Economics Letters 137 (2015) 45–49

Contents lists available at ScienceDirect

Economics Letters

journal homepage: www.elsevier.com/locate/ecolet

Eliciting beliefs in beauty contest experiments

Yaron Lahav *

Department of Business Administration, Ben-Gurion University of the Negev, P.O. 653, Beer Sheva, Israel

HIGHLIGHTS

- We elicit beliefs from subjects in beauty contest experiments.
- Beliefs and choices of subjects are inconsistent.
- Subjects are more sophisticated when measured with beliefs rather than choices.
- Our results cast doubt on previous findings made with choices rather than beliefs.

ARTICLE INFO

Article history:
Received 12 May 2015
Received in revised form
17 August 2015
Accepted 26 August 2015
Available online 25 September 2015

JEL classification:
C9
D7
D8

Keywords:
Beauty contest
Belief elicitation
Higher order beliefs
Level of sophistication
Reasoning
Strategic play

ABSTRACT

Using elicited beliefs in several treatments of the beauty contest experiment, we show that (a) belief elicitation does not affect decisions, (b) beliefs and choices are inconsistent, and (c) players are more sophisticated compared to previous work.

1. Introduction

According to Keynes (1936), the stock market is no different from a beauty pageant, where observers try to guess which contestant will win. The trick to correctly guessing the pageant winner is to determine which contestant others are most likely to choose. Recognized as the beauty contest phenomenon, this mode of thinking is exploited by investors, who form second-order beliefs to predict the assets most likely to be purchased by others.

Research on second-order beliefs has mostly been experimental and typically entails asking each subject to pick an integer from the set $\{0, 1, 2, \dots, 100\}$, and the experimenter then calculates an average of the chosen numbers and multiplies it by p, where $0 < p \neq 1$. The closest number to the product wins. The unique Nash (everybody chooses "0") is rarely achieved, but average choices converge to it with repetition. Previous studies considered the chosen integers to be estimators of beliefs in average choices and in the choices of others. Consequently, most conclusions were driven by choices, rather than by actual beliefs.

We postulate that the beauty contest experiment cannot thoroughly assess or reflect players' beliefs. To investigate the relationship between beliefs, decisions, and accuracy, we elicit players' beliefs by using beauty contest experiments probabilistically. By employing a unique study design, we know not only what numbers the players choose, but also what they think other players will choose.

2. P-beauty contest experiments and second-order belief elicitation

The beauty contest experiment is inadequate for explaining beliefs and decisions for several reasons. First, existing beauty

* Tel.: +972 54 9107098.
E-mail address: ylahav@som.bgu.ac.il.

http://dx.doi.org/10.1016/j.econlet.2015.08.024
0165-1765/

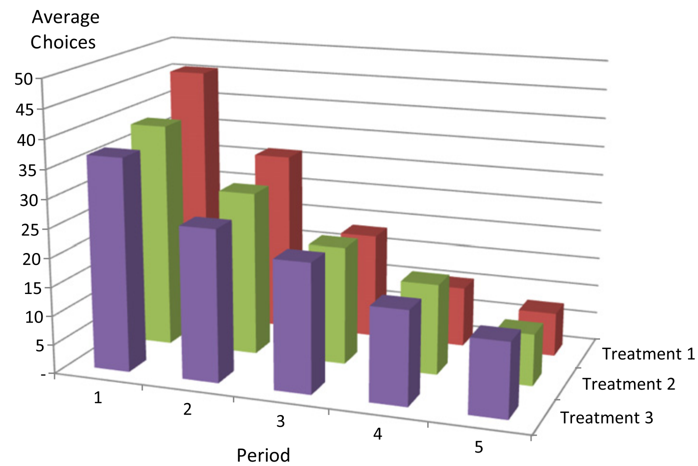

Fig. 1. Averages of choices per treatment.

contest experiments reveal nothing about how certain participants feel about their beliefs. Second, we cannot learn much about the beliefs of participants merely by observing their choices. And third, the compensation scheme ('winner takes all') is yet to be established as a viable way to elicit true beliefs. When eliciting beliefs in experiments, researchers should use an effective scoring rule (Selten, 1998). We employ the *quadratic* scoring rule, which is widely used in belief elicitation and is considered incentive-compatible (Selten, 1998), thus incentivizing participants to provide true beliefs.

Most studies on second-order beliefs are experimental and focus on differences in people's level of sophistication.[1] In recent works, however, researchers focus on how subjects react to variations in the experimental design, such as player heterogeneity (Guth et al., 2002), groups effect (Sutter, 2005; Kocher and Sutter, 2005), and two-player games (Grosskopf and Nagel, 2008).

3. Experimental design

In a computer laboratory, we use z-Tree (Fischbacher, 2007) to conduct nine experimental beauty contest sessions, as described above. Each session comprised five repetitions (periods) and a maximum of 20 participants, who were undergraduate students from various disciplines.

Sessions were divided evenly between three different treatments. In the control treatment, each period began with each participant choosing a number between 1 and 100, but no beliefs were elicited probabilistically. The second treatment began like the control treatment, but the participants were then asked to guess how many of the other participants they believe chose numbers in each of the following intervals: 0–10, 11–20, 21–30, . . . , 91–100. The third treatment was similar to the second treatment, but in each period, after the choice frequencies were submitted, participants could change the number that they initially chose.

The prize for making an accurate choice was 20 experimental points, split between all winners. In addition, each participant received compensation for accurate beliefs as follows:

Compensation

$$= 3 - \sum_{m=1}^{10} \left(\frac{Actual\ Frequency\ in\ m - Expected\ Frequency\ in\ m}{Number\ of\ Players} \right)^2$$

where m represents a choice interval. This formula is non-negative and represents the quadratic scoring rule.

[1] See Nagel (1995), Duffy and Nagel (1997), Camerer et al. (2004), Costa-Gomes and Crawford (2006), Grosskopf and Nagel (2008), and others.

Table 1
T-tests for periodic differences in choices between treatments.

Period	Differences between treatments		
	1 vs. 2	1 vs. 3	2 vs. 3
1	6.34	8.70*	2.36
2	2.98	5.22	2.24
3	2.31	3.88	1.57
4	5.20*	5.60*	0.40
5	1.44	5.22	3.78

* Differences are significant at the 5% level.

4. Research hypotheses

Our first hypothesis regards the possible effects that belief elicitation may have on participant choices:

H1: *The process of belief elicitation does not affect choice. Accordingly, choices in all treatments should not be significantly different from each other.*

The second hypothesis tests the accuracy of participants' beliefs by comparing them with their actual choice:

H2: *Participants' choices are consistent with their beliefs.*

The third hypothesis tests whether level of sophistication, as measured by choices, is consistent with that based on their beliefs:

H3: *Sophistication level based on choices is consistent with that based on beliefs.*

5. Results

As shown in Fig. 1, the average choices decrease with experience. Three sets of five independent t-tests for differences in average choices between treatments in each period (Table 1) show that only three differences are significant at the 5% level and that none of these differences is between treatments 2 and 3. We, therefore, accept hypothesis H1 and conclude that belief elicitation does not affect subsequent choices.

Testing for consistency[2] between choices and beliefs, we estimate the calculated belief of participant i in period t as follows:

$$\bar{B}_i^t = \sum_{k=1}^{10} \tilde{k} * n_k$$

[2] Inconsistency between choices and beliefs has been investigated experimentally in previous studies. Examples are Costa-Gomes and Weizsacker (2008) and Sutter et al. (2013).

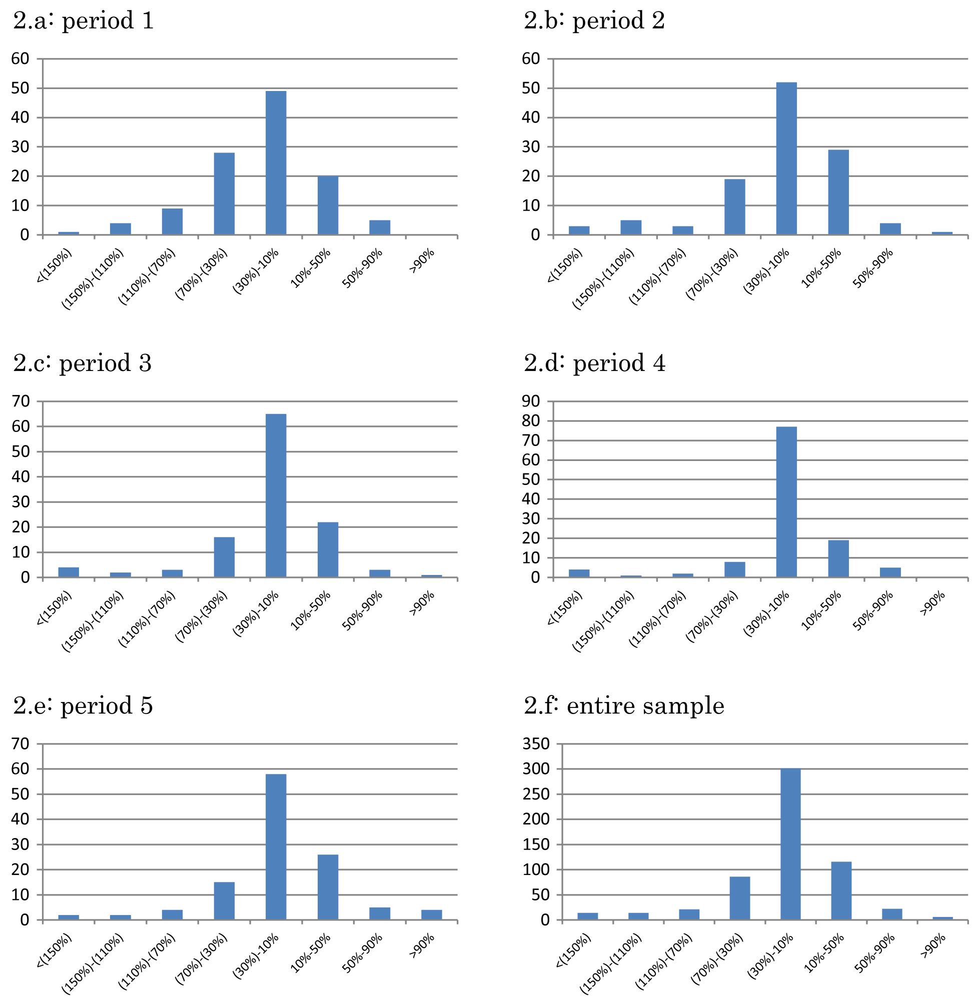

Fig. 2. Frequencies of normalized differences between choices and beliefs per period and over all periods.

where k represents the ten intervals $\{0\text{–}10, 11\text{–}20, \ldots, 91\text{–}100\}$, n_k is the number of participants that participant i in period t believes will choose a number within interval k, and $\tilde{k}$ is the average of possibilities in interval k ($\tilde{k} = \{5, 15.5, 25.5, \ldots, 95.5\}$). We then measure the normalized difference between beliefs and choices of participant i during period t as follows:

$$\varphi_i^t = \left(\frac{2}{3}\bar{B}_i^t - C_i^t\right) \Big/ \bar{C}^t$$

where C_i^t is the actual choice of participant i during period t and $\bar{C}^t$ is the average of choices during period t. Results pooled from treatments 2 and 3 are shown in panels (a)–(f) of Fig. 2. Column 1 of Table 2 shows the average $\bar{\varphi}^t = I^{-1}\sum_i \varphi_i^t$ for each period. If beliefs and choices are consistent, then these averages should not differ significantly from zero.[3] With the exception of period 5, all averages in column 1 are significantly different from zero, implying that, on average, participants' beliefs are consistently higher than their actual choices.

It is possible that, as a result of misunderstanding the rules of the experiment, participants guess the average instead of two-thirds of it. To test this possibility, we calculate a new measure as follows:

$$\delta_i^t = \left(\bar{B}_i^t - C_i^t\right) / \bar{C}^t.$$

Table 2
Periodic averages of differences and normalized differences between choices and beliefs.

Period	Column 1	Column 2
1	(19.8%)*	20.3%*
2	(14.8%)*	27.8%*
3	(17.4%)*	23.8%*
4	(19.4%)*	20.9%*
5	(10.0%)	35.1%*

* Numbers are significant at the 5% level.

Column 2 of Table 2 presents the calculation of $\bar{\delta}^t = \sum_i \delta_i^t$. All averages are positive and significantly different from zero. Therefore, we conclude that beliefs lie between perceived averages and actual choices of the participants. Accordingly, we reject hypothesis H2.

Hypothesis H3 posits that participants with lower levels of sophistication tend to choose higher numbers. A participant's level of sophistication is thus measured by what he believes to be the lowest number that other participants will choose. For example, in period 1, Level-1 participants do not believe that others will choose 30 or lower. In period $t > 1$, the actual average of the previous period is the new reference for all sophistication levels.[4] Although in periods 2–5 it is possible that an exact level of sophistication will not be found, some conclusions can be made.

[3] We are aware that this is an approximation, but we do think that errors can be considered random.

[4] Consistent with Nagel (1995) or Camerer et al. (2004).

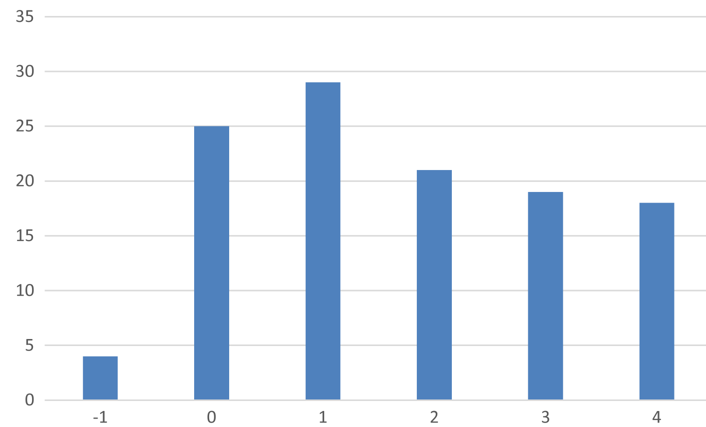

Fig. 3. Distribution of the differences between estimated and elicited level of sophistication.

Table 3
Frequencies of level of sophistication measured with elicited beliefs per session and period.

<table>
<tr><th>Treatment/session</th><th>Period</th><th colspan="5">Level of sophistication</th></tr>
<tr><th></th><th></th><th>0</th><th>1</th><th>2</th><th>3</th><th>4</th></tr>
<tr><td>2/1</td><td>1</td><td>1</td><td>2</td><td>3</td><td>7</td><td>7</td></tr>
<tr><td></td><td>2</td><td>1</td><td>5</td><td colspan="2">7</td><td>7</td></tr>
<tr><td></td><td>3</td><td>0</td><td>2</td><td>3</td><td colspan="2">15</td></tr>
<tr><td></td><td>4</td><td>0</td><td colspan="2">2</td><td colspan="2">18</td></tr>
<tr><td></td><td>5</td><td colspan="2">1</td><td colspan="3">19</td></tr>
<tr><td>2/2</td><td>1</td><td>0</td><td>3</td><td>2</td><td>2</td><td>10</td></tr>
<tr><td></td><td>2</td><td>0</td><td>1</td><td>3</td><td colspan="2">13</td></tr>
<tr><td></td><td>3</td><td>0</td><td colspan="2">1</td><td colspan="2">16</td></tr>
<tr><td></td><td>4</td><td colspan="2">2</td><td colspan="3">15</td></tr>
<tr><td></td><td>5</td><td>0</td><td colspan="4">17</td></tr>
<tr><td>2/3</td><td>1</td><td>0</td><td>1</td><td>3</td><td>6</td><td>9</td></tr>
<tr><td></td><td>2</td><td>0</td><td>2</td><td colspan="2">10</td><td>7</td></tr>
<tr><td></td><td>3</td><td>0</td><td colspan="2">7</td><td colspan="2">12</td></tr>
<tr><td></td><td>4</td><td>1</td><td colspan="4">18</td></tr>
<tr><td></td><td>5</td><td colspan="5">19</td></tr>
<tr><td>3/1</td><td>1</td><td>1</td><td>0</td><td>2</td><td>5</td><td>12</td></tr>
<tr><td></td><td>2</td><td>0</td><td>2</td><td colspan="2">4</td><td>14</td></tr>
<tr><td></td><td>3</td><td>0</td><td>0</td><td>4</td><td colspan="2">16</td></tr>
<tr><td></td><td>4</td><td>0</td><td colspan="2">1</td><td colspan="2">19</td></tr>
<tr><td></td><td>5</td><td colspan="2">3</td><td colspan="3">17</td></tr>
<tr><td>3/2</td><td>1</td><td>2</td><td>2</td><td>6</td><td>5</td><td>5</td></tr>
<tr><td></td><td>2</td><td>0</td><td>2</td><td>7</td><td colspan="2">11</td></tr>
<tr><td></td><td>3</td><td>0</td><td>3</td><td colspan="3">17</td></tr>
<tr><td></td><td>4</td><td>0</td><td>0</td><td colspan="3">20</td></tr>
<tr><td></td><td>5</td><td>0</td><td>0</td><td colspan="3">20</td></tr>
<tr><td>3/3</td><td>1</td><td>2</td><td>3</td><td>6</td><td>3</td><td>6</td></tr>
<tr><td></td><td>2</td><td>1</td><td>3</td><td>7</td><td colspan="2">9</td></tr>
<tr><td></td><td>3</td><td>0</td><td colspan="2">4</td><td colspan="2">16</td></tr>
<tr><td></td><td>4</td><td>0</td><td>0</td><td colspan="3">20</td></tr>
<tr><td></td><td>5</td><td>0</td><td colspan="4">20</td></tr>
</table>

Table 3 provides the number of participants in each period, categorized according to the level of sophistication. First, during period 1, the majority of players are characterized as level-4 in five of six sessions. Furthermore, in sessions where data for level-0 players are available, they do not exceed two players (one player for $t > 1$). Data for level-4 players are available in 11 periods, in nine of which those players are the majority.

To calculate what we refer to as "the measured level of sophistication", we use the model of Nagel (1995) by fitting choices into intervals bounded by a geometric average of $50\,(2/3)^n$ and $50\,(2/3)^{n+1}$, where n is the level of sophistication. For each participant, we compare this measure to the "elicited level of sophistication". We limit our measurements and comparison to period 1, where the level of sophistication of the participants can be accurately assigned.

Fig. 3 presents the distribution of the differences between measured and elicited numbers. The measured numbers are biased for 78% of the participants (91 of 116), and are positively biased for 75% of the participants. These findings cast doubt on any previous findings acquired with measured numbers.

6. Conclusion

We conduct beauty contest experiments with belief elicitation. In contrast to previous research, we show that choices do not reflect actual beliefs. First, we show significant difference between what participants think about the choices of other participants and what experimenters conclude about their beliefs using actual choices. We also show that participants are, on average, more sophisticated than assumed, and that naïve play is rare.

Second, we find inconsistencies between participants' choices and beliefs. We reject the possibility that participants fail to calculate two-thirds of the average.

The first finding casts doubt on previous findings that used actual choices to estimate level of sophistication. The second finding casts doubt on the consistency of people's choices and beliefs. Whether this is a behavioral anomaly or a misunderstanding of the instructions should be further investigated.

Acknowledgment

The author wishes to thank the Marie Currie International Re-Integration Grants Program (IRG) for the financial support (grant FP7-PEOPLE-2010-RG, No. PIRG07-GA-2010-268360).

References

Camerer, C.F., Ho, T., Chong, J., 2004. A cognitive hierarchy model of games. Quart. J. Econ. 119, 861–898.
Costa-Gomes, M., Crawford, V.P., 2006. Cognition and behavior in two-person guessing games: An experimental study. Amer. Econ. Rev. 96, 1737–1768.
Costa-Gomes, M.A., Weizsacker, G., 2008. Stated beliefs and play in normal-form games. Rev. Econom. Stud. 75, 729–762.
Duffy, J., Nagel, R., 1997. On the robustness of behaviour in experimental 'beauty contest' games. Econ. J. 107, 1684–1700.
Fischbacher, U., 2007. Z-Tree: Zurich tool-box for ready-made economic experiments. Exp. Econ. 10, 171–178.
Grosskopf, B., Nagel, R., 2008. The two-person beauty contest. Games Econom. Behav. 62, 93–99.
Guth, W., Kocher, M., Sutter, M., 2002. Experimental 'beauty contests' with homogenous and heterogeneous players and with interior and boundary equilibria. Econom. Lett. 74, 219–228.
Keynes, J.M., 1936. The General Theory of Interest, Employment and Money. Macmillan, London.
Kocher, M.G., Sutter, M., 2005. The decision maker matters: individual versus group behaviour in experimental beauty-contest games. Econ. J. 115, 200–223.

Nagel, R., 1995. Unraveling in guessing games: An experimental study. Amer. Econ. Rev. 85, 1313–1326.
Selten, R., 1998. Axiomatic characterization of the quadratic scoring rule. Exp. Econ. 1, 43–62.
Sutter, M., 2005. Are four heads better than two? An experimental beauty-contest game with teams of different size. Econom. Lett. 88, 41–46.
Sutter, M., Czermak, S., Feri, F., 2013. Strategic sophistication of individuals and teams. experimental evidence. Eur. Econ. Rev. 64, 395–410.