Does seeing deeper into a game actually increase ones chances of winning?

Carl Nicholas McKinney Jr. and John Van Huyck

January 2005

Abstract: This paper investigates whether the variation in estimated rationality bounds is correlated with the probability of winning when playing against another person in games that exceed both players estimated rationality bound. Does seeing deeper into a game matter when neither player can see to the end of the game? Subjects with higher estimated bounds win 63 percent of the time and the larger the difference the more frequently they win.

Key Words: bounded rationality, perfect information, Nim, human behavior, experiment.

JEL Classification: c72, c92, d82

Acknowledgments: Chris Ball provided able research assistance. Patrick Williams and Derek LeBlanc helped with software development. Jana Joswiak helped with recruiting and accounting. Private Enterprise Research Center and the National Science Foundation provided financial support. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the Private Enterprise Research Center or the National Science Foundation.

© 2000-2005 by the authors. All rights reserved.

I. Introduction

Perfect information games (PI-games) can be solved using a backward induction algorithm. Solve the terminal decisions. Truncate the game by replacing the terminal decision nodes with the solution. Repeat until one comes to the initial decision node or root of the game.

Selten (1978) observed that while people could be persuaded by the induction algorithm in simple games people rejected it when thinking about more complex games. The Chain Store Paradox is the observation that even when the induction argument is understood it does not provide a persuasive model of how to play sufficiently complex perfect information games. He proposed a model of bounded rationality to resolve the paradox. Because the theory of substantive rationality does not take account of players' cognitive capacity it makes inaccurate predictions and unpersuasive prescriptions.[1]

One measure of the complexity of a perfect information game is rank. A game's rank is the longest play path from the root of the extensive form to a terminal node. Johnson, Camerer, Sen, and Raymon (2002) demonstrate directly that untrained people with high analytical ability, college students, don't acquire the information they need to find the induction solution in rank 6 alternating offer bargaining games.

McKinney and Van Huyck (2004) measure bounded rationality using a class of strictly competitive perfect information games with two outcomes. All of the games had a substantively rational way to play. The value of the game was, from the perspective of the first mover, either a "win" or a "loss". Subjects played against a procedurally rational algorithm. The average subject had the ability to win winnable games up to rank 6. Beyond this capacity performance quickly deteriorated. This bound is much less than the complexity of most economic decisions.

While the theory of substantive rationality does not account for variation in peoples ability to reason about PI-games, McKinney and Van Huyck (2004) found statistically significant variation both within and across subject pools. The highest estimated bound they report is 14. The economic significance of being able to see a few moves deeper into a game more complex than that bound is not obvious.

This paper investigates whether the variation in estimated rationality

[1] See for example Binmore, McCarthy, Ponti, Samuelson, and Shaked (2002). Whether backward induction should be a requirement of substantive rationality is controversial, see Aumann (1996) and Binmore (1996).

bounds is correlated with the probability of winning when playing against another person in games that exceed both players estimated rationality bound, that is, does seeing deeper into a game matter when neither player can see to the end of the game? Regardless of the value of the game predicted by an analysis based on substantive rationality, the observed win frequency is almost exactly 50 percent. The value of the game does not help predict the outcome at all. However, the subject who can see deeper into the game wins 63 percent of the time and the larger the difference the more frequently they win.

II. Analytical Framework

To focus the analysis consider the game of Nim. Nim is a two player game in which the players alternate taking one or more stones from one of m rows. The player who takes the last stone wins. A game of Nim can be summarized by a $1 \times m$ vector of natural numbers, g, with elements g_i denoting the number of stones in row i. Let G denote the set of all Nim games, where $G = \{g \in \mathbb{N}^m \mid m \in \mathbb{N}\}$ and $\mathbb{N}$ denotes the natural numbers. The rank of Nim game g, $r(g)$, is equal to the number of stones used in the game:

$$r(g) = \sum_{i=1}^{m} g_i$$

Since the rules of Nim allow one to vary the complexity of the game as measured by rank, it is ideally suited to the study of bounded rationaltiy.

Consider using rank as a complexity measure. Intuitively, the longer the play paths the more difficult it is to think through. By the rank measure, game (5,3,4,5), which has rank 17, is more complex than game (1,1,1), which has rank 3. While the rank provides a useful and general measure of complexity, it may miss essential elements limiting human rationality. A game with a large rank may be easy to solve if it has a short winning play path.

An alternative measure is the complexity introduced by the shortest play path through the extensive form. The shortest play path would predict that (1,1,1) with a shortest play path of 3 was more complex than (0,0,3) with a shortest play path of 1, which is counter intuitive. It is not possible to lose game (1,1,1) as the first mover. The game contains no decision nodes in which it is possible to make a choice that changes the value of the game. It is a trivial game.

Denote decision nodes that can't change the value of the game trivial decision nodes. It is easy to see that all games with $r = m$ are trivial games and it is only slightly more difficult to show that all games with $r > m$ are non-trivial games. An alternative measure of complexity, denoted NT-complexity, is obtained by counting the number of non-trivial decision nodes in the extensive form of the game. The choices made at non-trivial decision nodes matter because they can potentially change the value of the game. For example, the following games all have the same rank of 7: (1,1,1,1,1,1,1), (1,1,1,4), (1,1,5), and (7). The NT-complexity measures for the four rank 7 games are 0, 220, 124, and 32 respectively, which corresponds to our intuition about their relative complexity.

A fourth measure of complexity is the probability that a blunderer wins the game playing against a substantively rational opponent, where a blunderer is a player who chooses a uniformly random feasible action at every information set they are assigned. A blunderer wins (1,1,1,4), (1,1,5), or (7) with probability 0.143, but wins the trivial game (1,1,1,1,1,1,1) for sure.

A. Substantive and Procedural Rationality

Every Nim game $g \in G$ has a value. The value is either a win for the first mover, W, or a loss, L. Let the value function be denoted $v(g) : G \to \{W,L\}$. Not only does Nim have a value when played by substantively rational players, but it also has a procedurally rational algorithm that can be used to compute the value and recommend optimal play.

A general computable algorithm to determine the value of g must not require construction of the extensive form representation of g. There exist several algorithms for computing v(g) that are procedurally rational. Here, we describe Bouton's (1902) algorithm.

Convert the decimal representation of the natural numbers in g into the equivalent binary representation. Let $b(g)$ denote the binary representation. Let $d_j(g_i)$ denote the digit in the 2^j position of $b(g_i)$, where $d_j(g_i) \in \{0,1\}$.

The set of balanced Nim games, B, is

$$B = \{g \in G \mid \sum_{i=1}^{m} d_j(g_i) \text{ is even } \forall j\} \quad .$$

The set of unbalanced Nim games, U, is

$$U = \{g \in G \mid g \notin B\} \quad .$$

To understand why the algorithm works we review the following propositions[2]. The first mover in a balanced game cannot win on his first move, because a balanced game has stones in at least two rows and a player

[2] See Mckinney and Van Huyck (2004) for references and details.

may only remove stones from one row during a turn.

Proposition 1: Any balanced game *g* must be left unbalanced.

Proposition 2: Any unbalanced game *g* can be left balanced.

Proposition 3: The first mover in an unbalanced game of Nim wins and in a balanced game of Nim loses.

Using the three propositions above, the value function can be expressed in our notation as follows:

$$v(g) = \begin{cases} W \ if \ \ g \in U \\ L \ if \ \ g \in B \end{cases}$$

Substantively rational players win all unbalanced games and when playing against other substantively rational players lose all balanced games.

B. The Wall Model of Bounded Rationality

The plausibility of the assumption that the players behave as if they are substantively rational grows weaker as the rank of the game being analyzed increases. Substantive rationality has survived as the foundation for economic analysis primarily because of the intractability of limited foresight paradigms, see Rubinstein (1998, section 7.4). However, in the case of Nim it is possible to construct tractable models of bounded rationality.

One possibility is that players have a bound on their ability to reason deductively about PI-games. A player is able to solve games up to some rank denoted $\bar{r}_i$ but not beyond this reasoning depth measure. Specifically, if playing a game *g* such that $r(g)\ \#\ \bar{r}_i$ they use a substantively rational strategy. However, if $r(g) > \bar{r}_i$, then they select an action uniformly from the set of feasible actions available at their current information set. They continue playing as if they were a blunderer until they reach a subgame with a rank that is less than or equal to $\bar{r}_i$ after which they use a substantively rational strategy for the subgame. Denote this model of bounded rationality the *wall model*.

If the bound on both players rationality is greater than the rank of the game, then *v(g)* characterizes the predicted outcome even when played by wall model players. However, when this condition does not hold the wall model gives a very different prediction for the likelihood the first mover wins the game.

The wall model can be used to determine the probability the first mover wins a game *g* as a function of $\bar{r}_1$ and $\bar{r}_2$:

$$w(g,\ \bar{r}_1,\ \bar{r}_2)$$

Table 1 presents this probability for game (5,3,4,5), a typical unbalanced game of rank 17, for the range of walls we have observed to be typical in our subject pool. Since the game is unbalanced, the first mover should always win when played by players with reasoning depth greater than 17.

Inspection of the table reveals that the difference in players' walls has more to do with the predicted probability Player One wins. The theoretical value of the game is completely swamped by considerations of bounded reasoning depths. Even when both players have the same wall the probability Player One wins is far from the substantively rational prediction of 1. Moreover, when played by two unusually good players of equal strength, Player One can actually be disadvantaged: when $\bar{r}_1 = \bar{r}_2 = 14$ Player One wins only 46 percent of the time.

<Table 1>

The Wall Model predicts that players with deeper reasoning depths are more likely to win complicated Nim games. The size of the advantage depends on the game.

Table 2 reports Player One's win probabilities for a typical balanced game of rank 14, Game (7,4,3). Since the game is balanced, assuming substantive rationality one predicts Player One should have zero probability of winning. Again, the influence of players relative walls swamps the influence of the games structure. The exception being those columns in which Player Two's wall is high enough to see to the end of the subgame they inherit after Player One moves, that is, when $\bar{r}_2 \geq 13$.

<Table 2>

III. Experimental Design

The experiment consists of two treatments. In the first treatment, subjects play against each other. In the second treatment, subjects play against the Bouton algorithm. The second treatment essentially replicates the baseline treatment in McKinney and Van Huyck (2004), which was designed to allow one to estimate bounded rationality.[3]

In treatment 1, 20 subjects played a pair of games, one balanced and one unbalanced game against each of the other subjects in the session. The 38 Nim games had ranks of from 12 to 17, which exceeds 95 percent of the bounds previously estimated for the TAMU subject pool. The role of first or second mover was not changed during a pairing. Each participant should win the unbalanced games when they are the first mover and win the balanced games when they were the seconded mover. Subjects were paid \$0.60 for a win and \$0.10 for a loss.

The subjects were paired through a round robin algorithm. Each participant played one round against each of the other participants in the session. At the beginning of each round, subjects were assigned either the role of first mover or second mover. This assignment is based on each subjects' prior assignments. The subject who had been assigned the role of first mover fewer times was always assigned the role of first mover at the beginning of each new round. In the event of a tie, the role of first mover was assigned at random.

Figure 1 is a screen grab of the graphical user interface used in treatment one. Subjects play the game by clicking on the black circles to remove them and everything to the right. Instructions are found to the right of the game. A dialog reminding the subjects of the current mover and what just happened appears below the game grid. Records are kept at the bottom of the screen. The record grid displays both games for the current pair and all games played in previous pairings.

<Figure 1>

Treatment 2 uses the design of McKinney and Van Huyck (2004) to measure $\overline{r}_i$. The subjects played 27 Nim games against Bouton's algorithm. 18 games are unbalanced and 9 games are balanced. Subjects were always in the role of the first mover. The games range in complexity from rank 3 to rank 17 and have shortest play paths ranging from 1 to 5. Subjects earned \$0.60 for a win and \$0.10 for a loss. Figure 2 is a screen grab of the graphical user interface used in treatment two. Records are kept at the bottom of the screen. All 27 games are displayed on the record grid and the subjects choose to play the games in any order. Subjects play at their own

[3] McKinney and Van Huyck (2004) also investigated how subjects priced performance uncertainty, which is the uncertainty arising from an awareness of one's limited ability to execute a substantively rational strategy.

pace.

<Figure 2>

The experiment consisted of three sessions. It was conducted in the Economic Research Laboratory. Each session consisted of twenty participants. The subjects were recruited from the Texas A&M University student body. Sessions take about three hours and a subject could earn as much as $34.50.

IV. Experimental Results

A. Treatment 1 Results: Round Robin Nim

The value, $v(g)$, of an unbalanced Nim is a win for Player One and the value of a balanced Nim games is a loss for Player One when played by substantively rational players. Table 3 reports the number of games won and lost contingent on $v(g)$. The Fisher exact score of 0.425 reveals no statistically significant correlation between $v(g)$ and the outcome. Player One won 51 percent of the unbalanced games and 50 percent of the balanced games. The substantively rational value of the game does not help predict subject performance in the experiment.

<Table 3>

The wall model uses information about the game and the estimated walls from treatment 2 in order to predict an outcome for the match. Figure 3 graphs the games value, the probability the first mover wins predicted by the wall model, and the frequency subjects in the first mover role win the game. The graph is sorted by the predictions of the wall model. The wall model makes much more accurate predictions than substantive rationality. The predicted value of the game is not included within a 95 percent confidence interval around the observed frequency for any game. On the other had the wall model is only rejected for 3 of the 38 games, which is about what one would expect given a 95 percent confidence interval.

<Figure 3>

The aggregate figures may mask either systematic differences across games or across individuals. Table 4 presents the data by game pairs played in a subject match. Recall a pair of subjects played an unbalanced and a balanced game in the same role of either Player One or Player Two before being rematched. Substantive rationality predicts that Player One wins the first game and loses the second. Substantive rationality makes the precise prediction of 30 W/L outcomes out of the 30 observations per game pair.

Rather than 100 percent, W/L outcomes were only observed 20 percent of the time, which is actually less than one would expect if either outcome were equally likely. About 61 percent of the time the first mover either won both games or lost both games. Hence, the data also reject the hypothesis that either player is equally likely to win the game.

The final column of table 4 reports ?^{2} tests of the wall model. For the most part the wall model is rejected. Specifically, the wall model is rejected in 12 of 19 cases. Comparing the distribution of the wall model and the data suggests that the wall model overestimates the advantage a subject has contingent on having a higher estimated wall in treatment 2.

<Table 4>

The wall model consists of two parts that can be tested directly. The first is the prediction subjects make uniformly random selections from the set of feasible actions at the root of subgames with a rank greater than their estimated wall. The second is the prediction subjects behave according to the theory of substantive rationality once the rank of the subgame falls below their estimated wall.

Figure 4 is a histogram comparing actual subject behavior in all of the subgames with a rank greater than their wall with the blunderer model. There is a slightly higher tendency for subjects to make extreme choices, either take one stone or take seven, than predicted by the blunderer model. Testing the hypothesis that the empirical distribution function was derived from the theoretical distribution function gives a Kolmogorov statistic of 0.0332, which fails to reject the null hypothesis of blunderer play in the first period at the 5 percent significance level.

<Figure 4>

Conducting the same test on individual subject behavior fails to reject the null hypothesis of blunderer play in 53 of 60 cases. Of the seven rejections, 4 are rejected for removing one stone too often, 2 are rejected for removing two stones too often, and one for removing four stones too often.

The second part of the wall model can also be directly tested. When the complexity of the game falls below a subjects' wall they are predicted to do at least as well as the substantively rational outcome in the remaining subgames. If the subgame is unbalanced, they should win it with probability one. Subjects actually win the first unbalanced subgame less complex than their estimated wall 75 percent of the time. A 95 percent confidence interval is [0.73, 0.78], which does not include the substantively rational prediction. The estimated wall found in treatment 2 overstates subjects ability to solve

Nim games in treatment 1.

Table 5 reports a 3x2 continency table comparing the outcome of the game stated from the first movers perspective to each players relative estimated wall. The subject with the highest estimated wall wins over 63 percent of the round robin pairings. The Fisher Exact test gives a p-value of 0.000, which rejects the hypothesis of no correlation between relative estimated walls and outcome.

<Table 5>

We searched for the best empirical model for predicting the outcome of the game. Including the difference between the first and second movers' estimated walls swamped the influence of *r(g),* round, *v(g),* and *w(g).* The empirical model is reported in table 6. Constraining the full model to conform with either the rational model or the wall model does lead to a significant worsening of the likelihood. Eliminating the constant, *r(g),* round, *v(g),* and *w(g)* does not worsen the likelihood significantly. The best empirical model only depends on the difference between the subjects estimated walls.

<Table 6>

A one rank advantage for a subject increases his probability of winning by 4.9 percent. This estimated advantage is only about one third of the marginal advantage predicted by the wall model.

B. Treatment 2 Results: Estimating Bounded Rationality

In the second treatment the participants played 27 games against a substantively rational algorithm. 18 of the games were unbalanced, that is, winnable, and 9 of the games were balanced, that is, unwinnable. Subject performance lies between blunderer and substantively rational play. Subject performance is statistically different from rational play for all games with a non-trivial decision node. As games increase in complexity, performance moves away from substantively rational play and towards blunderer play.

A notable feature of the treatment two data is that 4 of the subjects won all games up to rank $\hat{r}_i$ and lost all games more complex than $\hat{r}_i$. The bound on their performance was like a wall. The four perfect walls were at rank 6.5, 7.5, 7.5 and 9.5.

Depending on how strict one is about the criteria most of the other baseline subjects appear to have a bound on their rationality in the interval between rank 5 and rank 14. In order to measure these less than perfect bounds, we estimate a logit regression of the probability a subject wins against the rank of the game. The estimated rationality bound for subject *i*, $\hat{r}_i$, is either the perfect bound on *i*'s performance or if less than perfect then it is the point at which a logit regression for *i* estimates the subject as being equally likely to win or lose games of that rank.

Figures 5, reports a histogram of $\hat{r}_i$. The estimates of $\hat{r}_i$ include the four perfect walls as well as the 56 bounds estimated with the logit model. The median bound is 8. The estimated bounds for all of the subjects are contained in the interval [5,14].

<Figure 5>

V. DISCUSSION

Nim is a game with a computable algorithm that if used by both players would implement the substantively rational outcome of the game. The algorithm almost always wins against humans if the complexity of the game is not low. The bound on our subject's ability to implement substantively rational strategies is surprising low. Moreover, we observed a rather small range for the estimated bound when compared with the complexity of most popular parlor games or most life decisions, like whether to go to graduate school, accept a job offer, save for retirement, or have children.

In previous work, we observed people with perfect bounds on their ability to implement a substantively rational outcome. They win all winnable games of rank *r* or less and lose all winnable games of rank *r*+1 or more. In this experiment, 4 people had a perfect bound. For most people their performance deteriorates rapidly in games beyond a certain rank.

The wall model assumes people implement a substantively rational outcome up to their bound, and play like a blunderer (uniformly random) beyond their bound. This model makes more accurate predictions about the outcome of a game played by people than the value of the game derived from the theory of substantive rationality. However, the wall model over-estimates the advantage the person with a higher estimated bound has when actually playing the game.

Bouton's algorithm does not depend on Von Neumann and Morgenstern's (1943) concept of a pure strategy. Instead it assigns a value to any position by analyzing the binary representation of the position and identifying whether it is balanced or unbalanced. It is relatively easy to compute whether a position is balanced or unbalanced.

Sometimes people defend an analysis that uses the theory of substantive rationality to make predictions on the grounds that people find the equilibrium through some adaptive process. For example, Nash (1950, p.21) wrote, "We shall now take up the "mass-action" interpretation of equilibrium points... It is unnecessary to assume that the participants have full knowledge of the total structure of the game, or the ability and inclination to go through any complex reasoning processes. But participants are supposed to accumulate empirical information on the relative advantages of the various pure strategies at their disposal."

In thinking about Nim, it seems to us that Nash goes wrong in assuming that people can collect empirical information about their "pure strategies." For example, in the (5,4,3,5) game the cardinality of the space of "pure strategies" is so large that even the concept of a googol would not help us express the number here. Adaptive behavior takes place in a completely different and much, much smaller space than the space of "pure strategies."

After using backward induction to show that chess has a value, Von Neumann and Morgenstern (1980 [1943], p.125) remark, "...our proof ...gives no practically usable method to determine [the value]. This relative, human difficulty necessitates the use of those incomplete, heuristic methods of playing, which constitute 'good' chess..."[4] To the extent that our subjects are accumulating empirical information it is over a set of "incomplete, heuristic methods" of play. It is probable that the intersection of the set of heuristic methods imagined by our subjects and the set of pure strategies is empty for Nim games with more than 14 stones in 3 or more rows.

REFERENCE

Robert Aumann, "Reply to Binmore," *Games and Economic Behavior* 17(1), November 1996, p. 138-146.

Ken Binmore, "A Note on Backward Induction" *Games and Economic Behavior* 17(1), November 1996, p. 135-137.

Ken Binmore, John McCarthy, Giovanni Ponti, Larry Samuelson, Avner Shaked, "A Backward Induction Experiment," *Journal of Economic Theory*, 2002.

Charles L. Bouton, "Nim A Game with a Complete Mathematical Theory,"

[4] Theproof that chess has a value is frequently attributed to Zermelo, but see Schwalbe and Walker (2001).

The Annals of Mathematics 1901–1902,3(1), 35-39.

Eric Johnson, Colin Camerer, Sankar Sen, and Talia Rymon, “Detecting Failures of Backward Induction: Monitoring information search in sequential bargaining experiments,” *Journal of Economic Theory*, 104, 2002, p. 16-47.

Carl N. McKinney, Jr. and John B. Van Huyck, “Estimating Bounded Rationality and Pricing Performance Uncertainty,” laser-script 2004.

John F. Nash, *Non-cooperative Games: A Dissertation* May 1950; reprinted in *The Essential John Nash*, edited by H.W. Kuhn and S. Nasar (Princeton, Princeton University Press, 2002).

Ariel Rubinstein *Modeling Bounded Rationality*, (Cambridge, MA: The MIT Press, 1998).

Ulrich Schwalbe and Paul Walker, “Zermelo and the Early History of Game Theory,” *Games and Economic Behavior*, 34, 2001, 123-137.

Reinhard Selten, “The Chain Store Paradox” *Theory and Decision*, 1978; reprinted in *Models of Strategic Rationality* (Dordrecht, Kluwer Academic Publishers, 1988).

John Von Neumann and Oskar Morgenstern, *Theory of Games and Economic Behavior*, (Princeton, Princeton University Press, 1980 [1944]).

Table 1: Player One's win percentage for Game (5,3,4,5,0)

$\bar{r}_1$ \ $\bar{r}_2$	5	6	7	8	9	10	11	12	13	14
14	.9984	.9984	.9933	.9933	.9493	.9458	.9259	.8067	.5666	.4610
13	.9948	.9948	.9879	.9879	.9307	.9157	.8751	.7285	.5205	.4149
12	.9848	.9848	.9748	.9636	.8777	.8364	.7545	.5635	.3555	.1930
11	.9652	.9652	.9382	.9033	.7732	.6933	.5595	.3685	.2182	.1091
10	.9362	.9289	.8768	.8053	.6242	.4951	.3614	.2181	.1143	.0455
9	.8900	.8697	.7864	.6709	.4830	.3539	.2548	.1476	.0757	.0314
8	.8207	.7795	.6530	.4813	.2934	.1786	.1148	.0535	.0198	.0096
7	.7280	.6639	.4889	.3172	.1866	.1100	.0699	.0353	.0198	.0096
6	.5994	.4984	.3234	.2024	.1153	.0658	.0420	.0240	.0120	.0053
5	.5024	.4014	.2663	.1695	.1015	.0619	.0420	.0240	.0120	.0053

Table 2: Player One's win percentage, game (7,4,3)

$\bar{r}_1$ \ $\bar{r}_2$	5	6	7	8	9	10	11	12	13	14
14	.9999	.9999	.9992	.9992	.9930	.9930	.9231	.9231	.0000	.0000
13	.9386	.9386	.8651	.8026	.7276	.5990	.3942	.1978	.0000	.0000
12	.9386	.9386	.8651	.8026	.7276	.5990	.3942	.1978	.0000	.0000
11	.9311	.9300	.8533	.7861	.7035	.5650	.3590	.1625	.0000	.0000
10	.9099	.9056	.8181	.7360	.6323	.4714	.2653	.1176	.0000	.0000
9	.8694	.8590	.7481	.6387	.5111	.3501	.1936	.0780	.0000	.0000
8	.8130	.7865	.6426	.4956	.3680	.2442	.1307	.0531	.0000	.0000
7	.7263	.6790	.5017	.3547	.2632	.1763	.0931	.0401	.0000	.0000
6	.5858	.5112	.3338	.2282	.1694	.1166	.0628	.0255	.0000	.0000
5	.5057	.4310	.2834	.1982	.1525	.1014	.0533	.0210	.0000	.0000

Table 3: Player One's performance by *v(g)*

	Won	*Lost*	
$v(g) = W$	289 (.507)	281 (.493)	570
$v(g) = L$	283 (.496)	287 (.504)	570
	572 (.503)	568 (.497)	1140

Fisher exact = 0.425

Table 4

Game Pairs	*Outcome*				*?^{2}Test*		
Unbalanced/Balanced	*W/W*	*W/L*	*L/W*	*L/L*	*v(g)*	*Uniform*	*Wall*
(5,3,4,5,0)/(7,4,3,0,0)	9	6	7	8	0.000	0.083	0.170
(5,4,0,4,4)/(2,0,5,3,4)	9	3	9	9	0.000	0.040	0.038
(7,0,3,7,0)/(2,5,0,5,2)	11	4	4	11	0.000	0.189	0.834
(6,3,7,1,0)/(1,5,1,0,5)	8	6	7	9	0.000	0.024	0.033
(3,0,2,7,5)/(7,2,4,1,0)	15	5	3	7	0.000	0.189	0.516
(4,7,2,4,0)/(6,0,7,0,1)	10	3	5	12	0.000	0.126	0.711
(6,4,5,2,0)/(3,3,3,0,3)	6	7	7	10	0.000	0.131	0.038
(6,5,6,0,0)/(3,0,4,6,1)	6	3	7	14	0.000	0.149	0.106
(6,0,4,3,4)/(6,6,0,0,0)	8	7	8	7	0.000	0.200	0.012
(5,3,0,3,6)/(2,7,0,0,5)	7	8	6	9	0.000	0.062	0.254
(4,6,1,0,6)/(4,4,2,2,0)	6	8	8	8	0.000	0.134	0.021
(2,0,2,7,6)/(4,3,3,4,0)	12	6	1	11	0.000	0.077	0.412
(5,0,5,5,2)/(1,5,6,2,0)	13	5	6	6	0.000	0.083	0.298
(6,0,6,3,2)/(1,6,6,0,1)	11	11	2	6	0.000	0.272	0.044
(2,7,1,0,7)/(7,0,0,0,7)	6	7	5	12	0.000	0.075	0.404
(4,6,1,5,1)/(3,4,4,1,2)	7	5	8	10	0.000	0.045	0.041
(4,7,1,2,3)/(1,1,6,2,4)	13	3	6	8	0.000	0.172	0.153
(2,2,5,3,5)/(3,2,1,3,3)	7	11	6	6	0.000	0.041	0.001
(2,2,5,7,1)/(1,5,2,2,4)	10	7	4	9	0.000	0.068	0.429
Total (Percent)	174 (31)	115 (20)	109 (19)	172 (30)	0.000	0.000	0.000

Table 5: Outcome contingent on relative estimated walls.

	Won	*Lost*	
$\hat{r}_1 > \hat{r}_2$	362 (.633)	210 (.367)	572
$\hat{r}_1 < \hat{r}_2$	203 (.366)	351 (.634)	554
$\hat{r}_1$ ' $\hat{r}_2$	7 (.500)	7 (.500)	14
	572 (.502)	568 (.498)	1140

Fisher exact = 0.000

Table 6: Random Effects Logit Models

	Full Model	**Rational Model**	**Wall Model**	**Empirical Model**
v(g)	0.183 (0.639)	-0.047 (0.708)	-	-
w(g)	.075 (0.896)	-	1.781 (0.000)	-
$\hat{r}_1 \& \hat{r}_2$	0.192 (0.001)	-	-	0.200 (0.000)
round	0.000 (0.982)	-	-	-
r(g)	0.065 (0.540)	-	-	-
Constant	-1.105 (533)	0.035 (0.786)	-0.846 (0.000)	.011 (0.909)
s	0.27	0.31	0.49	0.27
Log likelihood	-724.610	-763.58	-729.80	-724.90
Likelihood ratio test	-	77.94 (0.000)	10.38 (0.035)	0.59 (0.965)

(p-values are in parenthesis.)

Figure 1: Treatment 1 user interface

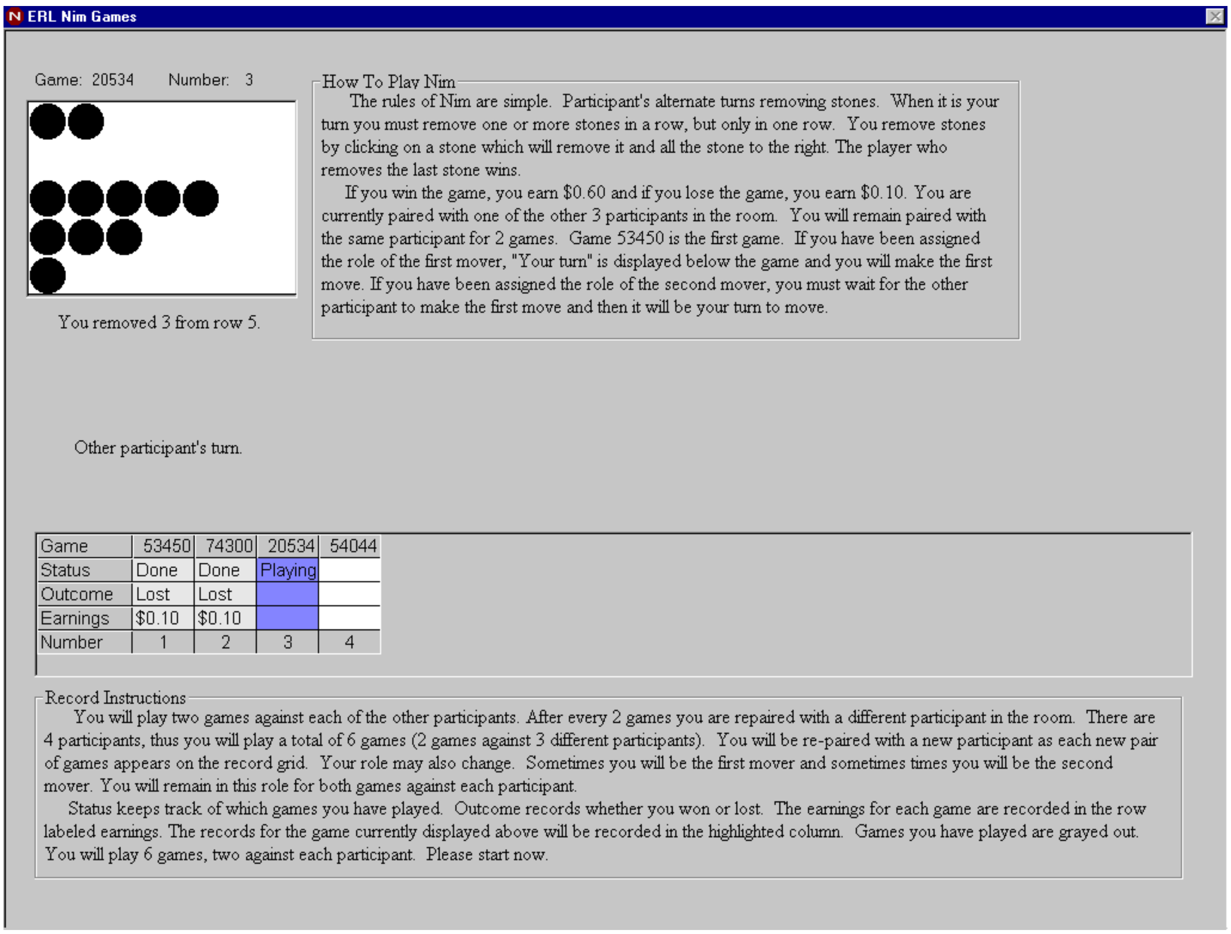

Figure2: Treatment 2 user interface

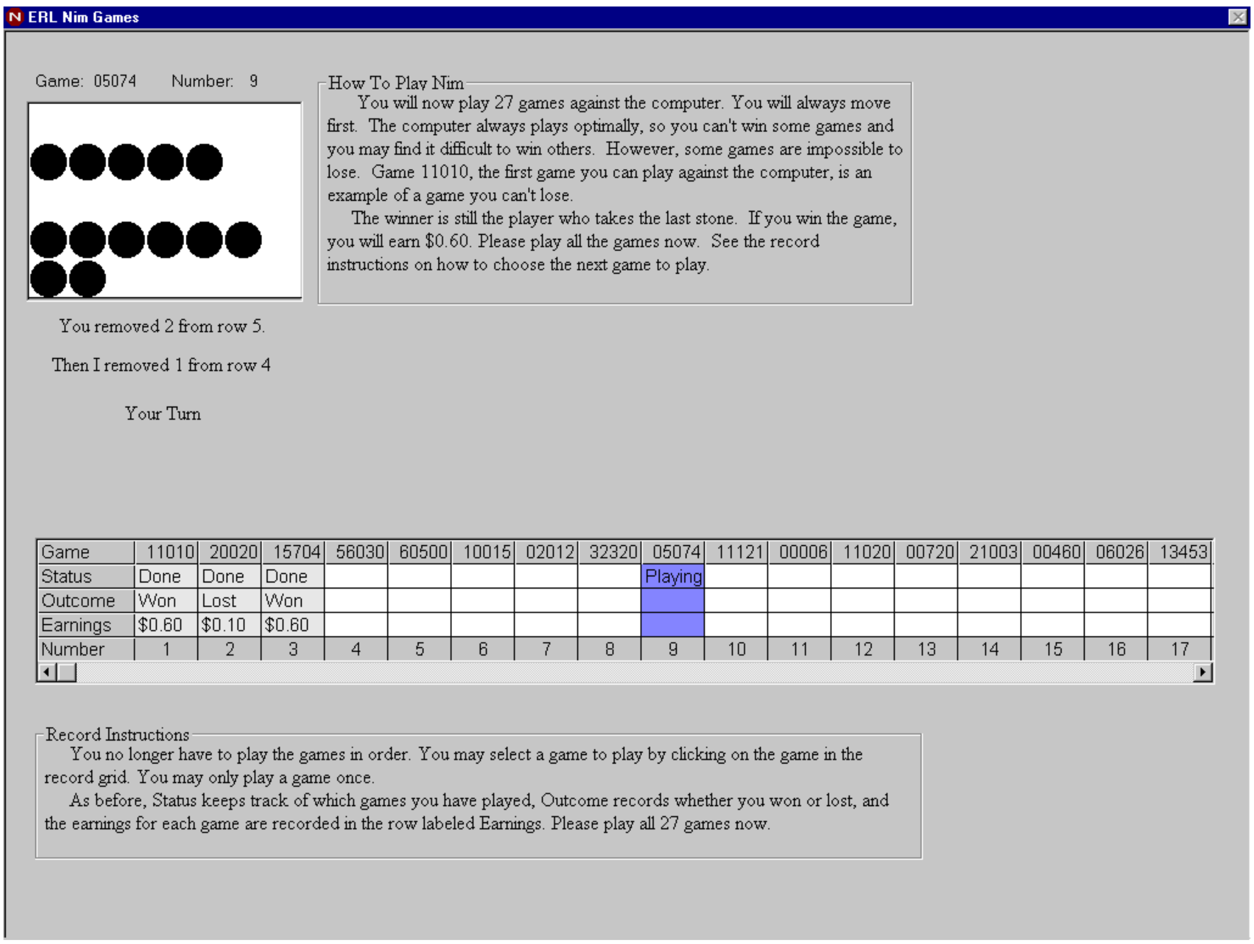

Figure 3: Comparing model predictions to subject data.

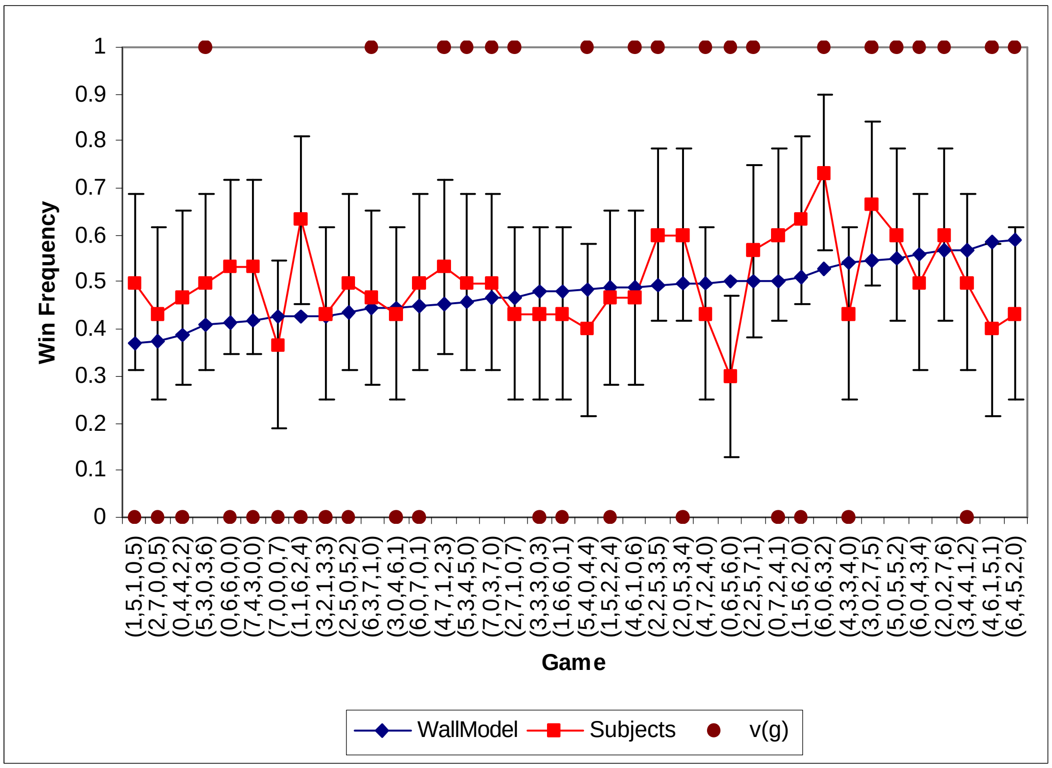

Figure 4: Frequency of initial actions compared

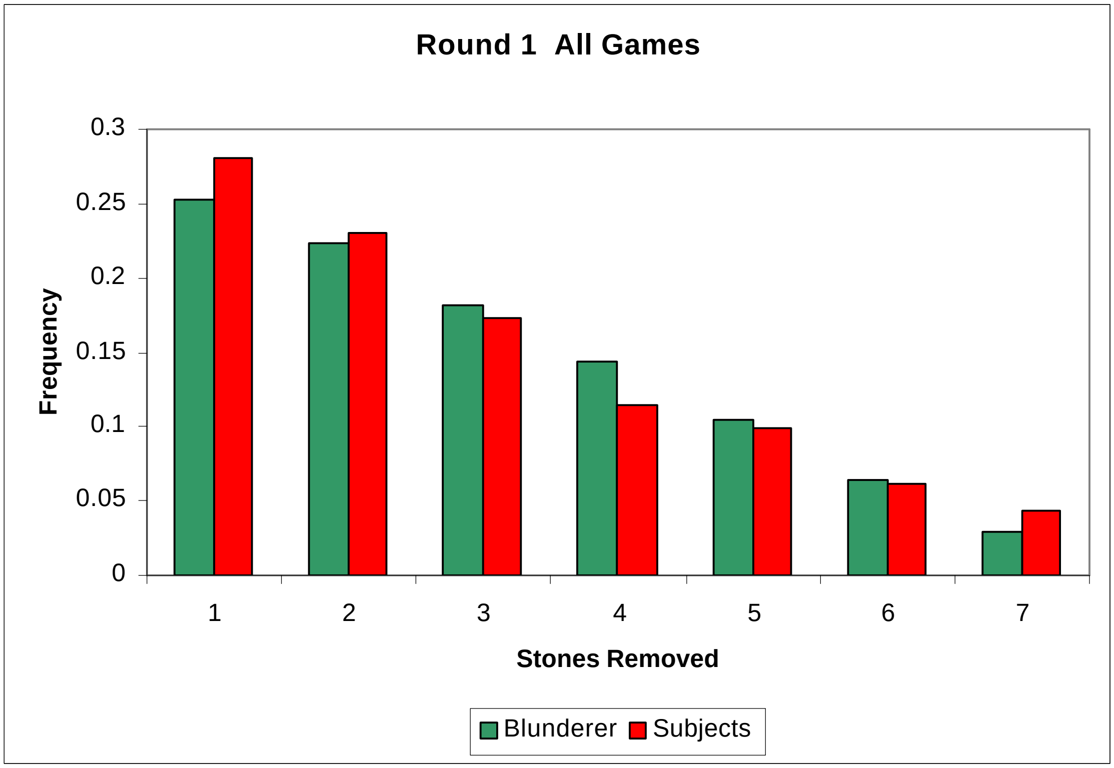

Figure 5: Histogram of $\hat{r}_i$