Documento de Trabajo/
Working Paper

IESA 16-04

Communication, coordination and competition in the beauty contest game: Eleven classroom experiments

Virtudes Alba Fernández
Universidad de Jaén y CentrA

Pablo Brañas Garza
Universidad de Jaén, IESA-CSIC y CentrA

Francisca Jiménez Jiménez
Universidad de Jaén y CentrA

Javier Rodero Cosano
CentrA

manuscript No.
(will be inserted by the editor)

Communication, coordination and competition in the beauty contest game: Eleven classroom experiments*

Virtudes Alba Fernández[1], Pablo Brañas Garza[2], Francisca Jiménez Jiménez[3], Javier Rodero Cosano[4]

[1] Dep. Statistics, University of Jaén and centrA
[2] Dep. Economics, University of Jaén, IESA-CSIC and centrA
[3] Dep. Economics, University of Jaén and centrA
[4] centrA and LINEEX

The date of receipt and acceptance will be inserted by the editor

Abstract This paper introduces some new features in the standard experimental design of the beauty contest in order to allow communication among participants. With that aim, we use the mode instead of the mean and non-rival payoffs. This design encourages students to communicate their guessed number, with a higher probability if subjects know the 0 Nash equilibrium. The lack of communication can only be explained by subjects endowed with competitive other-regarding preferences. Experiments are run in 11 classrooms ranging from 11 to 60 students in size. Participants are given at least one week to submit their guesses and a questionnaire explaining their choice. Results indicate that: i) communication induces coordination in the responses, ii) communication does not guarantee any improvement in the average reasoning level, iii) there exist significative differences according to classroom size and duration of degree.

J.E.L. Classification: C91, D83. I21

Key words Beauty Contest Games — Experiments — Communication — Coordination — Competition — Meta-analysis

* The authors are grateful to Rosemarie Nagel, Rosario Gómez, Amparo Urbano, Karl Schlag, Steven Brams and João Gata and the participants at the Lisbon Workshop on Game Theory 2001, the Twelfth International Conference on Game Theory at New York, the Economic Science Association Conference at Barcelona and the VII Encontro de Novos Investigadores de Anàlisis Económica at Santiago de Compostela and a anonymous referee for their helpful suggestions; our thanks to Ana & Pili López López for their contribution to data gathering and processing and Martha Gaustad for the language revision. We gratefully acknowledge the financial support from the University of Jaén R+D program (# 20210/148).
Correspondence to: Javier Rodero. centrA. C/Marqués de Larios, 3. 29015. Málaga. Spain. tel. + 34 951 03 96 10 e-mail: `jrodero@fundacion-centra.org`

1 Introduction

Over the last decade, the Beauty Contest Game (BCG hereafter) has been successfully used as an experimental tool in Economics. Despite its seemingly *ex-ante* complexity, it turns out to be an interesting game to study how subjects learn specific rules due to its *ex-post* simplicity.

BCG is a simple guessing game that facilitates the evaluation of individuals' level of reasoning. The basic BCG is as follows: a certain number of subjects are invited to play a game in which all of them must simultaneously choose a number from an interval (generally $[0,100]$). The winner is the player whose number is closest to p-times the mean of all the numbers chosen, where $0 < p < 1$. The winner (one of them if there are several) receives a fixed prize, the losers get nothing.

In an attempt to understand how people solve this game, two basic theoretical explanations can be given: the iterative elimination of dominated strategies and best response behavior. Empirical evidence provided by many experimental investigations [see for example Nagel, 1995, Ho et al., 1998, Bosch-Domènech et al., 2002] suggests that: i) few people know the zero solution (the Nash equilibrium), ii) the answers are widely spread on the interval and, iii) the average reasoning level rarely exceeds three iterations.

Since the main purpose of prior experimental research has been to study how many iterated levels subjects actually apply, BCG experiments are often run with isolated individuals. However, in real life situations people do not make decisions in an isolated setting such as a laboratory. Instead, they communicate among themselves, exchange their opinions and then make decisions. Therefore, many decisions are made in coordination with other individuals.

This paper addresses a new issue: how communication among subjects affects BCG findings. In our investigation, we let students communicate with one another before eliciting their choices in classroom experiments.

Given the peculiarities of BCG, the communication factor is not simple to implement. *Effective* communication requires giving the right incentives so people will say their true guessed numbers and coordinate with one another to come up with a single answer. For example, if the payoffs are rival, letting subjects communicate with each other may not be enough.

Consider the following example to illustrate the problem of incentives to induce communication in our context:

> John, Adam and Peter's father proposes that they play the following game: if at least two (the mode) of them say the same supermarket name, then the father will buy one coke for each winning son; if not, he will choose one randomly and will buy him a coke.

What will John, Adam and Peter do? If it is a simultaneous game[1] with pre-communication, the three brothers will say some names of supermarkets (strategies) and will try to coordinate at one of them. Assuming rationality (self-interest)

[1] Coordination is immediate in a sequential game. If John, for example, chooses strategy A first, then Adam will also select strategy A to coordinate. Consequently, if Peter wants to get the coke, he should choose the same strategy as his brothers.

and zero communication costs, the best response is the truth-telling strategy, that is, tell your opponent the strategy that you are going to use. If, in contrast, the communication costs are high, coordination may be difficult to achieve.

But even with costless communication, why would one of the brothers not be informed about the true strategy chosen by the others? It is obvious that two people are sufficient enough to constitute a winning coalition. Assume that both partners consider not only their own payoffs, but also their outsider brother's payoff in their utility function. This last argument affects their preferences negatively in the sense that they prefer their brother not to get the coke. That is, subjects are endowed with other-regarding preferences, which are not altruistic *a la* Fehr and Schmidt [1999], but envious, as these individuals like advantageous inequality.

What would happen if the number of brothers was very large, for example 15, and the winners were those who choose the most voted option? Either you get great pleasure from inequality or you will try to convince everybody (the more the better) that your strategy is the best or the most popular or use any other argument to increase your chances of winning. In this case, inequality-seeking would be the sole reason for not giving someone the relevant information (if communication costs are zero).

Therefore, in order to achieve this communication effect, we modify the game in two ways: 1) the mode instead of the mean is used as the order statistic and 2) non-rival payoffs are given. This means that the winner will be the person whose number is the closest to p-times the mode and all winners will get a fixed prize.

Why use the mode? Because we are looking for collusive behavior among students. When the mode is the reference, they have a clear incentive to coordinate: the larger the coalition the greater the probability of winning. This variation is an extension on the existing experimental literature on BCG.

It is important to highlight the fact that we use non-rival payoffs. This is because we are interested not only in analyzing self-interest behavior (the absolute position), but also the relative position in the group (other-regarding preferences). Let us suppose a weak equilibrium that is different from zero[2] . Under non-rival payoffs, we would find coordination at any number different from zero only in the case in which students are concerned solely about their own marks. In contrast, individuals endowed with competitive[3] other-regarding preferences would not communicate their true choice and betray the coalition (at a number different from zero).

Moreover, if any rational student knows the theoretical solution of the game, that is 0, she should communicate it to everyone to avoid the possible trembling-hand problem. Note that any weak equilibrium (even with non-rival payoffs) may suffer from hand trembling.

The experiments were run in 11 sessions held on the same day at the University of Jaén, Spain. All of them were conducted in the classroom by the same professor.

[2] Any number in the interval could be a weak equilibrium if all subjects choose the same number and they show self-interest behavior.

[3] Note that with rival payoffs the incentives to betray the coalition are obvious.

After the sessions, participants were asked to fill in a short questionnaire to shed some light on how they had reached the chosen number.

As we will see below, COMMUNICATION gives rise to a spectacular result: most of the guessed numbers are concentrated in one single number, that is, there is COORDINATION. However, this number is not always the Nash equilibrium. The statistical analysis, using meta-analysis, provides us with an interesting explanation for these discrepancies: sometimes, although the proportion of subjects who know the theoretical solution is quite low, they communicate it to everybody and coordination at zero is produced. Other times, however, there is a high percentage of individuals who know the right answer but do not communicate it, that is, there exists COMPETITIVE other-regarding preferences.

This paper is structured as follows: the theoretical and experimental research on BCG is summarized in the next section, while the experimental design and procedures are described in detail in the third section. In the fourth section, the principal results are presented and the data is analyzed using an innovative statistical technique in experimental economics: meta-analysis. Lastly, conclusions are reached in section five.

2 The beauty contest game: theoretical and experimental background

The original idea behind the BCG was first mentioned by Keynes [1936] in an attempt to express that a clever investor has to "anticipate the basis of conventional valuation a few months hence, rather than . . . over a long term of years" (pg. 155) so he can act in the stock market before other investors do. As explained in the introduction, the formal game model was introduced by Moulin [1986].

In the standard game ($0 < p < 1$, mean), the unique Nash equilibrium is 0. This solution can be reached two ways: i) by a best-response argument and ii) by the iterated elimination of dominated strategies.

Best response may be summarized as follows. Given that each subject has to submit a number between 0 and 100, the distribution of the chosen numbers lets us analyze the depth of reasoning of the population. Following Stahl [1996], we classify people into different levels. Level 1 includes people who expect that the other players will behave randomly so they choose $p * mean$ ($mean = 50$ if the choice distribution is uniform), level 2 contains people who expect that the other players' depth of reasoning is level 1 and thus choose $p^2 * mean, \ldots$. Generalizing, level K people are those who choose $p^K * mean$ because they believe that the other people are at level $K-1$. If K is large, $p^K * mean \simeq 0$, then if we repeat the process *ad infinitum* ($K = \infty$) we reach the theoretical solution[4], 0, the highest level of reasoning (figure 1 represents this process with $p = 2/3$). Random answers are called level 0 of reasoning.

[4] Although all answers (numbers) are possible game solutions (weak equilibria), if *all* the subjects choose the same number, only 0 is a Nash equilibrium.

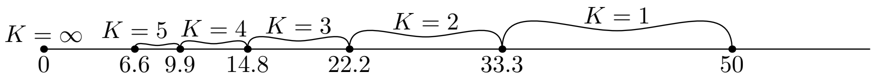

Fig. 1 An example of different reasoning levels

Figure 2, taken from Ho et al. [1998, pg. 951], also shows the convergence to the zero solution, but from a *dominance iterative* point of view[5]. Any number chosen between 66.7 and 100 is dominated by 66.6 ($100 * 2/3$, if $p = 2/3$), so they say that the interval $[66.7, 100]$ corresponds to irrational behavior ($R(0)$ for us). Rational individuals will always choose a number in the $[0, 66.6]$ interval. Applying the same reasoning, $R(1)$ players will choose a number below 66.7 (but above 44.4). Since 44.4 will again dominate any number between 44.4 and 66.7, we say that any number below 44.4 (and above $44.4 \cdot 2/3 = 29.6$) corresponds to a $R(2)$ individual. Following this iterative reasoning level process *ad infinitum*, we reach the theoretical Nash equilibrium (0, with $R(\infty)$). In game theory this process is called the iterated elimination of dominated strategies. Therefore, this game is dominance solvable[6].

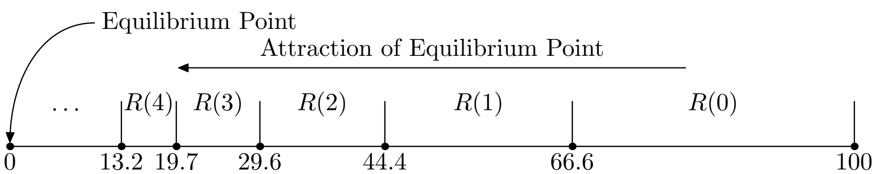

Fig. 2 Iterated reasoning of individuals eliminating dominated strategies

In very recent years some experimental researchers have opened up a new area of investigation regarding BCG or *p-beauty* [see Ho et al., 1998]. After the seminal papers by Nagel [1994, 1995], intensive research into BCG was begun. This research covers many different variations of the game [see Nagel, 1998, for a survey], some of which include:

- Subject pool: sometimes subjects are students, other times professors or newspaper readers —who presumably have different levels of education [see for instance Bosch-Domènech et al., 2002].
- Payoffs: a fixed amount of money, a variable amount of money, even a present (a beautiful *pteridium aquilinum* in our preliminary test or airplane tickets Bosch-Domènech et al. [2002] in the case of the newspaper experiment).

[5] See Rapoport and Amaldoss [2000] for an experimental study of iterative elimination of (strongly) dominated strategies.

[6] There is even a superior rationality level. An individual who knows the zero-solution by any of these methods can guess that most people would not reach the zero solution, thus the “intelligent” individual will link her answer to her estimation of the average rationality level. We call this phenomenon ∞-plus reasoning level to differentiate it from those individuals who find the theoretical solution, but fail to consider the importance of other people’s decisions [see also Alba-Fernández et al., 2003].

- Time to answer: immediate answers in the lab [Ho et al., 1998] versus long-time answers in post-card experiments [Bosch-Domènech and Nagel, 1997].
- One-shot vs. repeated games [see Ho et al., 1998, Alba-Fernández et al., 2003].
- Order statistic used: usually it is the mean, but Duffy and Nagel [1997] use the median and the maximum.
- Several feed-back learning environments [see Weber, 2003].

All these experiments bring some facts to light: individual level of reasoning is rarely larger than 3 [see Bosch-Domènech et al., 2002] and Iterated Best Replay Behavior (IBRB) is quite common [Ho et al., 1998].

3 Design, procedures and preliminaries

3.1 Design: A BCG with communication

Our experimental design expands upon the previous literature in a number of ways:

- Subjects are given a long time span to answer. Instructions were given and explained during one normal class period. The students were then told to bring back their answers along with a questionnaire either one, two or three weeks later (see below for details). Although students were allowed to communicate (freely), they were not explicitly encouraged to do so.
- Payoffs are not rival. All winners received 0.5 points towards their final grades, regardless of how many students got the right answer.
- Mode is used as the order statistic.

As Kocher and Sutter [2001] emphasize, most papers on beauty contest games only study individual decision making. However, considering a BCG with COMMUNICATION is not completely new: Kocher and Sutter [2001] have also studied interaction among individuals, although they use teams as a decision maker. In contrast, in our experiment we let students communicate freely among themselves and then choose their own strategy. Observe that the objective of the author above is to test if groups learn faster than isolated individuals.

In the standard BCG, with rival payoffs, individuals have no way and no incentives to collude since this reduces their own probability of winning. Moreover, when subjects value their payoffs more than those of their rivals, there is no room for altruistic behavior [Ho et al., 1998] and all individuals have incentives to betray their comrades. The only exception would be the perfect equilibrium, 0 in our case. Note that this would mean that someone would have solved the game and would like to share the prize with the other subject(s).

For this reason we introduce NON-RIVAL PAYOFFS. Under the assumption that there are self-interested students, any number in the interval could be a (weak) equilibrium. Moreover, if everyone coordinated at the same number (with costless communication), this would imply purely collusive behavior and could be reduced to a matching game [see Camerer, 2001].

Nevertheless, the design continues to cause certain problems since the mean generates some noise: an individual choosing 0 may not be sure of winning because other individuals might not be intelligent enough to calculate the equilibrium and, therefore the order statistic could be greater than 0. The only way to ensure victory is to have the correct information about other people's reasoning levels. But since this is not public information, people have to collude if they want to ensure that they will win. The use of the MODE as the order statistic strengthens this phenomenon, as it reduces the possible impact of an individual defection or any other noise.

To sum up, under this BCG with mode, non-rival payoffs and costless communication, if any subject knows the right answer of 0 and does not share it with other individuals —even reducing her own probability of winning by doing so— we could thus infer that she negatively values other subjects' payoffs in her own utility function. Therefore, this subject can be said to behave competitively.

Consequently, there is a unique equilibrium solution: zero. Any attempt at coordination in any other number would imply both the chance of betrayal by any individual (and the failure of the coalition) and the trembling-hand problem.

The coordination in any number different from zero can only occur if the subjects are purely altruistic or, at least not-envious and that they are completely rational who believe that the other subjects will be also perfectly rational. In contrast, students who care about their relative position will always try to outperform the rest of the class by choosing a smaller number. If there is at least one individual like this, the only possible equilibrium is 0.

3.2 Procedures

The experiment was conducted in classrooms at the University of Jaén in May 2000. The subject pool was comprised of typical university students. The date of the experiment was selected randomly and students were given no prior notification.

Instructions were explained aloud by the same professor in all the groups. The same procedure was followed with all the groups:

1. Participants were asked to choose a real number in the interval $[0,100]$.
2. The winner was the person (or persons) whose number was closest to 2/3 of the mode of the guessed numbers.When more than one mode existed, all of them were used to compute the different reference points and determine all the winning numbers. The subjects were given an extra-clue: they were told that they would realize when they reached the solution. Payoffs for all the winners were given in the form of 0.5 extra-points towards the final grade in the courses selected to run the experiment[7].
3. Students were required to submit their choices in written form by a given date. They were also asked to fill out a questionnaire regarding their process of reasoning (see appendix A.1).

Although communication was not explicitly encouraged, our time span obviously gave students the chance to communicate with each other. The total sample in our experiment was divided into three treatments according to answering

[7] As we needed an academic reward for the experiment, we required the collaboration of colleagues willing to give some extra points to the winners.

time: one, two or three weeks. All our experimental subjects were first year undergraduate students from the different schools. All the possible classifications are summarized in Table 1.

Table 1 Experimental subjects' features

code	*school*	*time*	*n*	*years*
B1–A	Business Studies	1 week	60	5
B1–B	Business Studies	2 weeks	44	5
B1–C	Business Studies	3 weeks	32	5
L1	Law	1 week	11	5
A1–A	Labor Affairs	1 week	44	3
A1–B	Labor Affairs	2 weeks	27	3
M1–A	Management	1 week	18	3
M1–B	Management	2 weeks	50	3
M1–C	Management	1 week	21	3
P1–A	Public Administration	1 week	32	3
P1–B	Public Administration	2 weeks	41	3

where *code* is the classroom label, *school* is the specific name of the program (Spanish), *time* is the period given to answer, *n* is the group size, and *years* the duration of the degree .

3.3 Preliminary test

As the mode is not usually used in BCG experiments, we performed a preliminary test to study the possible impact of this variation[8]. With that aim, we used a similar design to Bosch-Domènech et al. [2002], but by e-mail. In this "standard" experiment, participants were colleagues from the Economics Departments of the Universities of Jaén, Jaume I, Vigo, Autónoma of Barcelona and Carlos III of Madrid[9].

Figure 3 (the k value in the graph corresponds to the mean level of rationality, see below) summarizes the results. As can be seen, the results are very similar to those shown in the previous literature. These participants reached an average of 34.7, that is, a $k = 1$ reasoning level, even though the mode was 0. Note that since each bar of the histogram comprises an interval, the mode does not have to coincide with the highest peak. The same applies for the rest of the figures.

Therefore, introducing the mode as the reference in a BCG without communication did not produce a significant effect on the responses. Consequently, we

[8] That is, we did not allow communication.

[9] By e-mail (to implicitly discourage communication), 43 unexperienced experimental subjects were given the instructions and asked to return their responses. At the end of the experiment, each of the five winners were "paid" a beautiful plant.

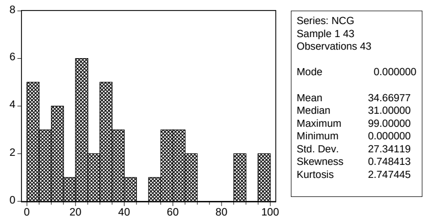

Fig. 3 Non-Communication BCG ($k = 1$)

analyzed the joint influence of the mode and communication factors in the following.

4 Results

4.1 First analysis

In this section we will compare the data set from our (eleven) sessions. First of all, we will show the immediate implications of allowing communication in our experiment. Figures 4 to 13 explore the different communication games. Group L–1 is not shown since all students chose 0. Remember that the previous figure 3 plots the non-communication game.

Initially, we can see clear differences in the dispersion of the responses. In contrast to the standard BCG experiments, our observations focus on some particular number (zero or any other one) in our communication sessions.

However the average level of reasoning is not homogeneous at all! We obtain levels from 0 to infinity:

- $k = 0$ for A1–B (mean 50.6),
- $k = 3$ for groups B1–A, B1–B and A1–A[10] (means 20.6, 21.6 and 21.5, respectively),
- $k = 4$ for P1–A (mean 12.5),
- $k = 5$ for B1–C and P1–B (means 8.1 and 7.7),
- $k = 6$ for M1–A (mean 5.0),
- $k = 7$ for M1–B (mean 4.2),
- $k \rightsquigarrow \infty$ for M1–C (mean 0.2) and,
- $k = \infty$ for L1.

Only one classroom shows a lower reasoning level than the former NCG. Note that our NCG showed an average of 34.7, that is, a reasoning level of $k = 1$.

From figures 4 to 13 we can infer that students coordinate in some number and that such a coordination leads to two different behavior patterns: i) to reach the Nash equilibrium (zero) and ii) to get a weak equilibrium (any number different

[10] Given that the coordination number was chosen randomly, this group must be properly considered as $k = 0$.

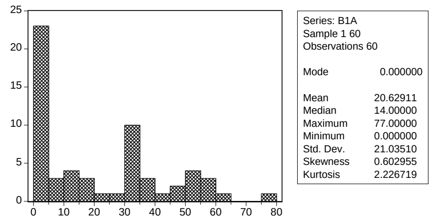

Fig. 4 Results from B1–A ($k = 3$)

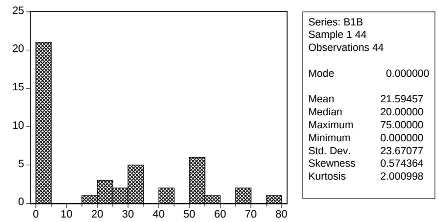

Fig. 5 Results from B1–B ($k = 3$)

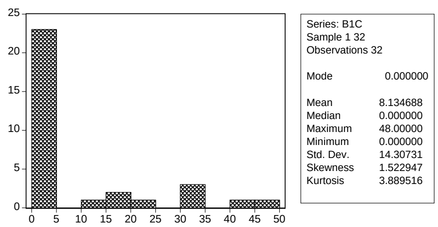

Fig. 6 Results from B1–C ($k = 5$)

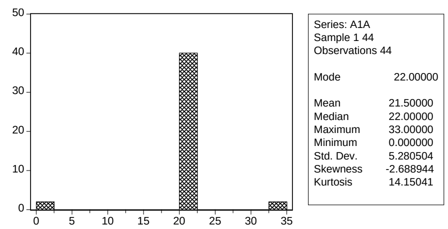

Fig. 7 Results from A1–A ($k = 3$)

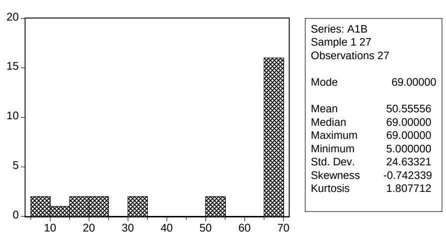

Fig. 8 Results from A1–B ($k = 0$)

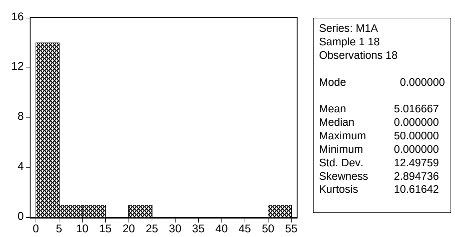

Fig. 9 Results from M1–A ($k = 6$)

from zero). Our students coordinated in the theoretical solution in 9 out of 11 sessions and colluded in another number in the remaining 2.

First, we will analyze the behavior observed in the two classrooms in which the participants did not collude in the Nash equilibrium. Figure 7 (A1-A group) shows an interesting and risky behavior. One person became a "leader". When we went to collect the answers, she cried out "write down 22" and most of the people followed her. Nobody noticed that $2/3 \cdot 22$ was a better option. Only two people got the right theoretical solution, but 22 was closer to $2/3 \cdot 22$ than 0; so all the students who said 22 won the game!

Figure 8 is even more interesting. Most of the class (25 out of 27 students) decided to concentrate at a point (69) far from any focal position. But two people chose (randomly) 50. Thus, they acted like involuntary "smart" traitors, thereby winning the game. Perhaps it is even more interesting to note that the vast majority of the sample coordinated in a dominated strategy.

Following Kocher and Sutter [2001, pg. 6], we can say that group behavior is not necessarily better than individual behavior. They offer two reasons for this:

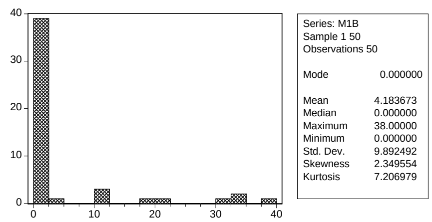

Fig. 10 Results from M1–B ($k = 7$)

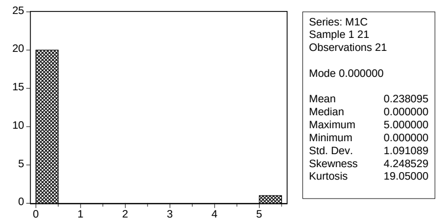

Fig. 11 Results from M1–C ($k \to \infty$)

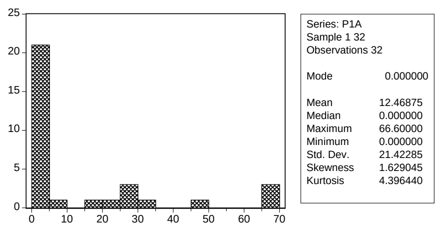

Fig. 12 Results from P1–A ($k = 4$)

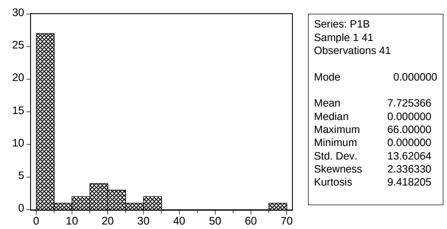

Fig. 13 Results from P1–B ($k = 5$)

first, team conformity and self-censorship in the group and, second, a tendency to polarize individual decisions. Moreover, Weber et al. [2000] maintain that the leadership effect depends on group size. Our results from A1–A and A1–B (Labor Affairs) seem to confirm these ideas.

The previous findings can be summarized into two different results:

Result 1 *Communication induces coordination in the responses.*

But, as not all groups have improved the usual results, we can add:

Result 2 *Communication* per se *does not guarantee an improvement in the average reasoning level.*

Nevertheless, the results in most of our groups are clearly better than the usual ones. However, when comparing the results from different sessions we observe that Law and Management students, in this order, are the most clever. Yet Business students (the only groups with game theory training) did not achieve good results. Why? Could the competition level be a possible explanation?

In order to answer the above question, in the rest of the paper we will focus on the competitive and collusive behavior shown in those groups which coordinated in the theoretical Nash solution, that is, whose mode was zero. Note that the difference between the mode and the mean precisely indicates the percentage of population that is left out of the winning coalition.

4.2 Collusive and competitive behavior

As we noted in the previous section, when we performed the experiment students were given a questionnaire to fill out. According to their answers, it is possible to classify the subjects into four categories:

Table 2 Distribution of right answers across groups

	I	%	L	%	F	%	O	%	Total	CI	NI
B1–A	4	6.7%	6	10.0%	9	15.0%	41	68.3%	60	25.0%	16.7%
B1–B	4	9.1%	3	6.8%	11	25.0%	26	59.1%	44	31.8%	15.9%
B1–C	2	6.3%	6	18.8%	11	34.4%	13	40.6%	32	53.1%	25.0%
M1–A	0	0.0%	2	11.1%	8	44.4%	8	44.4%	18	55.6%	11.1%
M1–B	3	6.0%	4	8.0%	32	64.0%	11	22.0%	50	72.0%	14.0%
M1–C	3	14.3%	4	19.0%	13	61.9%	1	4.8%	21	81.0%	33.3%
P1–A	1	3.1%	1	3.1%	19	59.4%	11	34.4%	32	62.5%	6.3%
P1–B	4	9.8%	4	9.8%	17	41.5%	16	39.0%	41	51.2%	19.5%
L1	0	0.0%	1	9.1%	10	90.9%	0	0.0%	11	100.0%	9.1%
A1–A	2	4.5%	0	0.0%	0	0.0%	42	95.5%	44	0.0%	4.5%
A1–B	0	0.0%	0	0.0%	0	0.0%	27	100.0%	27	0.0%	0.0%
Total	23	6.1%	31	8.2%	130	34.2%	196	51.6%	380		

being I: isolated, L: leaders, F: followers, O: other, CI: collusion index, NI: Nash index.

Leader (L) is the individual who guesses the right number (the perfect equilibrium) and shares it with other people (in our questionnaire these individuals answered Yes to # 3 and # 5. They also gave a "correct" explanation of the IBRB process in # 4).

Isolated (I) is the subject who does not share this information although he has reached it (that is, # 4 was answered correctly and these individuals said Yes in # 3, but No in # 5)

Follower (F) is the person who received the right answer from any leader (these individuals acknowledged this fact in # 3).

Other (O) are the remaining subjects.

This information allows us to define two indexes:

$$\text{Collusion Index } \equiv CI \equiv \frac{Leaders + Followers}{Population}, \tag{1}$$

and,

$$\text{Nash Index } \equiv NI \equiv \frac{Leaders + Isolated}{Population}, \tag{2}$$

The first index (CI) tries to capture the degree of cooperativeness within the population. The second index (NI) includes all the people who were able to solve the game[11]. Table 2 shows the distribution of subjects according to our classification[12].

The first surprising finding is the L1 group (Law, see row 9): only one student knew the theoretical solution, but she communicated it to everybody (she was the leader). The rest of people simply followed her. This implies purely collu-

[11] Note that this is an imperfect indicator as it is possible that a rational person who knows the theoretical solution also recognizes that the other members follow a (wrong) leader and, so, it is better to follow her too than to try on their own.

[12] Groups A1–A and A1–B are included solely for purposes of comparison.

sive behavior, that is, a matching game with an average reasoning level of $k = \infty$ ($mode = mean = 0$)[13].

The other 8 groups (the first 8 rows in the table) correspond to those with a zero mode and a mean greater than zero. A common feature of the 8 groups is that there is a high percentage of students who did not guess the right answer (column 4: others): a large percentage of subjects did not win anything -except in group M1–C in which only one person did not reach the solution. This reflects a lower degree of cooperation among students.

Column 6 illustrates collusive behavior: M1–B, M1–C and P1–A groups show the highest values (72%, 81%, 62.5%, respectively). Moreover, a comparison between the distribution of isolated individuals (first column) and leaders (second column) by groups does not show systematic differences. This means that among intelligent students we find subjects endowed with self and other-regarding preferences.

Nevertheless, a statistical analysis must be added to this preliminary descriptive study in order to draw more accurate conclusions.

4.3 Meta-analysis

To analyze this data we use the meta-analysis technique. As this statistical method is not widely known, in the appendix we have included a section dedicated to the specific instruments used here. Remember that we will analyze only three schools: Business, Management and Public Administration.

Table 3 shows the results from the homogeneity analysis regarding both indexes (CI and NI). The first column represents the indexes as defined above; the second column corresponds to the Q-statistic used to test general homogeneity among groups; the third column demonstrates the probability limits of acceptance of the null hypothesis and, finally, the last one shows if we accept the homogeneity of the groups at a significance level of $\alpha = 0.05$.

Table 3 Homogeneity tests for collusion and game theoretician indexes

Outcome measure	Q statistic	p-value	Homogeneity
CI	54.2302	$2.1177 \cdot 10^{-9}$	No
NI	10.0572	0.1853	Yes

The first important result is that the NI index is homogeneous across the different groups. The Q-statistic value is 10.05 and the p-value is 0.185. Therefore, with $\alpha = .05$, we ought to accept the homogeneous distribution of this index. Thus, we conclude that the game theoretician index is the same for all the experimental sessions. Therefore, the differences observed above among groups may not be

[13] It should be pointed out that this was the only group for which the subject was not compulsory.

explained by an unequal distribution of intelligent students. An estimation of the average NI is $\bar{y}_w = 0.185356$, that is, approximately 18.5% of individuals are able to reach the Nash equilibrium.

Result 3 *The ability to solve complex games is equally distributed among schools.*

Nevertheless, we find that there are significant differences among groups regarding the collusion level (CI). In particular, the value of the Q-statistic is 54.23 and the p-value $2.1177 \cdot 10^{-9}$. So, we can conclude that the CI does not show a homogeneous distribution among the eight groups studied. Thus, the level of collusion within each group is not the same.

We will now analyze this heterogeneity in depth, taking into account features such as (see table 1): answering time, degree, population size[14] and school.

Table 4 demonstrates the results of the homogeneity analysis of CI under these labels. The first column shows the categories of each attribute, while the second shows the Q-statistic value of the homogeneity test. The third column illustrates the probability limit to accept that all groups in each category are homogeneous and, finally, we show the suitable model for each case, that is, either the fixed effects model when homogeneity is accepted or the random effects model if homogeneity is rejected at a significance level of $\alpha = 0.05$.

Table 4 Testing homogeneity (CI) among groups by attributes.

Identification of subpopulations		Q-Statistics	p-value	Modelling
Time to answer	1 week	34.9857	$1.2267 \cdot 10^{-7}$	R. E. $\begin{cases} \bar{y}_w^* = 0.5557 \\ \hat{\sigma}^2 = 0.0714 \end{cases}$
	more	18.1018	$4.188 \cdot 10^{-4}$	R. E. $\begin{cases} \bar{y}_w^* = 0.5217 \\ \hat{\sigma}^2 = 0.0276 \end{cases}$
Degree	3 years	8.4825	0.0754	F. E. $\bar{y}_w = 0.6581$
	5 years	7.2738	0.0263	R. E. $\begin{cases} \bar{y}_w^* = 0.3544 \\ \hat{\sigma}^2 = 0.013 \end{cases}$
Population size	> 35	34.8425	$1.31 \cdot 10^{-7}$	R. E. $\begin{cases} \bar{y}_w^* = 0.4491 \\ \hat{\sigma}^2 = 0.0463 \end{cases}$
	≤ 35	5.9825	0.1124	F. E. $\bar{y}_w = 0.6419$
School	BS	7.2738	0.0263	R. E. $\begin{cases} \bar{y}_w^* = 0.3544 \\ \hat{\sigma}^2 = 0.013 \end{cases}$
	M	1.7294	0.4211	F. E. $\bar{y}_w = 0.6715$
	P	6.5794	0.0103	R. E. $\begin{cases} \bar{y}_w^* = 0.6587 \\ \hat{\sigma}^2 = 0.0374 \end{cases}$

First of all, we have divided the total observations into two groups according to the time span between the instructions and the decision: one week and more than one. In contrast to what we expected, the difference between these two subpopulations is not significant. The time to answer does not have a clear influence on

[14] We have labelled each group as $>$ 35 or $\leq$ 35 people sample.

the collusion level of the participants. Although there exists heterogeneity in the collusion index within each group, the averages are very similar between them. Therefore, we can say that the answering time does not affect the subjects' willingness to share their information with others. This would seem to suggest that in the first week the leaders had already decided how many followers they would share their right guess with and, after that, having more time at one's disposal to answer did not influence their initial behavior.

With regard to the degree programs, three-year-degree students (P1–A, P1–B, M1–A, M1–B and M1–C) are homogeneous in their low level of collusion (0.65), while five-year students (B1–A, B1–B, and B1–C) do not display any similarity. Thus, as the average collusion levels indicate (0.6581 vs. 0.3544), the former are less competitive than the latter.

If we discriminate by school, we only find homogeneity among the 3 classes from the Management School, with a mean of 0.67. Public Administration and Business schools are heterogeneous. Moreover, it is surprising that the average collusion index in Business is notably lower (0.35, nearly half) than in the others. This confirms our prior belief that the Business students are more competitive than the other Social Sciences students in our experiment.

Result 4 *Five-year-degree students are a bit more competitive than three-year-degree students. Nonetheless, the latter are equally cooperative.*

Finally, in relation to group size, the smaller groups (≤ 35) show a higher average collusion index (0.64) and homogeneity between them. However, the bigger groups (> 35) have heterogeneous behavior, with an average lower index (0.44). Thus, in line with Weber et al. [2000], larger size means smaller cooperation. These results can be summarized in the following outcomes:

Result 5 *Classroom size increases competitive preferences. Larger groups imply less collusion.*

Assuming costless communication, result 5 is consistent with our hypothesis of subjects endowed with competitive other-regarding preferences. To ensure victory in a large group it is necessary to coordinate with more people. However, our data shows that the CI is relatively small in these sessions. Therefore, this result can only be explained under the assumption that subjects have behaved competitively[15].

Finally, we should add that there are competitive individuals who sometimes prevent the correct information from reaching everyone. This behavior detracts from the general reasoning level.

5 Conclusions

In this paper we have run a modified version of the BCG standard experimental design in order to study cooperative behavior in a classroom experiment. The set

[15] Continuing with the metaphor in the introduction, if the brothers are not envious, the best thing is to coordinate in one point. If someone is left outside the pact, it is because his welfare is a negative argument in the other brothers' utility function.

of variations introduced in our experiment includes: i) the use of the mode instead of the mean in order to reduce the effect of each individual's guessed number, ii) non-rival payoffs to increase individual incentives to collude, and iii) allowing communication among participants.

This modified BCG encourages students to communicate their guessed numbers to each other. In order to increase the probabilities of winning, each subject will try to form the largest coalition possible, especially if subjects know the Nash equilibrium because they may avoid possible betrayals. Hence, if any subject knows the right answer and does not share it with other individuals —even if this means reducing her own probability of winning— we could infer that she negatively values other subject's payoffs in her own utility function.

Experiments were conducted in 11 classes ranging from 11 to 60 students in size. Participants were given at least one week to submit their guesses and a questionnaire explaining their choice. Our results may be summarized as follows:

1. In contrast to the standard BCG results (guessing widespread along the interval), our design with communication gives rise to a set of guessed numbers concentrated at any point: the 0 point or any weak equilibrium.
2. In this sense, we can say that communication does not guarantee any improvement in the average reasoning level. We need at least one subject who knows the true answer and is willing to share it with other subjects.
3. The percentage of people who are willing to share information decrease with classroom size (the larger the size, the greater the competitive behavior) and type of school (Business students are more competitive than others).

Thus, individual incentives to communicate —to share information— are crucial in order to achieve population reasoning level.

Future research lines include a laboratory experiment where communication can be controlled as well as a comparison between rival and non-rival incentives in our experimental framework.

A Appendix

A.1 Questionnaire

Translation of the questionnaire included on the answers sheet:

1 Name:
2 Number selected:
3 Did you make your choice on your own? Yes/No
4 If Yes, could you explain your process of reasoning?
5 Did you explain your process to anybody? How many people?

A.2 A summary of the meta-analysis method

Meta-analysis is a quantitative method used to combine and integrate the results of several studies that share a common aspect so that they can be treated in a statistical manner. If by k we denote the number of experimental sessions, and p_i, $i = 1,\dots,k$ is the measure outcome, that is, the proportion of people that verifies a specific property or adopts the behavior of interest (defined below), we can estimate each proportion by the corresponding laboratory results.

The point estimates of the measure outcome may differ between them. This variability can be modelled under two types of assumptions. The first assumes that differences are due to sampling error and thus the estimated proportions are considered to be homogeneous. This situation can be modelled by a *fixed effect model* (see Glass et al. 1981 or Sutton et al. 2000).

The second model considers that the variability of the estimations exceeds the expected sampling error, so there must be “real” and “systematic” differences between experimental sessions. This situation is modeled by a *random effect model*.

After deciding which approach is the most appropriate, we can draw some general conclusions about the behavior of proportions in all the sessions.

Briefly, the fixed effect model assumes that all the experimental sessions produce the same proportion. This hypothesis can be expressed as follows:

$$H_0 : p_1 = p_2 = \dots = p_k, \tag{3}$$

In this case the estimation of the population proportion is the weighted average given by,

$$\bar{y}_w = \frac{\sum\limits_{i=1}^{k} w_i p_i}{\sum\limits_{i=1}^{k} w_i},$$

where the weights are,

$$w_i = \left(\frac{p_i\,(1-p_i)}{n_i}\right)^{-1}, \tag{4}$$

and n_i the number of subjects in the ith experimental session.

These inverse-variance weights minimize the variance of the summary estimate defined as,

$$\mathrm{Var}(\bar{y}_w) = \frac{1}{\sum\limits_{i=1}^{k} w_i}.$$

Hence, to test the homogeneity of the results we consider the statistic:

$$Q = \sum_{i=1}^{k} w_i\,(RD_i - \bar{y}_w)^2\,. \tag{5}$$

Our test of this statistic supposes that under H_0, Q is approximately distributed as a χ^2 distribution with $k-1$ degrees of freedom. Hence, we reject H_0 if the value of Q exceeds the $1-\alpha$ percentile of the corresponding χ^2 distribution.

If we find heterogeneity (reject H_0), the proper formulation would be the random effect model. This assumes that proportions are randomly distributed, and typically follows the univariate normal distribution with unknown mean μ and variance σ^2.

The estimation of the population proportion is another weighted average of the proportions,

$$\bar{y}_w^* = \frac{\sum_{i=1}^{k} w_i^* p_i}{\sum_{i=1}^{k} w_i^*} \tag{6}$$

where the weights w_i^* are based on the corresponding fixed effects weights and the Q statistic,

$$w_i^* = \frac{1}{\left(D + \frac{1}{w_i}\right)}, \tag{7}$$

where

$$D = \max\left\{0, \frac{(Q-(k-1))\sum_{i=1}^{k} w_i}{\left(\sum_{i=1}^{k} w_i\right)^2 - \sum_{i=1}^{k} w_i^2}\right\}. \tag{8}$$

The estimation of the variance $\hat{\sigma}^2$ is calculated as

$$\hat{\sigma}^2 \begin{cases} 0 & \text{if} \quad Q \leq k-1, \\ \frac{(Q-(k-1))\sum_{i=1}^{k} w_i}{\left(\sum_{i=1}^{k} w_i\right)^2 - \sum_{i=1}^{k} w_i^2} & \text{otherwise.} \end{cases} \tag{9}$$

Note that these weights combine the variance inter and intra studies. The variance of the weighted average is as follows

$$\text{Var}\left(\bar{y}_w^*\right) = \frac{1}{\sum_{i=1}^{k} w_i^*}$$

Hence, to test if the random effect model is appropriate to the data, we state the null hypothesis $H_0' : \sigma^2 = 0$. Furthermore, we can see that it is equivalent to test H_0 in (3).

References

Virtudes Alba-Fernández, Pablo Bra nas Garza, Francisca Jiménez-Jiménez, and Javier Rodero-Cosano. Teaching nash equilibrium and strategy dominance: A classroom experiment on the beauty contest. Documento de Trabajo E2003/47, CentrA, 2003.

Antoni Bosch-Domènech, José García Moltalvo, Rosemarie Nagel, and Albert Satorra. One, two, (three), infinity, ...: Newspaper and lab beauty-contest experiments. *American Economic Review*, 92(5):1687–1701, 2002.

Antoni Bosch-Domènech and Rosemarie Nagel. Guess the number: Comparing the FT's and Expansión results. *Financial Times*, page 14, 30th June 1997.

Colin F. Camerer. Coordination. Mimeo, Jan. 2001.

John Duffy and Rosemarie Nagel. On the robustness of behavior in experimental beauty contest games. *Economic Journal*, 107:1684–1700, 1997.

Ernst Fehr and Klaus M. Schmidt. A theory of fairness, competition and cooperation. *Quarterly Journal of Economics*, 114:817–868, 1999.

Gene V. Glass, Barry McGraw, and Mary L. Smith. *Meta-analysis in social Research*. Newbury Park, CA: Sage, 1981.

Teck-Hua Ho, Colin F. Camerer, and Keith Weigelt. Iterated dominance and iterated best response in experimental ”p-beauty contests”. *American Economic Review*, 88(4):947–969, September 1998.

John M. Keynes. *The General Theory of Interest, Employment and Money*. MacMillan, London, 1936.

Martin G. Kocher and Matthias Sutter. The ‘decision maker’ matters: Individual versus group behavior in experimental ‘beauty-contest’ games. WP, University of Innsbruck, 2001.

Hervé Moulin. *Game Theory for Social Sciences*. New York Press, N. Y., 1986.

Rosemarie Nagel. *Reasoning and Learning in Guessing Games and Ultimatum Games with Incomplete Information: An Experimental Study*. Unpublished dissertation, University of Bonn, Bonn, 1994.

Rosemarie Nagel. Unraveling in guessing games: An experimental study. *American Economic Review*, 85(5):1313–1326, 1995.

Rosemarie Nagel. A survey on experimental “beauty-contest games”: Bounded rationality and learning. In D. Budescu, I. Erev, and R. Zwick, editors, *Games and Human Behavior, Essays in Honor of Amnon Rapoport*, pages 105–142. Lawrence Erlbaum Associates, New Jersey, 1998.

Amnon Rapoport and Wilfred Amaldoss. Mixed strategies and iterative elimination of strongly dominated strategies: an experimental investigation of states of knowledge. *Journal of Economic Behavior & Organization*, 42:483–521, 2000.

Dale O. Stahl. Rule learning in a guessing game. *Games and Economic Behavior*, 16:303–330, 1996.

Alexander J. Sutton, Keith R. Abrams, David R. Jones, Trevor A. Sheldon, and Fujian Song. *Methods for Meta-analysis in Medical Research*. John Willey and Sons, 2000.

Roberto A. Weber. ‘Learning’ with no feedback in a competitive guessing game. *Games and Economic Behavior*, 44:134–144, 2003.

Roberto A. Weber, Colin F. Camerer, Yuval Rottenstreich, and Marc Knez. The illusion of leadership: misattribution of cause in coordination games. Technical report, Carnegie Mellon, 2000.