The Economic Journal, **107** (*November*), 1684–1700. © Royal Economic Society 1997. Published by Blackwell Publishers, 108 Cowley Road, Oxford OX4 1JF, UK and 350 Main Street, Malden, MA 02148, USA.

ON THE ROBUSTNESS OF BEHAVIOUR IN EXPERIMENTAL 'BEAUTY CONTEST' GAMES*

John Duffy and Rosemarie Nagel

We report and compare results from several different versions of an experimental interactive guessing game first studied by Nagel (1995), which we refer to as the 'beauty contest' game following Keynes (1936). In these games, groups of subjects are repeatedly asked to simultaneously guess a real number in the interval [0, 100] that they believe will be closest to 1/2 times either the median, mean, or maximum of all numbers chosen. We also use our experimental data to test a simple model of adaptive learning behaviour.

> It might have been supposed that competition between expert professionals, possessing judgement and knowledge beyond that of the average private investor, would correct the vagaries of the ignorant individual left to himself. It happens, however, that the energies and skill of the professional investor and speculator are mainly occupied otherwise. For most of these persons are, in fact, largely concerned, not with making superior long-term forecasts of the probable yield of an investment over its whole life, but with foreseeing changes in the conventional basis of valuation a short time ahead of the general public. – John Maynard Keynes (1936, p. 154).

As Keynes observed, much of professional investment behaviour is concerned with attempting to keep one step ahead of average behaviour. The average investor, or in Keynes' view, 'the ignorant individual left to himself', is unlikely to disappear, as it is precisely the behaviour of this type of investor that market professionals are attempting to forecast. In this paper we study the evolution of behaviour in a number of market-type 'beauty contest' games involving groups of experimental subjects.[1] The behaviour of the subjects in these games has much in common, in flavour if not specifics, with Keynes' insight regarding the behaviour of investors in financial markets.

We study variations on the 'beauty contest' game that has been conducted experimentally and analysed previously by Nagel (1995), Stahl (1996) and Ho *et al.* (1996). In the version of the game studied in these papers, groups of

* This paper was written while Nagel was a post-doc in the Department of Economics at the University of Pittsburgh. We thank Colin Camerer, Ido Erev, Nick Feltovich, Al Roth, Dale Stahl, Nick Vriend and two anonymous referees for helpful comments and suggestions. Duffy acknowledges support from the U.S. National Science Foundation. The data from the experiments reported in this paper are available on request from the authors.

[1] The beauty contest analogy is also due to Keynes (1936, p. 156), who likened professional investment activity to newspaper beauty contests of the time where readers were asked to choose the six prettiest faces from among 100 photographs with the winner being the person whose preferences were closest to the average preferences. The reasoning process that Keynes describes for this beauty contest game applies to the p-mean, p-median and p-maximum games described below, where $p \neq 1$. For $p = 1$, as in Keynes' example, different levels of reasoning cannot be distinguished and therefore it is necessary to choose $p \neq 1$. See also Nagel (1995) and Ho *et al.* (1996) who also make use of Keynes' beauty contest metaphor. Nagel (1995) refers to the 'beauty contest' game as a guessing game.

subjects are repeatedly asked to simultaneously guess a real number in the interval [0, 100]. The winner is the person(s) whose guess is closest, in absolute value, to p times the mean of all numbers chosen, where p is some pre-announced fraction ($p \neq 1$). The winner of each round wins a fixed prize and the other participants receive nothing. If there are several winners, the prize is divided equally among them. For the case where $0 \leqslant p < 1$, iterated elimination of dominated strategies leads to a unique Nash equilibrium in which all players announce 0. For the case where $p > 1$, and a guessing interval of $[\epsilon, 100]$, $\epsilon > 0$, the unique Nash equilibrium is for all players to announce 100. When $p = 1$, the Nash equilibrium is no longer unique; any number in the guessing interval is a potential Nash equilibrium provided that all players announce the same number.[2] We shall focus our attention on the case where $0 \leqslant p < 1$. Nagel (1995) and Ho *et al.* (1996) report that after several repetitions of the p-mean game where $0 \leqslant p < 1$, the mean of all chosen numbers is consistently and significantly greater than zero, though there is some tendency for the mean to decline over time.

In this paper we test the robustness of these findings for the p-mean game by considering two additional experimental treatments and by playing the game for a longer number of rounds. In the two new treatments, we change the order statistic of the game, replacing the mean with either the median or the maximum. We shall refer to the three different games according to the order statistic used; for example, the p mean game will be referred to simply as the 'mean' game. For $0 < p \leqslant \frac{1}{2}$, both the 'median' and 'maximum' games are dominance solvable and have the same equilibrium as the mean game, namely that all players announce zero.[3]

We chose to examine how players behave in the median and maximum games in comparison with the mean game, because these two additional treatments assign very different weight to 'extreme guesses' – guesses that lie in the tails of the sample distributions. In the median game, these extreme guesses do not affect the value of the median as much as they affect the value of the mean, since the median is an unbiased measure of central tendency. Consequently, we hypothesise that by comparison with the mean game, players in the median game should be relatively less concerned about extreme guesses; they will therefore be capable of achieving greater coordination and will be relatively closer to the equilibrium strategy where all announce zero at the end of the game. By contrast, in the maximum game, the behaviour of the outlier – the player choosing the highest number – is the primary concern of all players. Since a single player determines the winning number, we expect that by comparison with the mean or median games, subjects in the maximum game will have achieved relatively less coordination, and will therefore be relatively

[2] The $p = 1$ case is similar to the coordination game studied by Van Huyck *et al.* (1990).

[3] For $\frac{1}{2} < p < 1$, the maximum game has no pure equilibria and for $p = \frac{1}{2}$, the equilibrium where all choose zero is a weak equilibrium. For $p \leqslant \frac{1}{2}$, the numbers in $[100p, 100]$ are weakly dominated by $100p$. By iterated elimination of weakly dominated strategies, 0 is the only non-excluded strategy. For the two player game, zero is a weakly dominant strategy.

further from the equilibrium strategy where all announce zero at the end of the game.

One interpretation of our experiment is that it provides a test of the importance of framing effects in group decision making.[4] Since all three treatments have the same equilibrium, it is clearly the framing of the objective of each guessing game treatment that must be responsible for any differences in observed behaviour across the three treatments. However, we believe that our three treatments also have important economic interpretations as well. As we have indicated, the three treatments provide a nice spectrum of environments that may be more or less conducive to the achievement of coordination by groups of economic agents. In this respect, our experiment is similar to the various coordination game environments that have been studied by Van Huyck *et al.* (1990, 1991). Alternatively one might view the different environments as giving more or less weight to 'fundamental' or 'speculative' behaviour. For example, one can view the median game environment as one in which players may be relatively more concerned with 'market fundamentals' such as the fraction p or the previous period's market average, and relatively less concerned with speculating about the future actions of other players. By contrast, one can view the maximum game environment as one in which players are perhaps relatively less concerned with such market fundamentals, and relatively more concerned with speculating about the future actions of those individuals who choose high numbers – Keynes's 'ignorant individuals'.

Another purpose of our experiment was to consider the performance of a simple theory of adaptive learning behaviour – learning direction theory – in explaining the behaviour of players from period to period across the three different treatments. According to learning direction theory, the direction of a player's action from the previous round to the current round can be explained by the player's previous round action in relation to the *ex-post* optimal action. This theory of learning behaviour was first proposed by Selten and Stoecker (1986) and has been applied by Selten and Buchta (1994), Nagel (1995) and others. In particular, Nagel (1995), Stahl (1996) and Ho *et al.* (1996) have all considered learning direction theory as a way of explaining the behaviour of subjects in the mean game, and so it seemed natural to consider the explanatory power of learning direction theory for the median and maximum games as well. While learning direction theory is not easily generalised, Stahl (1996) and Camerer and Ho (1996) have formulated more general learning models that incorporate the basic idea of learning direction theory.

I. BEHAVIOUR IN THE MEAN GAME: A MODEL OF ITERATED BEST REPLY

Nagel (1995) suggested the following 'descriptive model' of how subjects might play the mean game. In round 1, when there is no information about the

[4] For an introduction to the issue of framing and some empirical examples, see e.g. Tversky and Kahneman (1986).

behaviour of other players, subjects reason that a plausible reference or focal point is 50, which is the midpoint of the guessing interval [0, 100], and a good candidate for the mean if one believes that the first round guesses will be uniformly distributed over [0, 100]. While it is assumed that all players have the same initial reference point, players may differ according to the depth of reasoning they employ when determining their guesses for each round of the game. Denote the guess of player i in round t by $g_i(t)$. Player i's depth of reasoning in round t (the player's iterated best reply) is defined as the value of d that solves

$$g_i(t) = m(t-1)\,p^d,$$

where $0 < p < 1$ is the given fraction and $m(t-1)$ denotes the mean from the previous round; for $t = 1$, $m(0) = 50$, the hypothesised initial reference point. To simplify the analysis we will follow Nagel (1995) and group the continiuous d values into discrete categories based on neighbourhood intervals centred on certain integer values: $d = 0$, 1, 2 and 3. These neighbourhood intervals have the boundaries $m(t-1)\,p^{d+\frac{1}{2}}$ and $m(t-1)\,p^{d-\frac{1}{2}}$. Values for d below the lower bound for the $d = 0$ category are lumped together into a single category, referred to as $d < 0$, as are values for d above the upper bound for the $d = 3$ category, referred to as $d > 3$. Note that the Nash equilibrium prediction requires that all players employ a depth of reasoning $d = +\infty$. Using the six neighbourhood intervals, $d < 0$, $d = 0$, 1, 2, 3 and $d > 3$, Nagel (1995) observed that approximately 80% or more of players' guesses in the first four rounds of the mean game could be categorised as either $d = 0$, 1, 2 or 3.[5]

A potential difficulty with the mean game is that the mean is a biased measure of central tendency; it gives relatively more weight to extreme, 'outlier' observations. This bias may have affected the behaviour of subjects in the mean game treatment. Indeed, one observes that in the final rounds of Nagel's (1995) $\frac{1}{2}$- and $\frac{2}{3}$-mean game experiments, a few players, perhaps out of frustration, decided to choose very large numbers – numbers that were significantly larger than the numbers they chose in earlier rounds. Similarly, in the $\frac{1}{2}$-mean game sessions that we replicated for this paper, there were some players in the third and fourth rounds who chose larger numbers than they chose in either the first or second rounds.

With sample sizes of 13–16 players, these large final round deviations can and do affect the value of the mean. Subjects who recognise this possibility may want to avoid using too great a depth of reasoning when determining their guesses. Thus, while subjects may be capable of higher depths of reasoning – even at infinite depth of reasoning – the use of the mean as the order statistic may serve to alter their behaviour so that it appears as though they are employing a more shallow depth of reasoning than they might otherwise

[5] See also Stahl (1996) who develops a parametric approach that can be used to test the validity of this depth of reasoning approach. Ho *et al.* (1996) obtain similar results in the initial rounds of the mean game using smaller group sizes of three or seven subjects.

choose. This observation serves as our motivation for comparing behaviour in the mean game with behaviour in the median and maximum games.

II. EXPERIMENTAL DESIGN

To test our conjectures regarding behaviour in the different treatments, we conducted 12 experimental sessions. We conducted three sessions each of the $\frac{1}{2}$-median, $\frac{1}{2}$-mean and $\frac{1}{2}$-maximum games for four rounds, and one session of each of these three games for ten rounds. For each session we recruited a group of 13–16 subjects from the undergraduate population at the University of Pittsburgh (for a total of 175 subjects). Subjects were only allowed to participate in a single experimental session; no subject had any previous experience with any of the games.

Each subject was guaranteed a show-up fee of $5. In addition, the winner of each round was paid $20. If there was more than one winner, the $20 prize was divided equally among the winners. We explained to subjects in the mean and median games how the mean and median values are determined. Subjects were isolated from one another and were not allowed to communicate with each other. They were informed in advance how many rounds would be played. Subjects wrote their guesses for each round on cards. The cards were then collected and the numbers were read aloud as well as written on a black board without identifying any player. Depending on the game, subjects were informed of the median, mean or maximum, $\frac{1}{2}$ times the median, mean or maximum, and the winning number(s) – the number(s) closest to $\frac{1}{2}$ the median, mean or maximum. Once this information had been revealed, the next round of the game was begun.

III. EXPERIMENTAL RESULTS

Fig. 1 shows the cumulative frequencies with which numbers in the [0, 100] guessing range were chosen in the first four rounds of each treatment using pooled data from all sessions of a treatment. We observe that by the fourth round of all median games, 90% of guesses are less than ten while by the fourth round of all maximum games, only 13% of guesses are less than ten; in the mean games, 76% of guesses are less than ten by the fourth round.

First Round Behaviour

In the first round of every session, no subject chose zero. At the other extreme, the percentage of first round guesses that were weakly dominated (i.e. greater than 50) was 5% in all median games, 10% in all mean games and 15% in all maximum games.

To examine whether there were any differences in first round behaviour across treatments, we pooled data for all four sessions of a given treatment. We

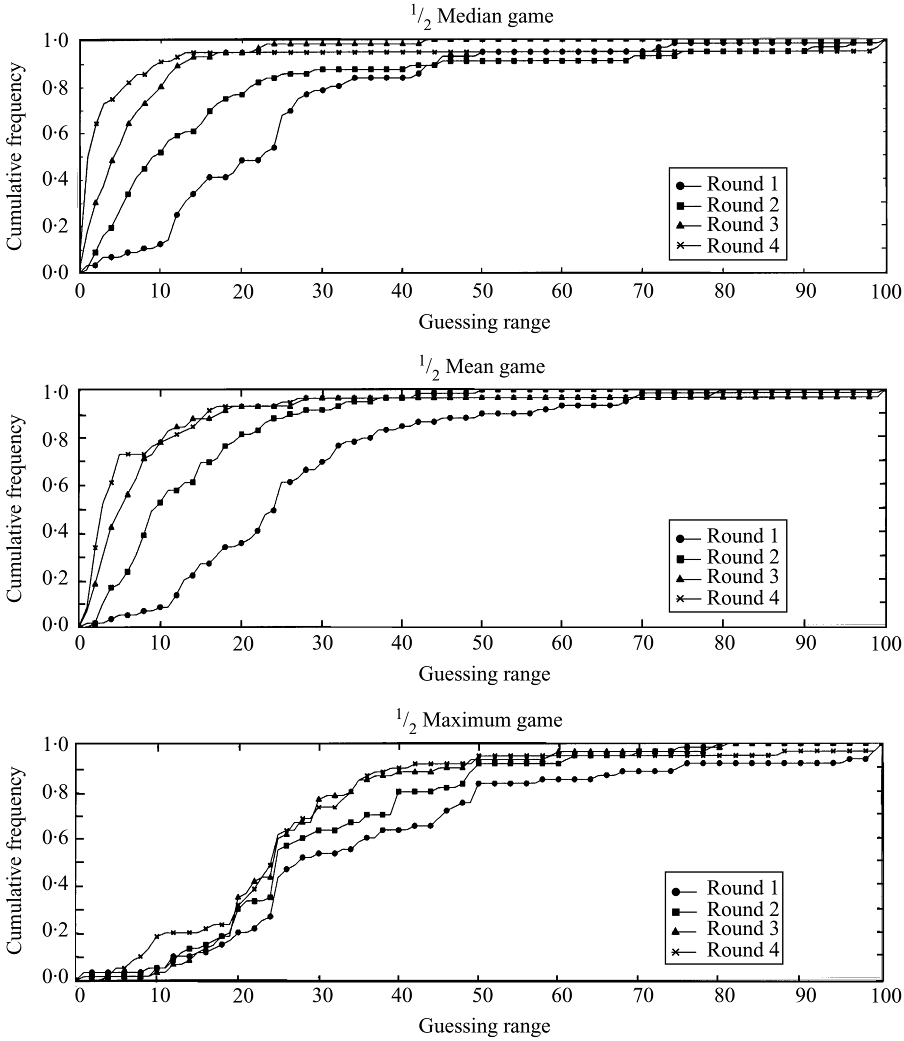

Fig. 1. Cumulative frequency of guesses in rounds 1–4 ½ median game.

then conducted non-parametric, robust rank order tests of the null hypothesis that there is no significant difference in the distribution of first round guesses between the mean and median games, the mean and maximum games, and the median and maximum games.[6] We find that we cannot reject the null hypothesis of no difference in the distribution of first round guesses between the mean and median games ($p = 0{\cdot}111$). However, we can reject this same null hypothesis in comparisons between the maximum and mean games ($p = 0{\cdot}003$) and between the maximum and median games ($p = 0{\cdot}000$) in favour of the

[6] See Siegel and Castellan (1988, pp. 137–43) for a description of the robust rank order test used here.

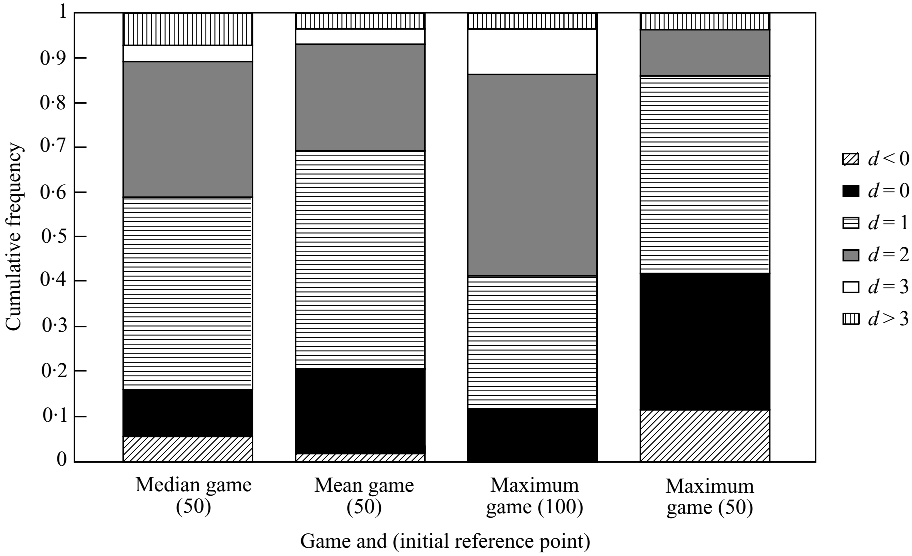

Fig. 2. Classification of first round choices according to depth of reasoning for given initial reference points.

alternative hypothesis that first round guesses are higher in the maximum game than in either the mean or median games.

In an effort to explain these differences, we classified subjects' first round choices according to the six discrete depth of reasoning categories discussed in Section I. For the mean and median games we used the hypothesised initial reference point of 50 while for the maximum game we considered both 50 and 100 as initial reference points. From a cognitive point of view, 100 is probably the more plausible initial reference point for the maximum game since it is the upper bound to the guessing interval and subjects are specifically instructed to choose the number they believe will be closest to $\frac{1}{2}$ times the maximum.[7]

Fig. 2 plots the relative frequency with which the six categories were observed in the first round of the three treatments. We see that with 50 as the initial reference point, 73% of first round choices in the mean and median game are categorised as either $d = 1$ or $d = 2$. Similarly, with 100 as the initial reference point, 75% of first round choices in the maximum game can be categorised as either $d = 1$ or $d = 2$; if 50 is used as the initial reference point, only 55% of first round choices in the maximum game can be categorised as either $d = 1$ or $d = 2$. If one assumes that the initial distribution of depths of reasoning (d values) should be roughly the same across all three treatments, then the appropriate first round reference point would appear to be 50 in the

[7] Nagel (1995) and Stahl (1996) found that for different p values in the mean game the most plausible initial reference point was always 50. See also Ho *et al.* (1996) who estimate the initial reference point for the mean game using Nagel's (1995) data.

Table 1

Percentage Change in the Median Guess From Round 1 to Round 4

Game/session	Percentage change in median guess
½-Median game	
Session 1	−84·0
Session 2	−98·9
Session 3	−92·3
Session 4	−94·7
½-Mean game	
Session 5	−91·4
Session 6	−40·0
Session 7	−91·3
Session 8	−87·3
½-Maximum game	
Session 9	−10·7
Session 10	−11·5
Session 11	−70·0
Session 12	−24·4

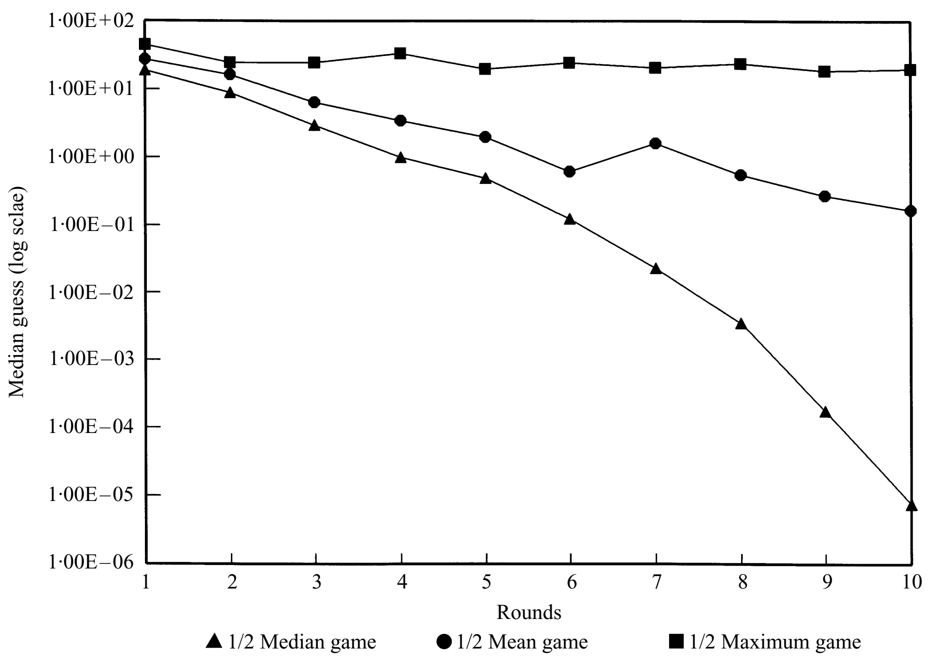

Fig. 3. The path of the median guess over 10 rounds of each game.

mean and median games, and 100 in the maximum game. This higher initial reference point in the maximum game may account for the statistically significant difference that was observed in the distribution of initial guesses between the maximum game and the mean and median games.

Table 2
Classification of Choices According to Depth of Reasoning

Round	Relative frequencies*									
	1	2	3	4	5	6	7	8	9	10
(a) ½-Median Game										
Session 1										
$d > 3$	0·00	0·00	0·00	0·00						
$d = 3$	0·00	0·00	0·00	0·00						
$d = 2$	0·27	0·27	0·20	0·47						
$d = 1$	0·47	0·27	0·53	0·27						
$d = 0$	0·13	0·13	0·13	0·13						
$d < 0$	0·13	0·33	0·13	0·13						
Session 2										
$d > 3$	0·27	0·07	0·07	0·13						
$d = 3$	0·07	0·27	0·33	0·27						
$d = 2$	0·33	0·27	0·27	0·47						
$d = 1$	0·33	0·20	0·13	0·07						
$d = 0$	0·00	0·07	0·00	0·00						
$d < 0$	0·00	0·13	0·20	0·07						
Session 3										
$d > 3$	0·00	0·00	0·00	0·00						
$d = 3$	0·00	0·00	0·08	0·46						
$d = 2$	0·23	0·31	0·23	0·23						
$d = 1$	0·38	0·31	0·54	0·08						
$d = 0$	0·31	0·23	0·15	0·08						
$d < 0$	0·08	0·15	0·00	0·15						
Session 4										
$d > 3$	0·00	0·00	0·00	0·15	0·08	0·00	0·15	0·23	0·62	1·00
$d = 3$	0·08	0·08	0·00	0·00	0·08	0·08	0·31	0·31	0·23	0·00
$d = 2$	0·38	0·31	0·62	0·46	0·31	0·54	0·54	0·46	0·15	0·00
$d = 1$	0·54	0·38	0·31	0·15	0·08	0·15	0·00	0·00	0·00	0·00
$d = 0$	0·00	0·23	0·08	0·00	0·08	0·08	0·00	0·00	0·00	0·00
$d < 0$	0·00	0·00	0·00	0·23	0·38	0·15	0·00	0·00	0·00	0·00
All Sessions										
$d > 3$	0·07	0·02	0·02	0·07						
$d = 3$	0·04	0·09	0·11	0·18						
$d = 2$	0·30	0·29	0·32	0·41						
$d = 1$	0·43	0·29	0·38	0·14						
$d = 0$	0·11	0·16	0·09	0·05						
$d < 0$	0·05	0·16	0·09	0·14						
(b) ½-Mean Game										
Session 5										
$d > 3$	0·00	0·00	0·00	0·00						
$d = 3$	0·06	0·06	0·00	0·06						
$d = 2$	0·13	0·50	0·25	0·63						
$d = 1$	0·44	0·13	0·63	0·19						
$d = 0$	0·13	0·19	0·06	0·00						
$d < 0$	0·25	0·13	0·06	0·13						
Session 6										
$d > 3$	0·00	0·00	0·00	0·00						
$d = 3$	0·07	0·21	0·36	0·21						
$d = 2$	0·43	0·36	0·36	0·07						
$d = 1$	0·43	0·36	0·07	0·36						
$d = 0$	0·00	0·07	0·00	0·21						
$d < 0$	0·07	0·00	0·21	0·14						
Session 7										
$d > 3$	0·13	0·00	0·00	0·00						
$d = 3$	0·00	0·07	0·00	0·00						

Table 2 (*cont.*)

Round	Relative frequencies*									
	1	2	3	4	5	6	7	8	9	10
$d = 2$	0·33	0·33	0·47	0·40						
$d = 1$	0·40	0·27	0·33	0·27						
$d = 0$	0·13	0·13	0·13	0·20						
$d < 0$	0·00	0·20	0·07	0·13						
Session 8										
$d > 3$	0·00	0·00	0·00	0·00	0·07	0·00	0·14	0·07	0·07	0·00
$d = 3$	0·00	0·14	0·00	0·00	0·07	0·14	0·29	0·14	0·14	0·00
$d = 2$	0·07	0·14	0·36	0·29	0·14	0·36	0·57	0·29	0·29	0·50
$d = 1$	0·71	0·57	0·43	0·57	0·71	0·29	0·00	0·29	0·29	0·29
$d = 0$	0·14	0·00	0·14	0·00	0·00	0·14	0·00	0·14	0·00	0·00
$d < 0$	0·07	0·14	0·07	0·14	0·00	0·07	0·00	0·07	0·21	0·21
All Sessions										
$d > 3$	0·03	0·00	0·00	0·00						
$d = 3$	0·03	0·12	0·08	0·07						
$d = 2$	0·24	0·34	0·36	0·36						
$d = 1$	0·49	0·32	0·37	0·34						
$d = 0$	0·10	0·10	0·08	0·10						
$d < 0$	0·10	0·12	0·10	0·14						
(*c*) $\frac{1}{2}$-Maximum Game										
Session 9										
$d > 3$	0·07	0·00	0·00	0·00						
$d = 3$	0·13	0·07	0·00	0·00						
$d = 2$	0·47	0·53	0·53	0·20						
$d = 1$	0·27	0·33	0·40	0·60						
$d = 0$	0·07	0·07	0·07	0·13						
$d < 0$	0·00	0·00	0·00	0·07						
Session 10										
$d > 3$	0·00	0·00	0·00	0·00						
$d = 3$	0·20	0·07	0·07	0·00						
$d = 2$	0·40	0·47	0·47	0·20						
$d = 1$	0·33	0·40	0·40	0·80						
$d = 0$	0·07	0·00	0·07	0·00						
$d < 0$	0·00	0·07	0·00	0·00						
Session 11										
$d > 3$	0·07	0·07	0·07	0·00						
$d = 3$	0·00	0·13	0·00	0·20						
$d = 2$	0·53	0·47	0·27	0·53						
$d = 1$	0·27	0·33	0·67	0·13						
$d = 0$	0·13	0·00	0·00	0·00						
$d < 0$	0·00	0·00	0·00	0·13						
Session 12										
$d > 3$	0·00	0·00	0·00	0·00	0·00	0·00	0·00	0·00	0·00	0·07
$d = 3$	0·07	0·13	0·00	0·00	0·00	0·00	0·00	0·00	0·00	0·07
$d = 2$	0·40	0·53	0·40	0·13	0·07	0·27	0·00	0·73	0·07	0·80
$d = 1$	0·33	0·33	0·40	0·87	0·67	0·73	0·80	0·27	0·67	0·07
$d = 0$	0·20	0·00	0·07	0·00	0·20	0·00	0·13	0·00	0·07	0·00
$d < 0$	0·00	0·00	0·13	0·00	0·07	0·00	0·07	0·00	0·20	0·00
All Sessions										
$d > 3$	0·03	0·02	0·02	0·00						
$d = 3$	0·10	0·10	0·02	0·05						
$d = 2$	0·45	0·50	0·42	0·27						
$d = 1$	0·30	0·35	0·47	0·60						
$d = 0$	0·12	0·02	0·05	0·03						
$d < 0$	0·00	0·02	0·03	0·05						

* Modal frequencies are underlined.

Behaviour Over Time

In Table 1 we report the percentage change in the median guess from the first to the fourth round of each individual session. We see that the percentage changes in the median guess from round 1 to round 4 are generally largest (in the direction of zero) in the median game sessions, and generally smallest in the maximum game sessions in accordance with our hypotheses. To test whether these percentage changes are significantly different from one another, we again performed a robust rank order test using the data of Table 1.

We find that the percentage changes in the median guess from rounds 1 to 4 are significantly larger in the median game than in the mean game ($p = 0.100$). Similarly, we find that the percentage changes in the median guess over rounds 1–4 are significantly larger in the mean game than in the maximum game ($p = 0.025$). It follows that the percentage changes in the median guess over rounds 1–4 are also significantly larger in the median game as compared with the maximum game ($p = 0.000$). These differences in the evolution of the median guess over time appear to persist beyond the first four rounds; Fig. 3 plots the path taken by the median guess in each of the single ten-round sessions that we conducted for each game. We see that at the end of ten rounds, the median guess in the median game is closest to the Nash equilibrium prediction while the median guess in the maximum game remains furthest from the equilibrium outcome.

To examine the reasoning process that subjects followed in rounds 2 to 10, we can use the same classification scheme for depth of reasoning that was used for the first round guesses. Now, however, we use as the reference point, the previous period's mean, median or maximum, depending upon the treatment. That is, player i's depth of reasoning in rounds $t = 2, 3, \ldots, 10$ is the value of d that solves

$$g_i(t) = m(t-1)\,p^d,$$

where $m(t-1)$ now represents the previous period's announced mean, median or maximum value. Tables 2(*a*)–(*c*) report for the median, mean and maximum game treatments, the relative frequency of choices in the neighbourhood categories of $d < 0$, $d = 0, 1, 2, 3$ and $d > 3$. We see that across all three treatments, the modal depth of reasoning categories (which are underlined) in rounds 2–4 are typically $d = 1$ or $d = 2$, as was the case for the first round. Indeed, over the first four rounds of every session, the majority of choices lie below the $d = 3$ category. If we consider the percentage of all guesses that are categorised as being either $d = 3$ or $d > 3$ in each round we cannot reject, using rank order tests at the 10% significance level, the null hypothesis of no significant difference in these percentages between any two of the three treatments. Thus, like Nagel (1995), we do not find any significant evidence that subjects employ increasing depths of reasoning over the first four rounds of the 'beauty contest' game. However, when we consider the ten round sessions, there is some evidence that players do employ greater depths of reasoning over time. This is most apparent in the final four rounds (7–10) of the $\frac{1}{2}$-median game, where we observe that all choices are classified as $d = 2$ or higher. By contrast, there is no evidence of increasing depth of reasoning in the longer, ten-round version of the maximum game. These findings suggest that if subjects are given enough time, and an environment where outliers have little influence, they may indeed learn to employ greater depths of reasoning.

IV. LEARNING DIRECTION THEORY

Following Nagel (1995), we examine whether players' behaviour from round to round follows the predictions of a simple adaptive learning scheme known as 'learning direction' theory (see Selten and Stoecker (1986)). This theory suggests that players adjust their decisions from round to round using an *ex-post* reasoning process based on the previous period's outcome. In the case of the 'beauty contest' game, if player i's guess in the previous round, $g_i(t-1)$, was higher than the winning number in the previous round, $p \times m(t-1)$, the theory predicts that player i will employ a greater depth of reasoning in round t by decreasing his adjustment ratio,

$$\frac{g_i(t)}{m(t-1)},$$

given the current reference point of $m(t-1)$, i.e.

$$g_i(t-1) > p \times m(t-1) \Rightarrow \frac{g_i(t)}{m(t-1)} < \frac{g_i(t-1)}{m(t-2)}.$$

Analogously, if player i's guess in the previous round was lower than the winning number in the previous round, then the theory predicts that player i will employ a more shallow depth of reasoning when forming his guess for the current round by increasing his adjustment ratio given the current reference point of $m(t-1)$, i.e.

$$g_i(t-1) < p \times m(t-1) \Rightarrow \frac{g_i(t)}{m(t-1)} > \frac{g_i(t-1)}{m(t-2)}.$$

Nagel (1995) found that a majority of players' actions in the first four rounds of the p-mean game were consistent with the predictions of learning direction theory. While there are a number of other learning models that could be applied to the data obtained from our experiments (e.g. naïve, adaptive or Bayesian learning models), our aim in this paper is not to test all of these competing theories, but rather to examine whether learning direction theory continues to be a good predictor of subject behaviour when the order statistic of the 'beauty contest' game is changed or when the number of rounds played of the game is increased.[8]

[8] See Stahl (1996) for an econometric analysis of how well learning direction theory compares with other learning theories using Nagel's (1995) data for the p-mean game. See also Ho *et al.* (1996) who compare the performance of learning direction theory with other theories of player behaviour using data collected from 10-round experimental sessions of the mean game involving small groups of subjects.

Table 3 (*a*) reports the frequencies of adjustment behaviour over rounds 1–4 using pooled data from all four sessions of each treatment. The data are categorised according to whether players' previous round guesses were greater than or less than the winning number and also according to whether players increased or decreased their adjustment ratios in the subsequent round. If in the current round, a player's guess turns out to have been greater than (less than) $\frac{1}{2}$ the reference point, learning direction theory predicts that in the following round, the adjustment ratio

$$\frac{g_i(t)}{m(t-1)},$$

will be decreased (increased) relative to subject's previous adjustment ratio. The frequencies of behaviour that are consistent with the predictions of learning direction theory are noted in boldface type and the total frequency of behaviour that is consistent with learning direction theory (LDT) is also noted.

We see that with a single exception (the transition from rounds 1 to 2 of the $\frac{1}{2}$-median game), an average of more than 65% of transition behaviour is consistent with the predictions of learning direction theory using pooled data for all three treatments. This finding is consistent with Nagel's (1995) finding for the *p*-mean game. A particularly interesting finding in Table 3 (*a*) is that players' depths of reasoning do not necessarily increase over time. In the first few rounds of the maximum game we see that a majority of players find that they are consistently below $\frac{1}{2}$ the reference point, and subsequently, most of these players choose to employ a more shallow depth of reasoning over time (they increase their adjustment ratio upwards). This type of behaviour is consistent with the predictions of learning direction theory, but runs counter to the common (game-theoretic) wisdom that players in all of these games should employ greater depths of reasoning over time.

In Table 3 (*b*) we disaggregate some of the findings reported in Table 3 (*a*) and note the frequency with which adjustment behaviour from one round to the next is consistent with the predictions of learning direction theory in each of our 12 sessions. Using this session-level frequency data, we can test whether learning direction theory provides a plausible characterisation of our data. The null hypothesis is that the frequency of adjustment behaviour that is consistent with learning direction theory should, on average, equal 0·5. That is, in Table 3 (*b*), for each transition period, we should observe roughly equal numbers of frequencies that are less than 0·5 as we observe frequencies that are greater than 0·5. We use the binomial test to determine whether we can reject this null hypothesis for each of the three round-to-round transitions using the 12 session-level observations reported in Table 3 (*b*).[9] We find that we can reject the null hypothesis for each of the three round-to-round transitions at a significance level that is always less than 10%. We conclude that learning direction theory is a plausible way of characterising transition behaviour over the first four rounds of all three treatments.

[9] See Siegel and Castellan (1988, pp. 38–44) for a description of the binomial test used here.

Table 3

Classification of Behaviour According to Learning Direction Theory

(*a*) Used pooled data on rounds 1–4

	Adjustment ratio is	Rounds 1–2	Rounds 2–3	Rounds 3–4
½-Median game guess was				
> ½ Ref. Pt.	Decreased	**0·36**	**0·59**	**0·59**
	Increased	0·50	0·23	0·34
< ½ Ref. Pt.	Increased	**0·13**	**0·14**	**0·07**
	Decreased	0·02	0·04	0·00
	Consistent with L.D.T.*	0·49	0·73	0·66
½-Mean game guess was				
> ½ Ref. Pt.	Decreased	**0·59**	**0·49**	**0·47**
	Increased	0·27	0·27	0·25
< ½ Ref. Pt.	Increased	**0·10**	**0·24**	**0·22**
	Decreased	0·03	0·00	0·05
	Consistent with L.D.T.	0·69	0·73	0·69
½-Maximum game guess was				
> ½ Ref. Pt.	Decreased	**0·22**	**0·21**	**0·39**
	Increased	0·03	0·14	0·06
< ½ Ref. Pt.	Increased	**0·54**	**0·57**	**0·49**
	Decreased	0·21	0·07	0·07
	Consistent with L.D.T.	0·76	0·78	0·88

(*b*) The frequency with which adjustment behaviour is consistent with the prediction of learning direction theory over rounds 1–4 of each session

	Rounds 1–2	Rounds 2–3	Rounds 3–4
½-Median game			
Session 1	0·47	0·80	0·73
Session 2	0·67	0·60	0·87
Session 3	0·46	0·77	0·62
Session 4	0·31	0·77	0·46
½-Mean game			
Session 5	0·75	0·63	0·83
Session 6	0·71	0·71	0·79
Session 7	0·60	0·80	0·53
Session 8	0·71	0·79	0·57
½-Maximum game			
Session 9	0·87	0·87	1·00
Session 10	0·80	0·73	0·93
Session 11	0·67	0·73	0·80
Session 12	0·73	0·87	0·73

Table 3 (*cont.*)

(*c*) Using data or rounds 5–10 from the single ten-round session of each treatment

	Adjustment ratio is	Rounds 4–5	Rounds 5–6	Rounds 6–7	Rounds 7–8	Rounds 8–9	Rounds 9–10
½-Median game guess was							
> ½ Ref. Pt.	Decreased	**0·62**	**0·69**	**0·77**	**0·38**	**0·69**	**0·62**
	Increased	0·23	0·00	0·15	0·46	0·08	0·31
< ½ Ref. Pt.	Increased	**0·15**	**0·31**	**0·08**	**0·08**	**0·15**	**0·08**
	Decreased	0·00	0·00	0·00	0·08	0·08	0·00
	Consistent with L.D.T.	0·77	1·00	0·85	0·46	0·84	0·70
½-Mean game guess was							
> ½ Ref. Pt.	Decreased	**0·57**	**0·50**	**0·07**	**0·21**	**0·43**	**0·36**
	Increased	0·14	0·36	0·00	0·64	0·21	0·21
< ½ Ref. Pt.	Increased	**0·21**	**0·14**	**0·07**	**0·14**	**0·36**	**0·43**
	Decreased	0·07	0·00	0·86	0·00	0·00	0·00
	Consistent with L.D.T.	0·78	0·64	0·14	0·35	0·79	0·79
½-Maximum game guess was							
> ½ Ref. Pt.	Decreased	**0·33**	**0·27**	**0·27**	**0·07**	**0·00**	**0·07**
	Increased	0·67	0·00	0·67	0·00	1·00	0·00
< ½ Ref. Pt.	Increased	**0·00**	**0·27**	**0·07**	**0·00**	**0·00**	**0·00**
	Decreased	0·00	0·47	0·00	0·93	0·00	0·93
	Consistent with L.D.T.	0·33	0·54	0·34	0·07	0·00	0·07

* Sum of the frequency numbers in boldface type which are consistent with the predictions of learning direction theory.

Table 3 (*c*) repeats the analysis of Table 3 (*a*) but for the remaining six round-to-round transitions in each of the ten-round sessions for each treatment. We see that when the number of rounds played is increased beyond four, learning direction theory may fail to predict the transition behaviour of the majority of subjects. Indeed, in the maximum game, an average of only 22·5% of adjustment behaviour over the last six rounds is consistent with the theory. However, in the median game, the behaviour of the majority of subjects remains consistent with the predictions of the theory in all but one transition period while in the mean game, the behaviour of the majority of subjects is consistent with the theory in all but two transition periods.[10]

These differences may be explained by the theory's assumption that subjects form their next round guesses conditional only on the outcome (the winning number) of the previous round. When the winning number of the previous round appears to be following a clear trend, then it is perhaps reasonable for subjects to condition their next round guesses solely upon the winning number of the previous round. This appears to be the case in the ten-round median game where extreme 'outlier' guesses have little influence, and consequently, the median guess and the winning number are monotonically decreasing over time. However, when the winning number of the previous round occasionally deviates from an established trend due to one or a few extreme guesses as in the

[10] A statistical test of whether learning direction theory is a good predictor of transition behaviour in the longer ten-round games would require that we have more than a single observation for each treatment.

ten-round version of the mean game, or when the winning number does not appear to follow any discernible trend, as in the ten-round session of the maximum game (see Fig. 3), then it may be less reasonable for subjects to form guesses that are conditional only on the winning number of the previous round. Thus, predictions such as those of learning direction theory that are based solely on the outcome of the previous round may continue to perform relatively well in those circumstances where aggregate behaviour adheres to a certain established trend, but may perform poorly otherwise.[11]

V. CONCLUSIONS

We have examined the robustness of behaviour across three different experimental treatments of the 'beauty contest' game. All three versions of the game have the same equilibrium prediction, namely that all players announce zero. Therefore, the differences in behaviour that we observe across the three treatments must be due to the various ways in which the problem is framed. We have suggested how the different ways of framing the 'beauty contest' game may lead to more or less coordination by groups of experimental subjects, or alternatively, how these different environments may encourage behaviour that is relatively more 'fundamental' or more 'speculative' in nature.

We find that behaviour in the mean and median games is similar with respect to first round choices, depth of reasoning and the direction of learning. However, we also find evidence suggesting that the percentage change in guesses over rounds 1–4 is significantly larger in the direction of zero in the median game than in the mean game. Furthermore, we observe that in the final rounds of the longer, ten-round median game session, subjects appear to be increasing their depths of reasoning whereas the same cannot be said for the mean game. We conclude that there is support for our hypothesis that guesses in the median game will be closer to the Nash equilibrium prediction as compared with guesses in the mean game.

We find even stronger support for our hypothesis regarding behaviour in the maximum game. First round choices in the maximum game are significantly higher than in either the mean or median games. Furthermore, percentage changes in guesses over rounds 1–4 are significantly smaller in the direction of zero in the maximum game as compared with either the mean or median games. Data from the ten-round sessions suggest that these differences between the maximum game and the mean and median games persist beyond the fourth round.[12]

[11] This explanation might well account for Ho *et al.*'s (1996) rejection of learning direction theory in their ten-round versions of the mean game using small group sizes.

[12] The relatively slower rate of convergence in the maximum game and our explanation for this finding may also account for Ho *et al.*'s (1996) finding that convergence to equilibrium in the mean game is slower with smaller group sizes of three players as compared with larger group sizes of seven players. In the smaller group size of three players, each individual player has a relatively greater influence on the value of the winning number. In this respect, the smaller group size treatment is most similar to our maximum game environment, where a single player determines the value of the winning number.

When we examine the direction of learning we find that the behaviour of the majority of subjects is consistent with the predictions of learning direction theory over the first four rounds of all three treatments and we are able to reject the null hypothesis that the predictions of learning direction theory are not useful in characterising adjustment behaviour in these early rounds. However, we find that learning direction theory frequently fails to predict the majority of subjects' actions in the later rounds of the maximum game, and is also somewhat less accurate in predicting behaviour in the mean game. This failure of learning direction theory in the later rounds of these games is likely due to the theory's use of the previous period's winning number as the reference point. Despite the limitations of learning direction theory, we believe that this theory serves as an important building block for more complex models of learning behaviour. It may be especially useful for characterising behaviour in the initial stages of games such as the 'beauty contest' game when players have little experience with the environment.

University of Pittsburgh

Universitat Pompeu Fabra

Date of receipt of final typescript: February 1997

References

Camerer, C. and Ho, T. H. (1996). 'Experience-weighted attraction learning in games: a unifying approach.' Working paper.

Ho, T. H., Weigelt, K. and Camerer, C. (1996). 'Iterated dominance and iterated best-response in experimental *p*-beauty contests.' Social Science Working Paper 974, California Institute of Technology.

Keynes, J. M. (1936). *The General Theory of Employment, Interest and Money*. New York: Harcourt Brace and Co.

Nagel, R. (1995). 'Unraveling in guessing games: an experimental study.' *American Economic Review*, vol. 85, no. 5, pp. 1313–26.

Selten, R. and Buchta, J. (1994). 'Experimental sealed bid first price auction with directly observed bid functions.' Discussion Paper No. B-270, University of Bonn.

Selten, R. and Stoecker, R. (1986). 'End behaviour in sequences of finite prisoner's dilemma supergames: a learning theory approach.' *Journal of Economic Behaviour and Organization*, vol. 7, no. 1, pp. 47–70.

Siegel, S. and Castellan, N. J. (1988). *Nonparametric Statistics for the Behavioural Sciences*. 2nd ed. New York: McGraw Hill.

Stahl, D. O. (1996). 'Boundedly rational rule learning in a guessing game.' *Games and Economic Behaviour*, vol. 16, pp. 303–30.

Tversky, A. and Kahneman, D. (1986). 'Rational choice and the framing of decisions.' *Journal of Business*, vol. 59, pp. S251–78.

Van Huyck, J. B., Battalio, R. and Beil, R. (1990). 'Tacit coordination games, strategic uncertainty, and coordination failure.' *American Economic Review*, vol. 80, no. 1, pp. 234–48.

Van Huyck, J. B., Battalio, R. and Beil, R. (1991). 'Strategic uncertainty, equilibrium selection principles, and coordination failure in average opinion games.' *Quarterly Journal of Economics*, vol. 106, pp. 885–911.