outcomes ordered 3-2-4-5-1, is the only ordering of 1-5 that permits 3rd-level reasoning to be distinguished behaviorally from 2nd-level reasoning. If Player II is a 1st-level reasoner and believes that Player I can be fooled at stage C into moving to D, then Player II can exploit this by moving from B to C. If Player I believes Player II to be a 1st-level reasoner, she can in turn exploit this

DOI: 10.1002/bdm

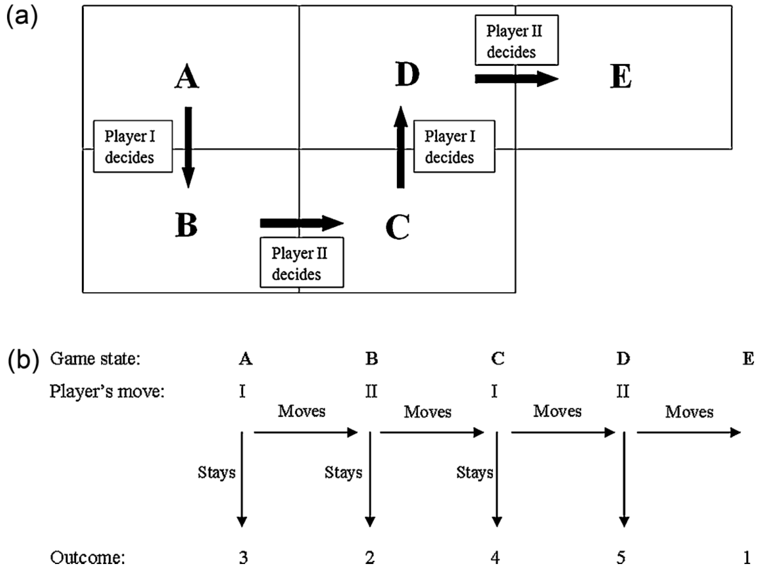


Figure 12. A four-stage fixed-sum sequential game in (a) matrix and (b) tree format.

by moving from A to B. However, if Player I believes Player II to be a 2nd-level reasoner, who would stay at B, then Player I would stay at A.

The 3-2-4-5-1 ordering does not distinguish behaviorally between 3rd- and 1st-level reasoning, as either 3rd- or 1st-level reasoners would stay at A. In order to distinguish between 1st- and 3rd-level reasoning, games with the structures 3-2-1-4-5 and 3-2-1-5-4 were also used, in which a 1st-level reasoner would move, but either a 2nd- or 3rd-level reasoner would stay at A.

In Experiment 3, participants played 30 trials, each comprising one 3-2-4-5-1 game, plus one game of either 3-2-1-4-5 or 3-2-1-5-4, presented in random order. Trials were separated by one "Catch" game of either 2-5-4-1-3 or 2-5-3-4-1. In Catch games, a reasoner of any level would move at A, which allowed us to ensure that participants were not rotely staying on every trial.

Choices on the first two games of each trial could be categorized according to the level of reasoning with which they were consistent. Staying at A in both games would be consistent with 3rd-level reasoning. Moving from A in the 3-2-4-5-1 game but staying at A in the other game (whether 3-2-1-4-5 or 3-2-1-5-4) would be consistent with 2nd-level reasoning; and staying at A in the 3-2-4-5-1 game but moving at A in the other game would be consistent with 1st-level reasoning. Moving on both trials is not consistent with any level of reasoning and is thus labeled "Chaotic."

Methods

Participants

We recruited 66 (31 female) participants who met basic criteria of learning the rules of the game. Three individuals failed to meet the basic learning criteria. All participants were recruited from the same population as those in the other experiments, and were compensated $0.50 for each game they won.

Trials

The first 25 trials comprised a training phase that did not allow participants to learn the reasoning level of their opponent. The test phase comprised 30 trials, each consisting of two games, with trials separated by 30 Catch games. The 30 trials are grouped into six blocks of five trials' length. Each trial consisted of a 3-2-4-5-1 trial in either the first or second serial position, and either a 3-2-1-4-5 or 3-2-1-5-4 trial in either the second or first serial position. Catch games consisted of either 2-5-4-1-3 or 2-5-3-4-1 types. As in the other experiments, the specific probabilities that were used were in the interval [.1,.9] in increments of .05. For any game, the difference between any probability and the next highest probability was 0.15, 0.2, or 0.25. There are fewer than 115 combinations of five probabilities meeting these criteria, making game-unique combinations impossible. In constructing games based on duplicated combinations of probabilities, we ensured that one instance was in the training phase, and the other was in the test phase.

The level of reasoning of the opponent was manipulated between-subjects, with participants assigned randomly to face a Myopic, Predictive or Superpredictive opponent.

Results and discussion

We define achievement score for this study as the proportion of trials on which *both* games were played optimally, given the opponent's level of reasoning. Participants competing against superpredictive opponents had the highest overall achievement, followed by those competing against myopic opponents. Those competing against predictive opponents never truly achieved an appropriate strategy. These differences in overall achievement were statistically significant ($F(2,63) = 110.557$, $p < .01$); all pairwise comparisons between conditions' marginal mean achievement scores were likewise statistically significant. Achievement scores across blocks are depicted in Figure 13.

DOI: 10.1002/bdm

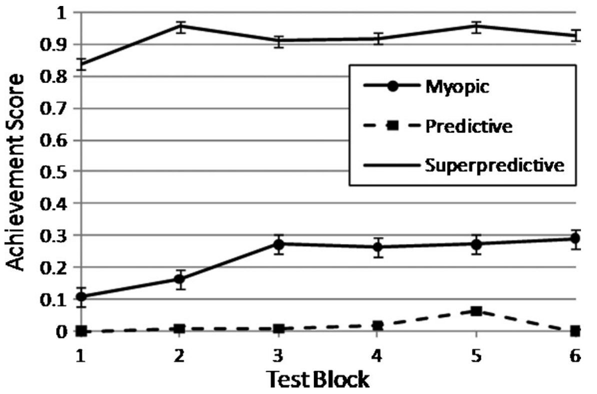


Figure 13. Achievement scores for Myopic, Predictive, and Superpredictive conditions in Experiment 3, reflecting proportion of trials that optimized outcomes.

Because in this more complex setting there were four rather than two behavioral strategies that could be distinguished, it is necessary to analyze the patterns of choice among the four strategies, beyond the correct-incorrect dichotomy that the achievement score reflects. Figure 14 presents these results with four panels that reflect proportions of choices consistent with participants' using 1st-level, 2nd-level, 3rd-level, and chaotic reasoning, respectively. All groups acted consistent with 3rd-level reasoning on most trials in the initial block. Those for whom this was optimal, because their opponent was superpredictive, increased their rates of acting in accordance with 3rd-level reasoning (Wilks $\Lambda = .831$, $F(5,59) = 2.41$, $p < .05$; partial $\eta^2 = .169$).

Those with a myopic opponent learned over time to respond accordingly more often (Wilks $\Lambda = .800$, $F(5,59) = 2.95$, $p < .05$; partial $\eta^2 = .200$) such that, starting with the second block, they responded consistent with 1st-level reasoning more than other groups (Myopic = .253, Predictive = .025, Superpredictive = .035; $F(2,53) = 6.59$, $p < .01$) and consistent with 3rd-level reasoning *less* than other groups (Myopic = .715, Predictive = .951, Superpredictive = .933; $F(2,53) = 5.75$, $p < .01$).

Participants in the Predictive group, whose opponents had an intermediate level of reasoning, not as low as the myopic opponent but not fully mutually rational as the superpredictive opponent was, showed an intriguing and consistent pattern of responding as if their opponent was superpredictive, to their detriment. Finally, it can be seen that responding consistent with chaotic reasoning was relatively rare in all conditions.

For believability questions, average responses again showed substantial believability at the beginning of the study ($M = 5.33$, $SD = 1.44$), which declined by the end of the study ($M = 3.32$, $SD = 1.99$). Between-group differences were not significant.

Rationality error rates diminished over blocks of trials (Wilks $\Lambda = .740$, $F(5,59) = 4.135$, $p < .01$) and were con-

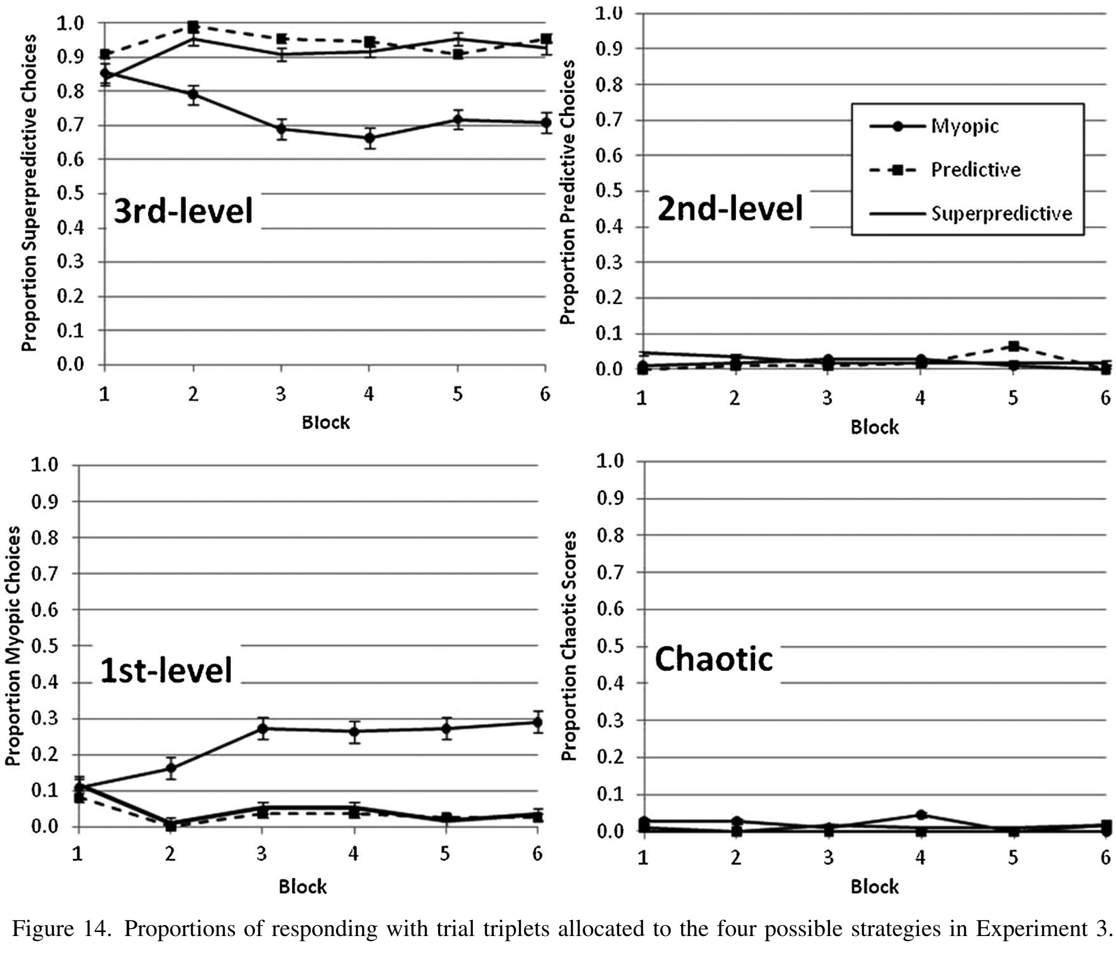


Figure 14. Proportions of responding with trial triplets allocated to the four possible strategies in Experiment 3.

sistent with the error rates observed in Experiment 2, with averages of .158 in Block 1, .111 in Block 2, .112 in Block 3, .091 in Block 4, .080 in Block 5 and .064 in Block 6. Finally, Catch game move rates for all conditions were high (Myopic = .947, Predictive = .970, Superpredictive = .914) and not significantly different from one another, suggesting that participants on the whole attended to the task.

In sum, participants showed a robust tendency by default to engage in 3rd-level reasoning when given the opportunity. Those for whom this was optimal quickly rose to even higher rates of 3rd-level reasoning. Those for whom 1st-level reasoning was optimal learned slowly and incompletely to do so; and those for whom 2nd-level reasoning was optimal showed little evidence of learning, instead retaining high rates of 3rd-level reasoning despite its inconsistent success. It is possible that participants categorized their opponents in a binary fashion, as either following or not-following mutually rational solutions. In all conditions in Experiments 1 and 2, opponents were consistently mutually rational (predictive) or consistently not so (myopic). The same applies to superpredictive and myopic opponents in Experiment 3. Opponents who were predictive, however, acted in a mutually rational manner on some trials (3-2-1-4-5 and 3-2-1-5-4) but not others (3-2-4-5-1). Participants with predictive opponents in Experiment 3 may have been confused by the failure to fall into a pattern of mutual rationality or its opposite.

EXPERIMENT 4: CONTROLLING FOR UNCERTAINTY AVOIDANCE AND DEFAULT STAYING

In the prior experiments, the response that reflected the highest-level reasoning at position A always consisted of staying. The only ordering in the three-stage fixed-sum game that distinguished levels of responding was 3-2-1-4; the only distinguishing ordering in the four-stage game was 3-2-4-5-1; and highest-level reasoning dictated staying in these games. This gives rise to two plausible alternative explanations, which were controlled for in Experiment 4. First, in the previous experiments, moving at A led to the uncertainty involved about the other player making a choice at B, whereas staying at A led to a certain outcome. Thus the tendency to stay at A could in part have been due to uncertainty avoidance. Second, it is possible that staying at A appeared to participants as a default strategy, to be enacted in the event they were uncertain of the choice.

In order to rule out these possibilities, in Experiment 4 participants chose between games, as shown in Figure 15. Because participants were required to choose one course of action or the other, neither choice could be construed as the default; and because both options led to a game in which their opponent made the first choice, there was equal uncertainty involved in both. If the participant believes the opponent to be predictive, then she will choose the game on the left to obtain 3; but, if she believes the opponent to be myopic, then she will choose the game on the right to obtain 5.

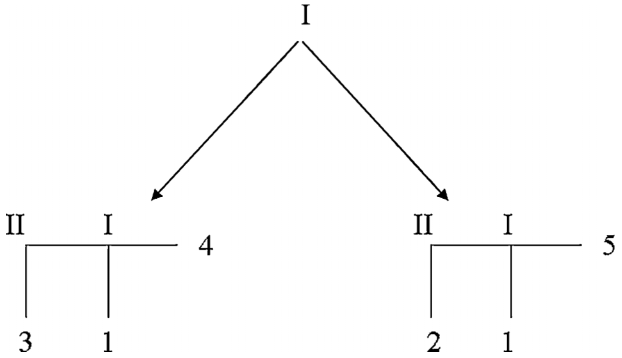


Figure 15. The design used in Experiment 4, wherein the participant (Player I) chooses which of two 2-stage games to play.

Methods

Participants

We recruited 22 (15 female) participants from the same population as those in the other experiments, and there was no external backing for correct choices.

Trials

Participants made 40 choices, which all had the structure depicted in Figure 15, and the 40 combinations of probabilities that had been used in Experiment 3. Options were counterbalanced by side. Up to four participants were studied at a time, sitting around a large table and filling out paper forms with an experimenter present. Participants did not learn the choices made by Player II, but simply expressed their choice on each game before moving on to the next game. Hence the results bear only on default reasoning and not on learning.

Results and discussion

On average, participants responded to 27 (67.4%) of the 40 problems in a manner consistent with 2nd-level reasoning rather than 1st-level reasoning. This is significantly greater than the chance level of 50% ($t(21) = 2.52$, $p < .05$). The individual-level proportions are given in the histogram in Figure 16. Most participants' choices were consistent with 2nd-level reasoning in a context where default responding and risk avoidance were eliminated as possible explanations.

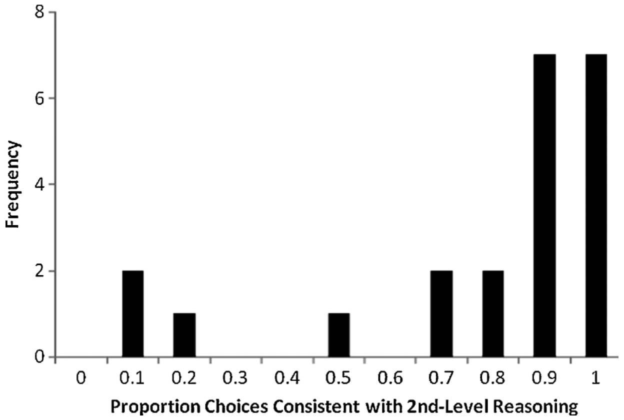


Figure 16. Histogram of responding in Experiment 4. Most participants' choices were consistent with 2nd-level reasoning.

GENERAL DISCUSSION

The prior literature on levels of reasoning in adult decision making was pessimistic, with results suggesting that reasoners attribute to others preferences but little reasoning, and particularly little reasoning about the reasoner's own pursuit of his or her interests. When playing with partners who reasoned at the first level, participants slowly and incompletely learned to both predict their partners' responses accurately and respond in more adaptive ways to these anticipated responses.

In four experiments, we replicated previous results and developed a new series of games that were simpler and more competitive by virtue of utilizing a fixed-sum rather than general-sum payoff structure. In this game, reasoners made default predictions that their opponents would reason about the reasoner's pursuit of their interests and acted accordingly. Those whose opponents matched these higher-level default assumptions quickly learned to predict and respond almost optimally. In these games, contrary to the prior literature, it was those whose opponents engaged in lower-level reasoning who performed less well initially, and learned slowly and incompletely to predict and respond optimally. Interestingly, participants engaged readily in 2nd-level reasoning when it was the highest level available (in Experiments 1 and 2), but demonstrated a notable shortage of 2nd-level reasoning when an even higher level of reasoning was available (Experiment 3). In sum, participants most readily engaged in the highest available level of reasoning that was available. In Experiment 4 we controlled for possible confounds of risk avoidance and default responding and continued to observe choices consistent with higher-level reasoning.

Individuals may possess higher-level TOM reasoning capabilities than was previously acknowledged, reflecting more realistic assumptions about their opponents in the domain of competitive and relatively simple games. These findings suggest that psychological models of recursive reasoning may deviate from normative models less than had previously been thought necessary, with implications for research in game theory and cognitive science as well as psychology.

We also note that there is an important distinction between the ability to reason at a high level, and the propensity to use it. When choices are made in concert with relatively high-order reasoning, as we observed in the fixed-sum games, it can be concluded that both the ability and the propensity are present. When choices are not consistent with high-order reasoning, this can reflect either inability or insufficient motivation. It is possible that individuals are capable of better performance in the more complex games in which we and others have found lower levels or reasoning, but choose not to deploy the greater cognitive effort that would be required. Furthermore, the appearance of a poor state of affairs in the prior literature may have been at least partially attributable to difficulties in interpreting non-competitive game behavior in game-theoretic terms. For example, in the game given in Figure 2, the participant who moves at A may be trying to increase the overall reward that is available in cells C and D, perhaps in the hope of establishing cooperative sharing of individual benefits for the greater benefit of dividing a larger combined reward (e.g., Konow, 2010).

It is a limitation of the present studies that, although participants were initially credulous that they were playing against a human, by the end of the study there was decidedly mixed opinion among participants in Experiments 1-3 regarding whether they were playing against a human or a computer. Our failure to be completely convincing on this may be due to the consistency of the opponent's strategy (being always consistent with myopic, predictive or super-predictive reasoning), effects in the time that the opponent took to respond, despite our efforts to make the time appear human, or other factors. Future research may explore whether responding differs systematically against opponents who are more strongly seen to be human or artificial; we are not aware of any current literature on this issue. However, it is also possible that the pairing of believability questions created a demand characteristic that appeared to the participants to call for diminishing credulity over time, and that the measurements reflecting end-of-session belief are uncharacteristically low because of this.

Also, we cannot exclude the possibility that, in the crucial 3-2-1-4 fixed-sum game used in Experiments 1 and 2, a preference for relatively equitable outcomes (e.g., Fehr & Schmidt, 1999) may have contributed to the findings. This possibility would suggest a general preference for outcomes of 2 and 3 rather than 1 and 4, which would become possible when moving from 3 to 2. One way to test this is to examine whether participants were more prone to move away from the initial "3" outcome when it reflected a relatively inequitable outcome, in other words when it was relatively distant from .5. There was only one trial in the game where the "3" outcome was less than .4, but there were 16 trials where it was greater than .6. Participants with a predictive opponent showed a ceiling effect, almost never moving under any circumstances, so we examined only those in the Myopic condition. In fact, in the 16 trials in which the "3" outcome was greater than .6, participants in the Myopic condition were actually less likely to move than in all trials combined (.395 compared with .446 in Experiment 1, .534 compared with .576 in Experiment 2), albeit to a non-significant degree. Thus, for the predictive case, when the "3" outcome corresponded to probabilities distant from 0.5, participants continued to stay at A, further suggesting the absence of the effect of inequity aversion.

Competition or simplicity?

The fixed-sum game was both simpler and more competitive than the general-sum game: Simpler because when outcomes in each cell are complementary rather than independent, the number of values defining the game declines from eight to four; competitive because the defining distinction between fixed- and general-sum games is that in the former, the players vie for shares of a limited resource.

Because competition and simplicity are confounded in this manipulation, we cannot reach firm conclusions about which may be more or less responsible for the increased level of reasoning that was observed. We speculate, however, that the competitive aspect may have been a relatively salient contributor to the effect, in comparison with the simplicity, based on two observations. First, in the context of the game-theoretic literature, even the more complex general-sum centipede game is a simple game. Second, we note that achievement scores against a myopic opponent in the fixed-sum game of Experiment 2 were notably low. If superior performance arose straightforwardly from using a simpler game, then greater learning should be expected to mitigate this error rate.

The 2nd- and 3rd-level reasoning observed strongly in these studies has implications for models of the many interactive games that have been studied. The conclusion that participants reach an average of 1.5 steps in their reasoning across many games including coordination games, market entry games, beauty contest games and others (Camerer, Ho, & Chong, 2004) may be appropriately moderated according to the simplicity and degree of competition involved in the game. Tor and Bazerman (2003) found that performance on a diverse set of interactive games was correlated with attention to aspects relating to other parties and their interaction with the rules. It may be that aspects of the fixed-sum game tend to focus attention on the other player. One possible mechanism by which this would happen is an increase in competition motivation (Garcia, Tor, & Gonzalez, 2006; Garcia & Tor, 2009), which can increase the amount of cognitive effort that is devoted to a task. Another possible mechanism for the comparatively poor results in a general-sum context is erroneously reasoning that, if the other player's outcomes are independent of one's own outcomes, then one's reasoning can be independent of the other player's.

Conclusions

The prior literature on adult recursive reasoning was pessimistic, suggesting that individuals reasoned at a low level, taking into account the desires of others but not their ability to reason strategically. We introduced a class of games that were both more competitive and simpler than had been used in the prior literature. Individuals were found to reason at a higher level by default, and to be quicker to learn against a higher-level reasoning partner under these circumstances. This finding suggests that individuals may not always systematically underestimate their opponents in strategic environments.

ACKNOWLEDGEMENTS

This research was supported by AFOSR research grant FA9550-08-1-0429 to PD and ASG.

REFERENCES

Bornstein, G., Gneezy, U., & Nagel, R. (2002). The effect of intergroup competition on group coordination: An experimental study. *Games and Economic Behavior*, *41*, 1–25.

Camerer, C., Ho, T.-H., & Chong, J.-K. (2004). A cognitive hierarchy model of games. *Quarterly Journal of Economics*, *119*, 861–898.

Case, D. A., Fantino, E., & Goodie, A. S. (1999). Base-rate training without case cues reduces base-rate neglect. *Psychonomic Bulletin and Review*, *6*, 319–327.

Cosmides, L. (1989). The logic of social exchange: Has natural selection shaped how humans reason? Studies with the Wason selection task. *Cognition*, *31*, 187–276.

Crawford, V. P. (2003). Lying for strategic advantage: Rational and boundedly rational misrepresentation of intentions. *American Economic Review*, *93*, 133–149.

Crawford, V. P., & Iriberri, N. (2007). Level-*k* auctions: Can a nonequilibrium model of strategic thinking explain the winner's curse and overbidding in private-value auctions? *Econometrica*, *75*, 1721–1770.

Erev, I., & Rapoport, A. (1990). Provision of step-level public goods: The sequential contribution mechanism. *Journal of Conflict Resolution*, *34*, 401–425.

Fehr, E., & Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation. *The Quarterly Journal of Economics*, *114*, 817–868.

Flavell, J. H., Mumme, D. L., Green, F. L., & Flavell, E. R. (1992). Young children's understanding of moral and other beliefs. *Child Development*, *63*, 960–977.

Flavell, J. H., Green, F. L., & Flavell, E. R. (1998). The mind has a mind of its own: Developing knowledge about mental uncontrollability. *Cognitive Development*, *13*, 127–138.

Flavell, J. H., Green, F. L., & Flavell, E. R. (2000). Development of children's awareness of their own thoughts. *Journal of Cognition and Development*, *70*, 396–412.

Garcia, S. M., & Tor, A. (2009). The *N*-effect: More competitors, less competition. *Psychological Science*, *20*, 871–877.

Garcia, S. M., Tor, A., & Gonzalez, R. (2006). Ranks and rivals: A theory of competition. *Personality and Social Psychology Bulletin*, *32*, 970–982.

Gigerenzer, G., & Hug, K. (1992). Domain specific reasoning: Social contracts, cheating and perspective change. *Cognition*, *43*, 127–171.

Goodie, A. S., & Fantino, E. (1995). An experientially derived base-rate error in humans. *Psychological Science*, *6*, 101–106.

Goodie, A. S., & Fantino, E. (1996). Learning to commit or avoid the base-rate error. *Nature*, *380*, 247–249.

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

Hedden, T., & Zhang, J. (2002). What do you think I think you think?: Strategic reasoning in matrix games. *Cognition*, *85*, 1–36.

Ho, T.-H., Camerer, C., & Weigelt, K. (1998). Iterated dominance and iterated best response in experimental *p*-beauty contests. *American Economic Review*, *88*, 947–969.

Johnson, D. W., Maruyama, G., Johnson, R., Nelson, D., & Skon, L. (1981). Effects of cooperative, competitive, and individualistic goal structures on achievement: A meta-analysis. *Psychological Bulletin*, *89*, 47–62.

Konow, J. (2010). Mixed feelings: Theories of and evidence on giving. *Journal of Public Economics*, *94*, 279–297.

Lieberman, D. A. (1997). Interactive video games for health promotion: Effects on knowledge. In R.L. Street , W.R. Gold , & T.R. Manning , *Health promotion and interactive technology.* Hillsdale, NJ: Lawrence Erlbaum.

Nickell, S. J. (1996). Competition and corporate performance. *The Journal of Political Economy*, *104*, 724–746.

Perner, J., & Wimmer, H. (1985). "John thinks that Mary thinks that" Attribution of second-order beliefs by 5- to 10-year-old children. *Journal of Experimental Child Psychology*, *39*, 437–471.

DOI: 10.1002/bdm

Rapoport, A., & Budescu, D. V. (1992). Generation of random series in two-person strictly competitive games. *Journal of Experimental Psychology: General*, *121*, 352–363.

Rindova, V. P., Becerra, M., & Contardo, I. (2004). Enacting competitive wars: Competitive activity, language games, and market consequences. *Academy of Management Review*, *29*, 670–686.

Stahl, D., & Wilson, P. (1994). Experimental evidence on players' models of other players. *Journal of Economic Behavior and Organization*, *25*, 309–327.

Stahl, D., & Wilson, P. (1995). On players' models of other players: Theory and experimental evidence. *Games and Economic Behavior*, *10*, 218–254.

Tor, A., & Bazerman, M. H. (2003). Focusing failures in competitive environments: Explaining decision errors in the Monty Hall game, the acquiring a company game, and multiparty ultimatums. *Journal of Behavioral Decision Making*, *16*, 353–374.

Wellman, H. M., & Banerjee, M. (1991). Mind and emotion: Children's understanding of the emotional consequences of beliefs and desires. *British Journal of Developmental Psychology*, *9*, 191–214.

Wellman, H. M., & Gelman, S. A. (1998). Knowledge acquisition in foundational domains. In D. Kuhn , & R. S. Siegler (Eds.), *The handbook of child psychology: Cognition, perception, and language* (Vol. 2, pp. 523–538). New York: John Wiley and Sons.

Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-analysis of theory-of-mind development: The truth about false belief. *Child Development*, *72*, 655–684.

Authors' biographies:

Adam S. Goodie is an Associate Professor of Psychology at the University of Georgia and Director of the Georgia Decision Lab (http://psychology.uga.edu/gdl/index.html). His research focuses on judgment and decision-making. Doctoral training was at the University of California, San Diego, and postdoctoral training at Max Planck Institutes in Munich and Berlin.

Prashant Doshi is an Associate Professor of Computer Science at the University of Georgia and director of the THINC lab (http://thinc.cs.uga.edu). His research focuses on multiagent decision making, behavioral game theory and its computational modeling. His doctoral training was at the University of Illinois at Chicago.

Diana L. Young is an Assistant Professor of Psychology at Georgia College & State University and Director of the Georgia College Decision Research Lab. Her research focuses on judgment and decision-making. Doctoral training was at the University of Georgia.

Authors' addresses:

Adam S. Goodie, Department of Psychology, University of Georgia, Athens, GA 30602-3013, USA.

Prashant Doshi, 539 Boyd Graduate Studies Research Center, Department of Computer Science, The University of Georgia, Athens, GA 30602.

Diana L. Young, PhD, Department of Psychological Science, Georgia College & State University, Milledgeville, GA 31061, USA.