Games and Economic Behavior 84 (2014) 39–57

Contents lists available at ScienceDirect

Games and Economic Behavior

www.elsevier.com/locate/geb

Out of your mind: Eliciting individual reasoning in one shot games

Konrad B. Burchardi[a,*], Stefan P. Penczynski[b]

[a] Institute for International Economics Studies, Stockholm University, 106 91 Stockholm, Sweden
[b] Department of Economics, University of Mannheim, L7, 3-5, 68163 Mannheim, Germany

ARTICLE INFO

Article history:
Received 15 December 2011
Available online 8 January 2014

JEL classification:
C91
D83

Keywords:
Individual reasoning
One shot games
Level-*k* model

ABSTRACT

We experimentally investigate the fundamental element of the level-*k* model of reasoning, the level-0 actions and beliefs. We use data from a novel experimental design that allows us to obtain incentivised written accounts of individuals' reasoning. In particular, these accounts allow to infer level-0 beliefs. Level-0 beliefs are not significantly different from 50, and almost 60% of higher level players start their reasoning from a level-0 belief of exactly 50. We also estimate that around one third of the participants play non-strategically. The non-strategic level-0 actions are not uniformly distributed.

1. Introduction

Equilibrium concepts have proven to be a powerful tool in economics, political science, international relations and other fields when trying to understand and predict strategic behaviour of individuals, firms, countries, and other entities. However, manifold experimental studies of human behaviour have shown that the equilibrium concept does poorly in predicting outcomes of one shot games, even when a unique equilibrium exists (see e.g. Nagel, 1995; Stahl and Wilson, 1995; see Camerer, 2003 for an overview). This class of games reflects *strategic situations without precedent* which are faced, for example, by consumers who – otherwise price-takers – buy a house and bargain with the seller about its price.

A leading explanation for the failure of equilibrium concepts to explain and predict behaviour in one shot games is that players may not believe that other players choose an equilibrium strategy. This explanation appears particularly pertinent in unprecedented strategic situations, where learning cannot cause a convergence of beliefs and strategies to equilibrium.[1] The level-*k* model of reasoning, as first proposed by Nagel (1995) and Stahl and Wilson (1995), postulates the following alternative belief structure: There exist so-called level-0 players, who do not play strategically. Strategic players form a belief about the distribution of actions of level-0 players. We refer to this as the player's 'level-0 belief'. The model defines level-1 players to best respond to their level-0 belief. Level-2 players form a belief about the fractions and strategies of lower level players and best respond to this. This process continues for higher level players. Hence the model assumes a

* Corresponding author.
E-mail addresses: konrad.burchardi@iies.su.se (K.B. Burchardi), stefan.penczynski@uni-mannheim.de (S.P. Penczynski).

[1] A distinct explanation could be that players fail to choose the best response to their belief about the other players' strategies. This is the idea behind the Quantal Response Equilibrium (QRE) proposed by McKelvey and Palfrey (1995) in which actions of higher expected payoff are more likely.

0899-8256/$ – see front matter
http://dx.doi.org/10.1016/j.geb.2013.12.005

hierarchy of types who best respond to non-equilibrium beliefs, referred to as *levels*. They distinguish themselves by the number of iterated best responses to the distribution of level-0 actions.[2]

Although there is extensive empirical work in support of the level-k model, little is known about its anchoring elements: What do strategic players believe about the actions of non-strategic players? Do non-strategic players exist? Do strategic players correctly anticipate the actions of non-strategic players? Answers to these questions are required for the model to have predictive power.

We formalise a level-k model which allows to discuss these questions in Section 2. To that specific end, our version of the level-k model allows for (i) a non-uniform level-0 action distribution, (ii) heterogeneous level-0 beliefs and (iii) the distribution of level-0 beliefs to be independent of the distribution of level-0 actions. None of these generalisations is a novel contribution of this paper. Ho et al. (1998), for example, were the first to parameterise a non-uniform level-0 distribution.

We then present two methods to investigate the above questions empirically. Both rely on our specific experimental design. Our laboratory experiment matches players in teams of two, but team partners remain anonymous. The teams participate in a one-shot game. In order to determine a joint team decision, both players decide individually on an action. One of these two actions is chosen (with probability one half) as joint team action. Before this stage, each player is allowed to send exactly one written message to their unknown team partner, and these messages are exchanged simultaneously. We analyse these message to infer the players' individual reasoning.

Three aspects of this design are central to our strategy. First, the written message is a participant's only opportunity to influence the decision of his team partner, who with probability one half decides on the joint team action. The design therefore provides incentives to the participants to state their reasoning fully and clearly. Second, the statements are written at the time of the actual process of reasoning. It is a standard theme in protocol analysis, the research field in psychology concerned with methods of eliciting verbal accounts from participants, that "[t]he closest connection between thinking and verbal reports should be found when participants were instructed to focus on the task while verbalising their ongoing thoughts" (Ericsson, 2002, p. 983). Third, the messages are written prior to receiving any information from the team partner. The written messages therefore reflect individual reasoning. This experimental design can be applied to any one shot game. It builds on the innovative study of strategic play by Cooper and Kagel (2005), who were first to use team communication as a means of inferring reasoning.[3] We believe the experimental design is the most important contribution of this paper.

Our approach relies on the assumption that the team players in our setting think as hard about the game as individual players in individual setups. Equally importantly, our approach relies on the assumption that the team players in our setup have an incentive to articulate their reasoning fully, clearly and truthfully. A failure of these assumptions might, in particular, lead to an upward bias in the estimate of the fraction of level-0 players. Our results need to be interpreted in light of these caveats, which we discuss in Section 3.2.

Using this design, we conduct a laboratory experiment in which participants play the standard 'beauty contest' game: each participant, in our case teams, is asked to state a number between 0 and 100 and the team whose number is closest to $p = 2/3$ of the average of all numbers wins a prize. The unique Nash equilibrium of this game is to play 0 for all teams. We discuss the experimental procedures in detail in Section 3.

The first empirical strategy is then to obtain an estimate of the distribution of level-0 beliefs and level-0 actions by letting research assistants analyse and classify the written accounts of reasoning. Higher level players frequently communicate a level-0 belief. While the majority of them start their reasoning at exactly 50, the level-0 belief exhibits heterogeneity. However, the mean of level-0 beliefs is not significantly different from 50. Further, we detect non-strategic reasoners and obtain an estimate of the distribution of level-0 actions. We find that at least 20% of the participants play non-strategically, in the sense of not attempting to best respond to any belief. The mode and median of the actions of these players are close to 60, but again the mean is not significantly different from 50. These results are discussed in detail in Section 4.

The second empirical strategy is to estimate the structural parameters of our generalised level-k model with a maximum likelihood estimation, using information about the players' sophistication from the written accounts. These estimates suggest that the distributions of level-0 actions and beliefs have an estimated mean of 58 and 54, respectively. These are close to the estimates obtained with the first empirical strategy, and again they are not significantly different from 50. We estimate about one third of participants to be playing non-strategically. The procedure and results are discussed in Section 5.

Related literature The paper relates to an extensive empirical literature on the level-k model of reasoning. Most of this literature is concerned with estimating the distribution of level-k types in a given population. A central challenge faced by empirical studies of the level-k model is that it can typically explain a given *action* in a single round of a game with several

[2] Applications of the level-k model include dominance solvable games (Costa-Gomes and Crawford, 2006; Nagel, 1995), normal-form games (Costa-Gomes et al., 2001; Stahl and Wilson, 1995), two-person zero-sum games with non-neutral framing (Crawford and Iriberri, 2007a), auctions (Crawford and Iriberri, 2007b), and coordination games (Crawford et al., 2008). For a more complete overview see Camerer et al. (2004) and Crawford et al. (2013).

[3] They let players communicate via a free instant messenger in order to observe the players' communication and learn about the strategic nature of the arguments. This method enabled them to investigate the dynamics of strategic sophistication of teams as a whole. Our aim, instead, is to infer individual level reasoning. Therefore we use a communication protocol in which a single message is written by subjects *prior* to any team interaction, which allows observing individual reasoning in one shot games. The use of a free communication protocol is not suitable for our purpose, since it would be difficult to disentangle individual reasoning from the influences of the statements of a player's team partner.

kinds of reasoning.[4] Several approaches exist to obtain additional information in order to infer the underlying reasoning. One powerful approach is to obtain for each player multiple choices made in subsequently played variants of a game without feedback, as in Stahl and Wilson (1995), Costa-Gomes et al. (2001) and Costa-Gomes and Crawford (2006). In these studies, a player's sequence of choices is matched to a typical 'fingerprint' of, say, a level-1 player. This provides estimates of individual levels of reasoning. In addition to the 'fingerprint', the latter two studies use information search data to identify types of players.[5] Another approach to collect more information is reported by Agranov et al. (2013) who incentivise and observe provisional choices over time and relate the initial considerations to the level-0 belief. This approach has in common with our approach that it adds information on the thought process when playing a single round of the game. In her original study, Nagel (1994) asked the participants to verbally state the reason for their chosen action and then classified and analysed comments of participants. A descriptive analysis of optionally given verbal comments received in the context of a newspaper experiment is presented by Bosch-Domènech et al. (2002).

Another approach is to impose more structure on the model to be estimated. Camerer et al. (2004) use a GMM estimator to pin down the single parameter of the cognitive hierarchy model, a one-parameter level of reasoning model. This is an elegant methodology to estimate the level-k distribution that assumes certain level-0 action and belief distributions. Our second empirical strategy is to estimate the parameters of a more general structural model using maximum likelihood estimation.[6] Conceptually, our econometrical model is closest to the one used by Ho et al. (1998). They model normally distributed level-0 actions and iteratively derive higher level beliefs as the mean of a number of draws from the distribution of actions one level below. They obtain estimates of the normal level-0 distribution and the fractions of level-k players.

Yet another approach is proposed in Bosch-Domènech et al. (2010), who use a large dataset to fit a set of beta distributions to action data from various 'beauty contest' games. Some of the super-imposed beta distributions are then associated to underlying levels of reasoning, which provides both estimates of the type distribution and the choices for individual types.

2. Generalised level-*k* model

This section presents a generalised level-k model. The purpose of these generalisations is to provide a framework to think about and distinguish some of the specific assumptions which have previously been made. Having a generalised model which nests the different specific versions of the level-k model is as well necessary to empirically evaluate these assumptions. It is not our purpose to introduce a 'new' level-k model since simpler models might be more suited tools in most applications of strategic thinking.

Consider a game in which N players, indexed by i, simultaneously choose an action from the set of actions X_i. Given a vector of actions $(x_1, \ldots, x_N)$, player i's payoff is given by the utility function $u_i(x_1, \ldots, x_N)$. We restrict attention to symmetric games, like those which Nagel (1995) studied, in which $X_i = X_j$ and $u_i = u_j$, for each i, j, and each u_i is invariant to permutations in opponents' strategies. We therefore drop player index subscripts from X_i and u_i.

Level-k type models deviate from equilibrium models in that they allow for heterogeneity in the belief about other players' actions across players. The belief formation is characterised as an iterative reasoning process, and players exhibit heterogeneous levels of reasoning k, $k = 0, 1, 2, 3, \ldots$. Denote the fraction of level-k types in the population with l_k and define $\boldsymbol{l} := \{l_0, l_1, l_2, \ldots\}'$.

Level-0 players are defined as playing non-strategically in the sense that their action does not depend on any belief over other players' actions. We denote the true distribution of level-0 actions across level-0 players by $f_0(x; \boldsymbol{\eta}^0) \in \Delta(X)$, where $\Delta(X)$ is the set of distributions over X and f_0 is fully characterised by the parameter vector $\boldsymbol{\eta}^0$. We refer to $f_0(x; \boldsymbol{\eta}^0)$ as 'level-0 action distribution', or just 'level-0 distribution'.

Players who exhibit a higher level of reasoning ($k \geqslant 1$) are playing strategically in the sense that they choose a best response to their belief about the actions of other players. In some symmetric games players only need to form a belief about a small set of statistics of the distribution of other players' actions, in the sense that conditional on knowledge of these statistics, information about the full distribution leaves the best response unchanged. In the 'beauty contest' game analysed in this paper, and for three players, Breitmoser (2012) shows that the best response to other players whose actions are independently drawn from identical uniform distributions with support $[0, 100]$ can be quite different from best responses to the mean of that uniform distribution. In our setting, the best response calculated as in Breitmoser (2012) for a game with 6 and 8 players are 33.0 and 33.2, respectively. These are very close to just choosing $2/3$ of the mean of the distribution of other players' actions. Therefore, we abstract from strategic considerations as in Breitmoser (2012), and

[4] For example, a player who chooses 33 in the standard 'beauty-contest' game might do so because he is a level-1 reasoner who believes that level-0 reasoners play on average 50. But he might just was well be a level-0 reasoner who has chosen the number at random or a level-2 reasoner who beliefs that the population is composed of a combination of level-1 and level-0 reasoners who on average choose 50.

[5] This method was introduced by Camerer et al. (1993).

[6] Note that the level-k model typically makes stark choice predictions when the distribution of beliefs is assumed to be non-heterogeneous and no noise is assumed. For example, a uniform level-0 distribution in the 'beauty contest' game, together with non-heterogeneous, consistent beliefs about the mean of level-0 actions, imply a degenerate distribution of level-0 beliefs at 50. Any action slightly off 33, 22, etc. will be attributed to level-0 play if no additional error structure is assumed. A maximum likelihood estimation then typically yields almost all players to be level-0 players and fits the level-0 distribution to the full-sample action distribution. We explicitly model heterogeneity in the level-0 belief, which causes heterogeneity in actions of players of the same level-k, $k > 0$, and allows us to separate this from level-0 play. In this sense the generalisation of the model is important for our estimation strategy.

work with the simplifying assumption that a player only forms a belief about the mean of the distribution of other players' actions and best responds to that belief.[7] When player i believes that the mean of actions of all of her opponents $j \neq i$ is g we write her expected utility as $u(x, g)$.

Denote with b_i the belief of player i of the mean of the distribution of actions of level-0 players, where b_i may vary across players. We refer to b_i as player i's 'level-0 belief' and denote with $h(\cdot; \boldsymbol{\xi})$ the 'distribution of level-0 beliefs' in the population, characterised by the parameter vector $\boldsymbol{\xi}$. Further, players of level k believe all other players in the population to be of level $k-1$. Consider a level-1 player with a level-0 belief b_i. He believes that he only faces level-0 players. He will choose his action as $x_1(b_i) = \arg\max_x u(x, b_i)$, assuming a unique best response. The best response of player i of level $k > 1$ with level-0 belief b_i is:

$$x_k(b_i) = \arg\max_x u\big(x, x_{k-1}(b_i)\big),$$

where we assume a unique solution. This formulation pins down $x_k(b_i)$ iteratively. Implicit in the formulation is the assumption that a level-k player believes all other players to be of level $k-1$.[8]

Note that while each level-1 player i has a unique best response, heterogeneity in b_i across level-1 players implies that generally there is a non-degenerate distribution of best-responses of level-1 players. Since b_i is heterogeneous, $x_k(\cdot)$ is generally heterogeneous across players of the same level k. Denote the distribution of actions of level-k players as $f_k(x; \boldsymbol{\eta}^k) \in \Delta(X)$, and let it be characterised by parameter vector $\boldsymbol{\eta}^k$. Importantly, the distribution of actions will depend on the distribution of b_i's in the population. Given assumptions on the level-0 distribution $f_0(x; \boldsymbol{\eta}^0)$, the distribution of b_i's in the population, the true level-k distribution $\boldsymbol{l}$, the level-k model makes a probabilistic prediction about the frequencies of actions. We will return to this in Section 5.

Our formulation of the level-k model is general in the sense that it nests many aspects of the level-0 specifications of models by Nagel (1995), Stahl and Wilson (1995), Costa-Gomes et al. (2001), Camerer et al. (2004) and Costa-Gomes and Crawford (2006) as special cases.[9]

3. Inferring individual reasoning

This paper presents two empirical strategies to analyse level-0 actions and beliefs thereof. Both of these strategies rely on our specific experimental design which generates written messages, and the classification of these messages. This section explains our design, its caveats, the experimental procedures and how we classify the written messages.

3.1. Experimental design

In order to elicit the reasoning underlying a player's action, we designed the following game structure: Individuals are randomly assigned in teams of two players. Their payoff in the game depends on a joint 'team action'. To determine this, both players are given the chance to choose an action – which we call the 'final decision'. Then one player's decision is chosen randomly, with probability one half as the team action. Consequently, each player faces a 50% chance of having her partner's final decision determining the team action. The players hence have an incentive to ensure that their team partner's final decision is as sound as possible. Importantly, the players are given the possibility to convince their partner of the optimal team action. In particular, players are allowed to write one message to their team partner, which consists of a 'suggested decision' and a justifying text. This text is unlimited in size and its writing is not limited in time. The messages are exchanged *simultaneously* once both players have entered their message, and thereafter the players take their final decision individually.

The simultaneous exchange of a single message ensures that every explanatory statement and suggested decision in the first round is written without any previous communication with the team partner, hence reflecting an individual's reasoning. It is therefore this first message and suggested decision which we will analyse in this paper in order to understand individual decision making in situations of strategic interaction.[10] Importantly, under the reasonable assumption that the best way to

[7] Without this simplifying assumption it is virtually impossible with conventional computing power to estimate our model. Both the derivation of the best response in the sense of Breitmoser (2012) in games with 6 and 8 players and estimating our empirical model are computationally involved.

[8] This assumption follows Nagel (1995). This is in contrast to Camerer et al. (2004) who introduced the 'cognitive hierarchy' model where the players' beliefs reflect the true relative frequencies of lower level types and the true distribution of types follows a Poisson distribution. Note that for players with $k \geqslant 2$ higher-order beliefs need to be specified: a player needs to form a belief over the lower-level players' beliefs of the level-0 distribution. We assume that beliefs and higher-order beliefs coincide. See Goeree and Holt (2004) for a set of more general assumptions. Similarly, a $k \geqslant 3$ player needs to form a belief about the population beliefs of player i, $2 \leqslant i \leqslant k-1$, i.e. a level-3 player needs to know what a level-2 player believes the relative proportions of level-0 and level-1 players are. Following the literature, our specification assumes these higher-order population beliefs are assumed to be consistent with the beliefs of the lower-level player.

[9] Nagel (1995) assumes level-0 play to be 50, the expected choice when playing randomly according to a symmetric distribution or as a result of salience. Players' level-0 belief is assumed consistent, which renders $h(\cdot)$ degenerate. Stahl and Wilson (1995) and Camerer et al. (2004) assume uniform randomly distributed level-0 actions and consistent, homogeneous level-0 beliefs. Costa-Gomes et al. (2001) and Costa-Gomes and Crawford (2006) work with the same hypotheses and their evidence, including a specification test, does not reject them. Various studies specify further types and different population beliefs, assumptions which are not nested in our setup.

[10] The experiment was programmed and conducted with the software z-Tree (Fischbacher, 2007).

convince one's team partner is by explaining him the own reasoning, this design gives an incentive to write down the reasoning process as fully and clearly as possible. It is generally applicable in one shot games and we believe it constitutes an important methodological contribution of our paper.

Note that in contrast to the suggested decision, the final decision does not reflect individual decision making, since it is taken after receiving the team partner's message. A companion paper (Penczynski, 2012) shows that lower-level players indeed tend to adopt the suggested decisions of higher-level team partners in their final decision. The final decision of the participants has a lower mean and median than the suggested decision. Moreover the standard deviation of decisions across participants decreases after the exchange of the message and outliers become fewer, consistent with the idea that level-0 players who potentially chose the outliers, become fewer. Throughout this paper we will therefore exclusively use the suggested decision for the subsequent analysis and classification of the individual reasoning.

3.2. Caveats of the experimental design

The experimental design has several potential caveats.

First, the team setup might induce other population beliefs than an individual setup does. The suggested and final decisions are taken while knowing that the opponents are teams of two players. If teams are believed to perform better than individuals, this might induce a population belief with more weight on higher levels.

Second, the experimental setup might induce higher or lower effort to think about the best response. The suggested decision is taken while justifying it in a message. This might increase the participants' level of reasoning relative to an individual setup, e.g. because the participants would examine the task at hand more thoroughly in order to state sensible arguments in the communication. On the other side, the team setup might induce the decision maker to think less hard about the game. The decision maker is matched with a team partner who might be better or worse than himself at thinking about the game. Obviously the decision maker does not know, when writing the message, whether his team partner is better or worse. Our experimental setup exploits that with some probability he is matched with a team partner who is worse at thinking about the game. But since with some probability he is matched with a better team partner from whom he might learn, his incentives to think about the game are lower than in an individual setup. Also, we cannot exclude the possibility that the requirement to explain one's decision can cause stress, leading to a lower level of reasoning.

These effects work in opposite directions. The sum of these effects might both increase or decrease the level of reasoning and therefore might lead to lower or higher actions of participants, respectively. Subsequently, we show that the distribution of suggested decisions is similar to results from non-communication treatments in the literature, e.g. Nagel (1995). If anything, the decisions we observe are slightly higher than in Nagel (1995), suggesting that some of the effects which lead to a lower level of reasoning might be present. However, the similarity of our data to Nagel (1995) is suggesting that none of the above effects seems to be particularly strong.

Third, we work under the assumption that the participants think the best way to persuade the team partner is to write down their own reasoning as fully and clearly as possible. A concern with our design is that individuals might send an 'intermediate' reasoning to their team partner, which they judge more persuasive than their real reasoning. For example, they might not articulate some of the first of k steps of reasoning: a level-2 player might choose to write a message which suggests a level-0 belief of 33 and exhibits one step of reasoning, suggesting an action of 22 to his team partner. To the extent this happened in the experiment, our classified level-bounds will underestimate the true level of reasoning. Ultimately, we cannot discard this possibility.[11]

Our results need to be interpreted in light of these caveats. In particular, the high fraction of level-0 players we find might be explained by some of these effects.

3.3. Experimental procedures

We conducted the experiment in the Experimental Economics Laboratory of the Department of Economics in Royal Holloway (University of London). In 6 sessions we played three rounds of the 'beauty-contest' game ($p = 2/3$). Since in this paper we are interested in individual reasoning, we only analyse the first round suggested decision and the accompanying first message, i.e. the activities that took place before any interaction of the team players.

At the start of each session the participants were made familiar with the structure of the experiment and the messaging system in two practice rounds. We used the same software as above for the practice rounds, but asked the teams to find the answer to two unrelated questions. Since we wanted to avoid any pre-treatment sensitisation to strategic considerations, we asked them to provide the year of two historic events. The questions in the test round were chosen to be relatively difficult to stimulate the use of the messaging system. The participants of the experiment were paid a show-up fee of £5 and in each round separately the winning team won a prize of £20 (£10 per team player).

[11] There is some weak evidence suggesting that this phenomenon might not be prevalent: If a higher level player chooses to skip the initial steps of his reasoning in his message, we would expect to observe relatively low stated level-0 beliefs. Instead we find level-0 beliefs which are on average higher than 50, and not significantly different from 50. For somebody with a prior belief that the level-0 is uniform random and beliefs about the mean are around 50, which we do not maintain, this suggests that few subjects skipped some initial steps of their reasoning.

A total of 84 individuals participated in our experiment. Sessions had 12, 14 or 16 participants. The participants were mainly undergraduate students in Royal Holloway and all of them were recruited by the host institution. Out of the 84 students 15 were studying Economics, 13 of them being in their first year of studies, one in the second year, one being in the third year. 16 of the 84 students had received some form of training in game theory, but only 5 had been confronted with the 'beauty contest' game. The majority of students had participated in an economic experiment before.

3.4. Classification of communication transcripts

We use the information elicited on individual reasoning in the following way: First, we identify the cases in which no level-k type of reasoning is undertaken. Then, conditional on a level-k reasoning being applied by the player, which includes non-strategic play, we uncover two sets of data: (a) the maximum and minimum number of steps of reasoning which can be interpreted into the message, including possibly 0, and (b) the "communicated level-0 belief" which a player states, if any.

In particular, two research assistants read the messages and classified the type of reasoning with the following procedure:

1. To investigate the prevalence of reasoning patterns different from level-k reasoning we asked the RAs to indicate whether the player puts forward equilibrium reasoning. For this it was not necessary that the player actually played the unique equilibrium strategy.[12] The RAs were further instructed to denote whether the player applied an iterated elimination of dominated strategies. For this it is necessary that *first* some actions are excluded and then a strategy is formed for the remaining action space.[13]

2a. If any level-k reasoning was explained in the message, we asked the RAs to indicate if the argument contained a belief about the mean of the others' play that served as a starting point for best responses, but was in itself not derived by choosing a best response. If so, we asked them to denote this belief about the mean of the others' play. We refer to is as 'communicated level-0 belief'.[14]

2b. Lastly, we were interested in how many steps of reasoning the player applies. When designing the classification procedure we were worried that in some cases it might not be possible to identify from the communication *exactly* how many steps of reasoning were applied.[15] We therefore asked the classifiers to only indicate the lowest level of reasoning which is clearly stated and the highest level of reasoning which could possibly be interpreted into the messages.[16] We refer to these as 'lower bound' and 'upper bound', respectively. We instructed the classifiers to consider as level-0 a player whose message does not exhibit "any strategic reasoning whatsoever". This might arise as a result of choosing a number randomly or based on non-strategic considerations such as taste. We emphasised that for this classification to be chosen, it was important that the player was not in any way best responding to what he thought others would play.

When designing the classification procedure we intended to avoid two potential concerns: First, the classifiers might try to extract more information than the messages actually contain. We therefore instructed the classifiers to only enter information when it was clearly contained in the message.[17] Second, we were concerned that in the 'beauty-contest' game in the case of an ambiguous statement relatively low suggested decisions might lead the classifiers to indicate a higher lower-bound on the level of reasoning than was clearly exhibited. In contrast when indicating the upper-bound, knowing about low choices should, if anything, lead the classifiers to indicate a higher upper-bound. We therefore split the classification of the messages into two parts. We did not reveal the choice data to the classifiers when asking for the lower-bound but revealed it subsequently when asking for the upper bound.[18]

The classification was undertaken by two PhD students in the Department of Economics at LSE. First they classified the transcripts individually. After this phase their classification of the lower bound coincided for 77% of all participants and the classification of the upper bound coincided in 76% of all cases. Then the two RAs met to reconcile their judgements and provide a joint classification, if possible. We only use data on which they could agree in the reconciliation.

---

[12] The classification in Bosch-Domènech et al. (2002) proceeds similarly by including level-∞ in the classification. We call a player a 'sophisticated' type, if she recognised the equilibrium, but played an action different from the unique equilibrium action.

[13] For example the statement "Everybody plays on average 50 so I should not play higher than 34" is not an iterated elimination of dominated strategies.

[14] For completeness, we also asked whether an argument revealed a communicated population belief distribution. If so, the classifiers were asked to indicate whether it was degenerate or non-degenerate.

[15] Think for example of the imaginary statement: "I presume everybody else will play 33, so let us play 22." This clearly exhibits one step of reasoning. But it seems possible, too, that the player skipped the first step of his reasoning when writing down his argument.

[16] The instructions specified that the classifiers, after writing down the lower bounds should be able to say to themselves: "It seems impossible that the players' level of reasoning is below this number!", and after writing down the upper bounds: "Although maybe not clearly communicated, this statement could be an expression of this level. If the player reasoned higher than this number, this was not expressed in the statement!"

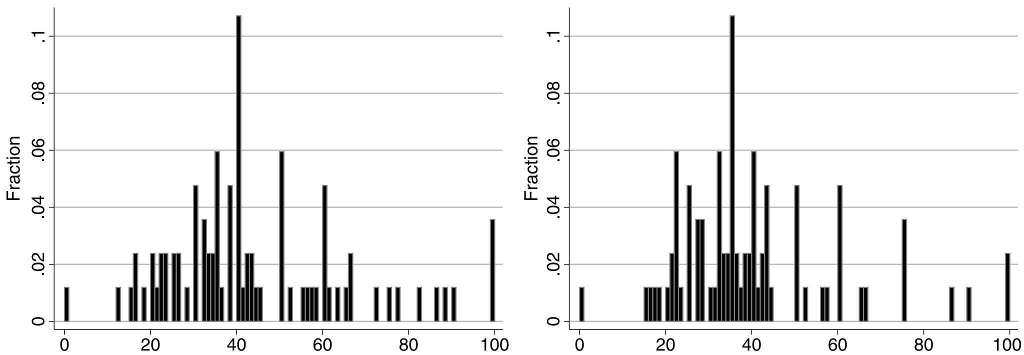
[17] The instructions were self-contained and were not complemented by verbal comments. The instructions were written by the two authors, of whom one had taken a look at the communication transcripts beforehand. They can be obtained from the authors upon request.

[18] Other studies applying a classification use a similar procedure in order to avoid any unconscious alignment of the classification with the choice data that might result from implicit assumptions (for example Rydval et al., 2009).

Table 1
Summary statistics.

	Mean	Std. Dev.	Median	Min.	Max.	*N*
PANEL A: Full sample						
Suggested decision	43.93	21.14	40	0	100	84
Final decision	39.73	18.75	35	0	100	84
Team action[a]	40.02	18.98	35.5	16	100	84
PANEL B: Level-0 players						
Suggested decision	62.35	22.39	60	16	100	17
PANEL C: Non-level-0 players with communicated belief						
Communicated level-0 belief	55.26	12.33	50	40	100	36

[a] The team action is a random draw of the two final decisions.



(A) Suggested decisions (B) Final decisions

Fig. 1. Individual decisions ($N = 84$).

Later we asked further 3 RAs to again classify the level bounds. Table A.5 in Appendix A.4 shows, for example, that for 76 out of 78 messages (∼97%), 3 or more of the 5 classifiers agreed on exactly one level lower bound.[19] We take this as strong evidence that our method of classification is robust, provides informative insights about individual reasoning and can easily be replicated.

4. Experimental results

4.1. Action data

Table 1 presents aggregate summary statistics for all 6 sessions. Fig. 1 shows histograms of the suggested decision and the final decisions aggregated over all sessions.

The suggested decision is comparable to the first period's decision of other individual level experiments, in that it is taken by the participants *individually* without having received any message from the team partner. The data on the suggested decision is similar to data generated in other comparable experiments in having similar means and a high fraction of choices between 20 and 50.[20] The original study by Nagel (1995) had an average of 36.73, which is slightly lower than ours. We believe that this difference is likely attributable to the different subject pool.[21]

In contrast to the suggested decision, the final decision does not reflect individual decision making, since it is taken after receiving the team partner's message. A companion paper (Penczynski, 2012) shows that lower-level players indeed tend to adopt the suggested decisions of higher-level team partners in their final decision. Also, the final decision of the participants has a lower mean and median than the suggested decision. Moreover the standard deviation of decisions across participants decreases after the exchange of the message and outliers become fewer, consistent with the idea that level-0 players who potentially chose the outliers, become fewer. Therefore we will exclusively use the suggested decision for the subsequent analysis and classification of the individual reasoning.

[19] The 6 subjects that did not write a message are dropped from this exposition.

[20] The spike at 40 might be unusual. The communication data reveals that this arises mainly as a result of two factors: some level-0 players chose 40 and some level-1 players chose 40 as they held the belief that the level-0 mean would equal 60.

[21] Table II in Camerer et al. (2004) illustrates that the average number in such a game depends on population characteristics. It seems that our subject sample performs more similar to the Caltech board and students of the Pasadena City College than to the students in Nagel's study.

Table 2
Level classification results.

		Level upper bounds					Total
		0	1	2	3	NA	
Level lower bounds	0	17	11	1	0	6	35
	1		26	3	0	2	31
	2			6	5	0	11
	3				1	0	1
	NA					6	6
	Total	17	37	10	6	14	84

Notes: The cells in this table indicate the number of subjects that were classified with the respective combination of lower and upper bound.

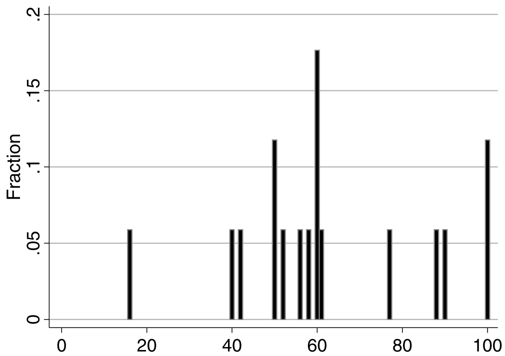


Fig. 2. Suggested decisions of level-0 players ($N = 17$).

4.2. Level-k bounds

Table 2 presents the lower and upper bounds on the level of reasoning of individuals. For 70 participants both a lower and an upper bound were indicated. Eight participants have a non-classified upper bound. Another 6 participants did not make any statement and could therefore not be classified.

For 50 of the 84 participants the lower and the upper bounds coincide ($\approx 60\%$) and hence the classification fully determines their level of reasoning. These are 17 level-0, 26 level-1, 6 level-2 and 1 level-3 players, corresponding to the diagonal of Table 2. For further 20 players the classification restricts the level of reasoning to be one of two possibilities. Only for one participant we have an interval between 0 and 2. None of those participants for whom both a lower and upper bound is indicated was classified as potentially reasoning higher than level 3.[22]

Two players identified the Nash equilibrium, one of them being an 'equilibrium'-type that suggested playing 0, another one being 'sophisticated' in the sense that she imitated a level-2 player and suggested 20. No upper bound was assigned to those players who identified the equilibrium. Further two participants were found to apply elimination of dominated strategies.

To the extent that the written messages reflect the players' true level of reasoning, the data suggests that at least 20% of the subjects are non-strategic reasoners, since 17 out of 84 participants are identified as level-0 reasoners. For additional 24 participants we cannot exclude the possibility that they are level-0 reasoners. When estimating the level-*k* model (see Section 5.3) we estimate 37% of the population to be level-0 reasoners.

4.3. Level-0 actions and beliefs

We analyse the actions chosen by those with a lower and upper bound of 0 in order to provide an estimate of the distribution of level-0 play. Panel B of Table 1 shows summary statistics of the suggested decisions of the 17 players who are identified as level-0 players from the communication, and Fig. 2 shows a histogram of these decisions. A one sample Kolmogorov–Smirnov test finds that the distribution is significantly different from uniform (p-value $= 0.038$). Level-0 players choose on average 62 and their median choice is 60, deviating from the uniform distribution in the direction of higher numbers. The null of the mean of level-0 actions being equal to 50 is not rejected at a 5% significance level (p-value $= 0.076$) when tested against the alternative that it is different from 50.[23]

[22] The data by subject can be obtained from the authors upon request.
[23] The p-value is calculated by simulating (with 100,000 runs) the distribution of the mean of 17 draws from a uniform distribution on the interval $[0, 100]$.

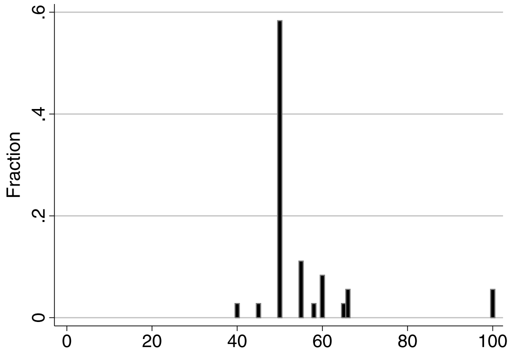


Fig. 3. Communicated level-0 beliefs ($N = 36$).

The analysis of the written accounts of individual reasoning further allows us to analyse the players' communicated beliefs about the average action of level-0 players. In 36 of the messages the players stated a non-derived belief about the average action of other players.[24] Table 1 shows summary statistics of the communicated level-0 beliefs in panel C. The mean of these level-0 beliefs is 55.25. Fig. 3 presents the distribution of the communicated level-0 beliefs.[25]

More than 20 participants started reasoning with a communicated level-0 belief of exactly 50, consistent with the assumption of a degenerate level-0 belief derived from a uniform level-0 distribution. Interestingly, players who choose a different starting point for their reasoning predominantly deviate in a direction consistent with the level-0 actions, causing the mean to be higher than 50, namely 55.26. 11 players have a level-0 belief between 55 and 66. The alignment might result from salience considerations, since $p = 2/3 \approx 0.667$ possibly primes towards higher numbers than 50.

5. Estimation of structural level-*k* model

This section discusses how a generalised level-k model, which allows for heterogeneous level-0 beliefs, can be estimated. We will use the suggested decision together with the upper and lower bounds obtained from the classification to estimate the model's structural parameters, in particular the distribution of level-k types and the distributions of level-0 actions and beliefs. Among others this allows us to estimate the fraction of level-0 players.

5.1. An estimable model

The level-k model outlined in Section 2 makes a probabilistic prediction about the observed actions. Let $f_k(x; \boldsymbol{\eta}^k)$ denote the probability mass function over the actions of a level-k player. Then the unconditional probability mass function of the action of some player i can be written as

$$p(x_i; \boldsymbol{\psi}) = \sum_{k=0}^{K} l_k f_k\big(x_i; \boldsymbol{\eta}^k\big) \tag{1}$$

where $\boldsymbol{\psi} = (\boldsymbol{l}, \boldsymbol{\eta}^0, \boldsymbol{\xi})$, $l_k \geqslant 0$ for all k and $\sum_{k=0}^{K} l_k = 1$. This is a convex combination of component densities denoted in the statistics literature as 'finite mixture distribution'. The l_k's give the weight of each component distribution, and the $\boldsymbol{\eta}^k$'s are the parameters which characterise each component distribution.[26]

The distribution of all actions is hence a finite mixture of the action distributions of level-0, level-1, level-2 players and so on. We will outline in Section 5.2 how to consistently estimate the parameter vector $\boldsymbol{\psi}$. In the following we will first describe which form $f_k(x; \boldsymbol{\eta}^k)$ takes, how it depends on the level-0 action and belief distributions and how we parameterise these.

Action distributions in the level-*k* model The action distribution of a level-0 player is simply $f_0(x; \boldsymbol{\eta}^0)$. The actions of higher level players, $f_k(x; \boldsymbol{\eta}^k)$ with $k \geqslant 1$, are derived from their level-0 belief b_i. As the level-0 belief is a random variable, the action distribution of a level-k player can be understood as the distribution of a transformed random variable.[27]

---

[24] An example of such a statement is: "I think there's got to be at least 8 or 9 teams, so I reckon thats got to come to at least 500 in total, the average of which would be about 55–60, which gives 40 as two thirds of it. So I say we go for between 40 and 50." This was classified as a level-0 belief of 57.5. Of these 36 messages for which a level-0 belief was classified, one was written by a player identified as level-0 player, who while stating a belief about the average of other players' actions did not best respond to it.

[25] Out of 36 players, 7 indicate a narrow interval for their belief about the mean of level-0 actions. The remaining 29 players indicate a single number for their belief about the mean of level-0 actions. This is in line with results that level-0 beliefs are degenerate in Ho et al. (1998).

[26] The parameters $\boldsymbol{\eta}^k$, $k = 1, 2, 3 \ldots$ are a function of $\boldsymbol{\xi}$.

[27] A non-degenerate level-0 belief distribution implies a non-degenerate action distribution for higher level players. In our model, without decision noise, a non-degenerate level-0 belief distribution is required to explain a non-degenerate action distribution of higher level players.

For a general form of population beliefs, this might be a complicated transformation. This is simplified by the fact that the messages predominantly exhibit degenerate communicated population beliefs in the sense that level-k players expect all others to be level-$(k-1)$.[28] We have therefore assumed degenerate population beliefs throughout. The action of a level-k player then follows from his level-0 belief as $p^k \cdot b_i$ or in our case $(2/3)^k \cdot b_i$.

Parametric assumptions We parameterise $f_0(x; \boldsymbol{\eta}^0)$ as bounded normal distribution defined on the interval $[0, 100]$ and characterised by the mean and standard deviation $(\eta^0_\mu, \eta^0_\sigma) = \boldsymbol{\eta}^0$. This allows for a concentration of the level-0 actions as well as an approximate uniform distribution when σ^0 is large. We parameterise the distribution of the level-0 belief $h(\cdot; \boldsymbol{\xi})$ as well as bounded normal defined on the interval $[0, 100]$ and characterised by the mean and standard deviation $(\xi_\mu, \xi_\sigma) = \boldsymbol{\xi}$. As a special case this allows for the mean of the distribution of level-0 beliefs $h(\cdot; \boldsymbol{\xi})$ to coincide with the mean of the level-0 action distribution. To find the action distribution of level-k players that follows from this level-0 belief distribution, Lemma 1 in Appendix A.1 is useful. It states that a random variable which is distributed as bounded normal on $[0, 100]$ with parameters (μ, σ) will – when applying a multiplicative transformation using a factor a – be distributed as bounded normal with parameters $(a\mu, a\sigma)$ and support $[0, a100]$. Therefore, in our parameterised version of the model, the action distribution of a level-k player, $f_k(x; \boldsymbol{\eta}^k)$, $k \geqslant 1$, is a bounded normal distribution with $\boldsymbol{\eta}^k = (2/3)^k \boldsymbol{\xi}$ and support $[0, (2/3)^k 100]$.[29]

We estimate the fraction of level reasoners for levels 0 to 3. No parametric structure is imposed on the distribution of level-reasoners, but for computational reasons we assume the highest level of reasoning to be 3. This is consistent with the classification of the messages which never indicate a level of reasoning higher than 3, and also with the findings in Arad and Rubinstein (2012).

5.2. Estimator and identification

Likelihood function Given the probability mass function in Eq. (1) for the action of a player of unknown level, we can write the log-likelihood of the data as

$$L(\boldsymbol{x}; \boldsymbol{\psi}) = \sum_{i=1}^{n} \log p(x_i; \boldsymbol{\psi}), \tag{2}$$

where $\boldsymbol{x}$ is the vector of actions observed. We use the information on the bounds by imposing $l_k = 0$ in $p(x_i; \boldsymbol{\psi})$ when the classification information is such that individual i is certainly not of level k.

Identification For the model to be identified, the mixture densities need to be linearly independent for all mixture probabilities $l_k \neq 0$. In the 'beauty contest' game with a parameterisation of the level-0 action and belief distribution as bounded normal this will necessarily be satisfied, irrespective of how many levels are estimated.[30] If the true data generating process is reflected by our econometric model, all parameters are therefore identified. However, if the true data generating process is different, this is not necessarily true; for example full identification might fail if the true data generating process comprises decision errors of level-k, $k > 0$, players. Since heterogeneity in the level-0 belief and higher-level decision error have similar implications for the distribution of actions of higher-level players, they are not separately identified without strong assumptions.[31] An empirical specification with both decision errors and heterogeneity in the level-0 belief would rely heavily on functional form assumptions to separately identify both.

Maximum likelihood estimator We estimate the parameter vector $\boldsymbol{\psi}$ with the maximum-likelihood estimator, denoted $\hat{\psi}_{MLE}$. The log-likelihood function is thrice differentiable and the expectation of the third partial derivative is finite. We are unable to calculate the true information matrix, but we calculate an estimate of it and verify that it is positive definite. The MLE for this estimation problem is hence consistent, asymptotically normal and with asymptotic variance given by the inverse of the Fisher information matrix. We find the global maximiser of the log-likelihood function numerically, since the likelihood equations cannot be solved analytically.[32]

[28] In our experiment, out of 12 players with a lower bound on k greater or equal to 2, the messages of 9 players are classified to have a degenerate population belief.

[29] Note that in Eq. (1) the component density f_k is indexed, allowing for the possibility that they are of different parametric families. In the case of the 'beauty contest' we could omit the subscript.

[30] Unless in the special case where $\boldsymbol{\eta}^0 = (2/3)^k \boldsymbol{\xi}$ for some $k \geqslant 1$. However, for any non-zero mean of the level-0 distribution, this cannot be true. Consider, for example, if the mean of the level-0 belief distribution equals the mean of the level-0 action distribution.

[31] With suitable functional form assumptions, decision noise and level-0 belief heterogeneity would be separately identified parametrically. The assumptions on the distribution of decision noise would need to imply a different action distribution for higher level players than the distribution of level-0 beliefs does.

[32] In order to find the maximum likelihood estimates, we numerically maximise the log-likelihood. We start out searching over a grid covering the full range of parameter values. All sets of possible parameter values are bounded except for the standard deviation of the bounded normal distribution. Once the vector of parameter values is found which gives the highest likelihood, a new, finer grid is defined around this set. The new grid includes the parameter values that were neighbours of the maximising set in the previous, coarser grid. The calculation is iterated until the mesh size is at least three orders of magnitude smaller than the parameter value.

Table 3
Estimated level-k distribution.

Parameter	l_0	l_1	l_2	l_3
Estimate	0.37 (0.057)	0.47 (0.058)	0.15 (0.042)	0.01 (0.016)

Notes: The table presents the results from a maximum likelihood estimation of the structural model as outlined in Section 5.1. This table only presents the results for the level-k distribution, but the level-0 action and belief distribution were estimated simultaneously. Those results are reported in Table 4. Bootstrapped standard errors are given in brackets. These are obtained from 200 iterations of our estimation when sampling 84 observations from our data.

Appendix A.2 shows by means of Monte Carlo experiments that with higher-level decision errors our estimator of ξ_σ is biased, while there is no such evidence for any other parameter. Given that decision noise and heterogeneity of level-0 beliefs of higher-level players are not separately identified, the estimator 'absorbs' decision errors of higher-level players as heterogeneity in the level-0 belief. This is why we do not explicitly specify decision errors of higher level players in our empirical model. Given this modelling choice, it is important to keep in mind that a high estimate of ξ_σ might reflect the presence of decision noise of higher level players.

5.3. Estimation results

Level-*k* distribution Table 3 shows the estimation results for the level-k distribution. We estimate 47% of the participants to be level-1 reasoners, and 15% to be level-2 reasoners. We estimate only 1% of the participants to be level-3 reasoners. This is similar to the classification results in Section 4.2. We estimate that more than one third of the players (37%) are level-0 players.

The estimates in Table 3 on the relative frequencies of level-k reasoners, conditional on $k \geqslant 1$, resemble the relative fractions of level-1, level-2 and level-3 reasoners found in the literature for various games (see Camerer et al., 2004; Costa-Gomes and Crawford, 2006; Crawford and Iriberri, 2007a, etc.). This is reassuring and adds to the earlier evidence that the level-k model can have predictive power for the distribution of actions as a function of population and game characteristics.

However, our estimate of the fraction of non-strategic reasoners is substantially higher than previously estimated. Nagel (1995) associates certain actions with level-0 play and estimates between 2% and 17% of the population to be level-0 reasoners. Camerer et al. (2004) parameterise the type distribution as Poisson and estimate the mean to be approximately 1.5, which by the distributional assumption corresponds to roughly 22% non-strategic play.

Some of the level-k literature suggests that level-0 players only exist in the heads of other players. The evidence presented in this paper sheds doubt on this assumption.

One reason for the heterogeneity of findings might be the ability of different methodologies to uncover the fraction of level-0 players. We believe that our methodology is well-suited to uncover this fraction in a given subject pool. Other reasons for the heterogeneity might be the subject pool and the amount of testing and training that is done before the experiment, which might influence the capability of players to play strategically.[33] In our study we did not play any test rounds involving strategic situations. We believe this approach to be appropriate for studying one shot games.[34] However, another explanation for the high fraction of level-0 reasoners is that our experiment by its design might induce lower levels of reasoning than a more conventional setup.

Level-0 actions and beliefs Table 4 presents our estimates of the parameters characterising the distribution of actions of level-0 players and the beliefs of higher level players regarding their play.

We estimate the mean of the level-0 action distribution to be 58.38 and the mean of the level-0 belief distribution to be 54.01. These are close to the estimates we obtained non-parametrically in Section 4 and again consistent with each other. None of the two is significantly different from 50. We estimate the level-0 action distribution to have a standard deviation of 19.73. This is as well similar to the non-parametric estimate obtained before.[35]

[33] Examples of studies which did perform some testing or training are: Stahl and Wilson (1995), who went through practice exercises aimed at learning how the own pay-off is calculated given ones "beliefs about the other participants' choices" (pp. 250–251), and find a fraction of level-0 players of 16% or 18%, depending on the estimation procedure (p. 239); Bosch-Domènech et al. (2002), who examined comments provided with responses to newspaper experiments, which had been published by the newspapers with explanatory cover stories (p. 1691 and Table 1), and find a fraction of level-0 players of 12% (p. 1692).

[34] One robustness check for these results is the use of the final decisions rather than the suggested decisions. The results from the estimation are similar, except for the mean of level-0 actions. The estimate of this is lower at 48.83 (see Appendix A.3 for all results). This is likely the case since level bounds are unchanged and the persuasion that occurs during the communication is mostly picked up in the level-0 action distribution.

[35] When estimating the level-0 action distribution with a beta distribution that nests the uniform distribution, the estimated distribution is very similar in shape to the one found here. We estimate the level-0 belief distribution to have a standard deviation of 16.28, but our estimator is biased for this parameter, as discussed above.

Table 4
Estimated level-0 actions and beliefs.

Parameter	η^0_μ	η^0_σ	ξ_μ	ξ_σ
Estimate	58.38	19.73	54.01	16.28
	(7.09)	(3.45)	(2.49)	(2.41)

Notes: The table presents the results from a maximum likelihood estimation of the structural model as outlined in Section 5.1. This table only presents the results for the level-0 action and belief distribution, but the level-k distribution was estimated simultaneously. Those results are reported in Table 3. Bootstrapped standard errors are given in brackets. These are obtained from 200 iterations of our estimation when sampling 84 observations from our data.

6. Concluding comments

In this study we present an experimental design which allows to investigate the anchoring element of the level-k model, the level-0 actions and beliefs. We use this design with a standard 'beauty contest' game. Consistent with a common assumption of many versions of the level-k model, we find that the communicated level-0 beliefs are mostly 50. However, in both the level-0 beliefs and level-0 actions, upward deviations are common. One plausible determinant for level-0 actions and beliefs might be salience. We also estimate that one third of the participants are playing non-strategically. This finding is subject to the caveat that participants might have reduced incentives to think hard about the game in a team setup and might not fully articulate their reasoning. The actions of these level-0 players are distributed significantly different from a uniform distribution.

In order to gain these insights we present a novel experimental design. The novel features of our design are that the participants (i) state their individual reasoning in close temporal proximity to the reasoning process itself, and (ii) are supplied an incentive to state their reasoning as fully and clearly as possible. The written accounts provide an immediate insight into the individual's reasoning in addition to the participant's action in the game. This relaxes the need to make assumptions on aspects of the reasoning in order to interpret the player's action. The level of reasoning, for example, can be determined without particular assumptions on the level-0 belief or population belief distributions. In addition, it is possible to infer level-0 beliefs without any preconception of their distribution. Obtaining an incentivised written account of individual reasoning also relaxes the need to design complex games with a view to drawing inference from actions alone. It allows us to learn about individual reasoning in games which are economically interesting even when actions alone are not informative. The design is generally applicable to one shot games and we hope can prove useful for other purposes in experimental economics.

Acknowledgments

We are indebted to Madhav Aney, Ayala Arad, Dan Benjamin, Colin Camerer, Mariano Cena, Syngjoo Choi, Vincent Crawford, Ziad Daoud, Erik Eyster, Nagore Iriberri, Gilat Levy, Rosemarie Nagel, Michele Piccione, Mireia Raluy i Callado, Ariel Rubinstein, Reinhard Selten, the editor, and four referees for helpful discussions and comments on earlier drafts. We are especially grateful to Georg Weizsäcker for his initial encouragement, ongoing support and many very helpful discussions. We thank seminar audiences at Cornell, ESA International Meeting 2009, ESA North America Meeting 2009, GfeW Meeting 2008, LSE, Nordic Behavioural and Experimental Conference 2008, Universitat Pompeu Fabra, and University of East Anglia for comments. Kara Contreary, Jan Zapal, and Piotr Zurawski provided excellent research assistance. We are grateful for the financial support of the Michio Morishima fund (STICERD), the London School of Economics and the German Association of Experimental Economists (GfeW, Heinz-Sauermann Förderpreis 2007). This work was partly accomplished at the Universitat Pompeu Fabra in Barcelona, Spain, and benefited from its hospitality. Stefan Penczynski gratefully acknowledges the financial support of the German Academic Exchange Service (DAAD Doktorandenstipendium Nr. D/05/44596) during his stay in Barcelona.

Appendix A

A.1. Transformation of bounded normal

Lemma 1. *A random variable which is distributed as bounded normal on* $[0, 100]$ *with parameters* (ξ_μ, ξ_σ) *will still be distributed as bounded normal when applying a multiplicative transformation with a factor a, with parameters* $(a\xi_\mu, a\xi_\sigma)$ *and support* $[0, a100]$.

Proof. For a monotonic, differentiable transformation of a random variable B, say $X = g(B)$, the distribution of X is found as $f_X(x) = f_B(g^{-1}(x)) \cdot |g^{-1'}(x)|$ where the support needs to be suitably changed. In our case $g(b) = p^k b$, $g^{-1}(x) = x/p^k$ and $g^{-1'}(x) = 1/p^k$ and

$$f(x; \bullet) = \frac{\frac{1}{p^k \xi_\sigma}\phi(x/p^k; \boldsymbol{\xi})}{\Phi(100; \boldsymbol{\xi}) - \Phi(0; \boldsymbol{\xi})} = \frac{\frac{1}{p^k \xi_\sigma}\phi(y; p^k \boldsymbol{\xi})}{\Phi(p^k 100; p^k \boldsymbol{\xi}) - \Phi(0; p^k \boldsymbol{\xi})}$$

Table A.1
Data generating process and results of Monte Carlo studies.

Parameter	l_0	l_1	l_2	l_3	ξ_μ	ξ_σ	η^0_μ	η^0_σ
Panel A:	**0.360**	**0.470**	**0.150**	**0.020**	**54.32**	**16.56**	**67.28**	**23.35**
Baseline	0.384	0.465	0.139	0.012	53.52	16.30	66.83	25.01
	(0.069)	(0.068)	(0.044)	(0.010)	(2.60)	(2.09)	(8.31)	(7.77)
Noise	0.384	0.462	0.139	0.015	53.23	20.35	66.76	24.21
	(0.063)	(0.065)	(0.045)	(0.011)	(3.63)	(2.83)	(6.68)	(6.13)
Panel B:	**0.000**	**0.734**	**0.234**	**0.032**	**54.32**	**16.56**	–	–
Baseline	0.017	0.744	0.212	0.027	53.75	16.54	31.91	1.16
	(0.012)	(0.053)	(0.052)	(0.016)	(1.98)	(1.54)	(15.05)	(2.43)
Noise	0.020	0.747	0.210	0.023	53.58	20.47	32.54	2.28
	(0.011)	(0.052)	(0.051)	(0.013)	(2.90)	(2.32)	(17.08)	(6.32)

Notes: The bold numbers show the parameter values taken to generate 100 datasets with this data generating process. The other rows of the table report the mean (and standard deviation) of the parameter estimates obtained when running our estimation on these 100 datasets.

where the equality of the nominators follows by straight-forward manipulation of $\phi()$ and the denominator follows from

$$\Phi(T;\boldsymbol{\xi}) = \int_{-\infty}^{T} \frac{1}{\xi_\sigma}\cdot\phi\left(\frac{x-\xi_\mu}{\xi_\sigma}\right)dz = \int_{-\infty}^{p^kT} \frac{1}{p^k\xi_\sigma}\cdot\phi\left(\frac{w-p^k\xi_\mu}{p^k\xi_\sigma}\right)dw = \Phi\left(p^kT;p^k\boldsymbol{\xi}\right)$$

after a change of variable $w = p^k z$. $\square$

A.2. Monte Carlo studies

This section presents results from several Monte Carlo studies, which analyse the properties of our maximum likelihood estimator. We use the model outlined in Section 5.1 to generate data for 84 participants. By design, the Monte Carlo results are specific to the parameters for which data is generated. We use the parameter estimates presented in Section 5 as parameters in the data generating process.

Using these fractions of different levels, we randomly assign levels of reasoning to the 84 synthetic data points. From these we generate level bounds as follows. Firstly, there are participants whose communication does not enable any classification. We reflect this by not giving bounds with probability 0.17, comparable to the 16.7% of subject in the experiment. In our sample, 27.1% of the level classifications have an interval of two possible level values and 1.4% of three possible levels. For those data points which are assigned level bounds, we generate an upper bound which is one level above the true level with probability 0.2 and similarly (and independently) one level below for the lower bound. The probability of 0.2 is chosen conservatively, since this data generating process gives slightly more noisy classifications than we really have: 32% of the generated data points have two and 4% have three possible levels.

Table A.1 presents results based on 100 Monte Carlo runs. Panel A reports results when the estimates from Tables 4 and 3 are used to generate the data. Panel B reports a robustness check when the data generating process assumes no level-0 players ($l_0 = 0$), and scales the remaining fractions up proportionally. In the “baseline” studies, the ‘true’ parameter is always within 1.5 standard deviations of estimates from the MC studies. Therefore our estimator generates estimates which are not significantly different from the true parameter values, despite having a low standard deviation, especially for the fractions of different level players. We conclude from these results that the estimator is indeed able to uncover the parameters of the model.

The above described Monte Carlo studies do not account for decision noise of higher level players. In order to investigate how such decision noise influences the characteristics of the estimator, we add decision noise in the data generating process. We calibrate the noise with the help of the 36 observations of communicated level-0 beliefs. We can calculate the difference between their suggested decision in the experiment and the one that would result theoretically from their level,[36] a degenerate population belief and their communicated level-0 belief. Table A.2 reports statistics of this difference, first for all level-k players with $k > 0$ and then by level of reasoning. We use the standard deviation of 6.75 to calibrate the (bounded) normally distributed decision noise in the data generating process for all decisions of level-k players with $k > 0$, irrespective of their level.[37]

The “noise” studies in both panels show that the majority of the results are unchanged, except for the estimator of the standard deviation of the level-0 belief distribution. The mean of the estimates of this parameters are more than a standard

[36] For 27 non-level-0 subjects, the lower and upper bounds of level of reasoning coincide. For 4 more we use the level with the highest likelihood contribution from the estimation.

[37] We assume the noise to be bounded such that the suggested decision is always in [0, 100]. We assume the mode to be 0. With the standard deviation being small compared to the range of the interval, the mean is very close to 0 for common non-equilibrium decisions.

Table A.2
Inferred decision noise.

Deviation	Mean	Std. Dev.	Median	Min.	Max.	N
All	1.55	6.75	0.0	−12	22	31
Level-1	1.86	7.20	0.5	−12	22	22
Level-2	1.25	6.04	0.5	−6	11	8
Level-3	−3.00	–	−3.0	−3	−3	1

Table A.3
Estimated level-k distribution.

Parameter	l_0	l_1	l_2	l_3
Estimate	0.361	0.477	0.148	0.015
	(0.054)	(0.057)	(0.037)	(0.009)

Notes: The results reported are obtained when using the final decisions in our estimation. Our original results are reported in Table 3.

Table A.4
Estimated level-0 actions and beliefs.

Parameter	η^0_μ	η^0_σ	ξ_μ	ξ_σ
Estimate	48.83	22.58	53.75	14.38
	(6.03)	(2.82)	(1.97)	(1.64)

Notes: The results reported are obtained when using the final decisions in our estimation. Our original results are reported in Table 4.

Table A.5
Classification match across research assistants.

	Coinciding RA classifications					Total
	5	4	3	2	1	
Lower bounds	44	19	13	2	–	78
Cumulative fraction	0.56	0.81	0.97	1.00	–	
Upper bounds	20	23	27	6	2	78
Cumulative fraction	0.26	0.55	0.90	0.97	1.00	

Notes: The first and third row of the table presents for how many subjects a number of x (between 5 and 1) research assistants' classifications of an upper or lower bound coincide, respectively. For example, for 44 subjects all 5 research assistants stated the same lower bound. In the second and fourth rows the table presents cumulative fractions, i.e. for how many subjects do x or more research assistants agree on a lower or upper bound, respectively.

deviation above the true parameter value. This implies that the estimator for the standard deviation of the level-0 belief distribution is biased upwards because it picks up the level-independent decision noise. However, the remaining estimates are similarly close to the true parameter values as without the noise.

A.3. Estimation with final decisions

Tables A.3 and A.4 report the estimation results corresponding to Tables 3 and 4 when final decisions are used instead of suggested decisions.

A.4. Classification agreement

In order to test the classification for stability and replicability, 3 further RAs individually classified the data. Table A.5 shows that in 97% of the messages, 3 or more out of 5 RAs put the exact same level lower bound. The messages are an excellent coordination device and the classification procedure is shown to be very well replicable.

Appendix B. Experiment and classification instructions

This appendix reproduces all instructions shown to the participants in our experiment, conducted in the experimental laboratory of Royal Holloway on the 29th of May 2008 and the 28th of November 2008 (Section B.1) as well as the instructions which were provided to the research assistants in order to explain how to classify the written accounts of reasoning (Section B.2).

B.1. Experiment

The experiment instructions were distributed sequentially before the Test Period, Part I and Part II and read aloud by the experimenter.

Welcome to the experiment!

Introduction

You are about to participate in an experiment in team decision making. The experiment is funded by the Michio Morishima fund, the London School of Economics and the German Society of Experimental Economic Research. Please follow the instructions carefully.

In addition to the participation fee of £5, you may earn a considerable additional amount of money. Your decisions and the decisions of the other participants determine the additional amount. You will be instructed in detail how your earnings depend on your and the others' decisions. All that you earn is yours to keep, and will be paid to you in private, in cash, after today's session.

It is important to us that you remain silent and do not look at other people's screens. If you have any questions or need assistance of any kind, please raise your hand, and an experimenter will come to you. If you talk, exclaim out loud, etc., you will be asked to leave. Thank you.

Since this is a team experiment, you will at various times be matched randomly with another participant in this room, to form a team that plays as one entity. Your team's earnings will be shared equally between you and your team partner.

The experiment consists of two parts (**Part I** and **Part II**) which are independent of each other and feature different tasks. Part I consists of three rounds and Part II consists of four rounds. However, the way you interact as a team to take decisions will be the same throughout the two parts.

Now, let us explain how your ***Team's Action*** is determined. In fact, both your team partner and you will enter a ***Final Decision*** individually and the computer will choose randomly which one of your two final decisions counts as your team's action. The probability that your team partner's final decision is chosen is equal to the probability that your final decision will be chosen (i.e. your chances are 50:50). However, you have the possibility to influence your partner's final decision in the following way: Before you enter your final decision, you can propose to your partner a ***Suggested Decision*** and send him one and only one text ***Message***. *Note that this message is your only chance to convince your partner of the reasoning behind your suggested decision. Therefore, use the message to explain your suggested decision to your team partner.* After you finish entering your suggested decision and your message, these will be shown to your team partner. She/he will then make her/his final decision. Similarly, you will receive your partner's suggested decision and message. You will then make your final decision. As outlined above, once you both enter your final decision, the computer chooses randomly one of your final decisions as your team's action.

If you have any questions at this point, please raise your hand. In order for you to get familiar with the messaging system, you will now try it out in a **Test Period**. Please turn the page for further instructions.

Test period

A participant in this room is now randomly chosen to be your team partner. The **Test Period** has two rounds, with one question to answer in each round. Since this is only a test, your earnings will not depend on any decision taken now. In both test rounds you will need to answer a question about the year of a historic event. The team that is closest to the correct year wins.

As described, you will be able to send one ***Suggested Decision*** with your proposed year and an explaining ***Message***. After having read your partner's suggested decision and message, you will enter your ***Final Decision***. As described earlier, either your or your partner's final decision will be chosen randomly to be your ***Team's Action***.

The messenger allows ***Messages*** of any size. However, you have to enter the message line by line since the input space is only one line. Within this line you can delete by using the usual "Backspace" button of your keyboard. By pressing "Enter" on the keyboard, you add the written sentence to the message. Please note that only added sentences will be sent and seen by your partner. *The words in the blue input line will* ***not*** *be sent*. You can always delete previously added sentences by clicking the "Clear Input" button. The number of lines you send is not limited. You can therefore send messages of any length. You finally send the message to your partner by clicking the "Send Message" button.

When you are ready, please click the "Ready" button to start the **Test Period**.

Start Part I

You are about to start Part I of the experiment. You are now randomly matched with a new team partner. For each of the next three rounds you will be matched with a new team partner, i.e. in each of the following rounds you will play with a different person.

In each round, once all teams' actions have been taken, the computer will let you know which of your final decisions has been chosen randomly as your team's action. It will also let you know all other teams' decisions, whether your team won the round and your personal earnings. In each round, the winning team earns £20 (£10 per team player).

Then the next round of the game follows. It will feature an identical task, but you will be matched with a new team partner.

Your task is the following:

Your team and all other teams will take their ***Team's Action*** by choosing a number between 0 and 100. 0 and 100 are also possible. Only whole numbers will be accepted. From all teams' actions, the computer will calculate the ***Average***. Two thirds ($\frac{2}{3}$) of this average will be your ***target*** number. The winning team will be the one that is ***closest to two thirds of the average***. If two or more teams are equally close, the prize will be randomly given to one of these teams.

As described earlier, you will send your team partner a ***Suggested Decision*** and a ***Message***. Remember to explain in the message your reasoning behind your suggested decision. (*And note again that the words in the blue input line will* ***not*** *be sent. Press "Enter" to add them to the message*.) After this information is exchanged, both of you enter your ***Final Decision***, from which the computer randomly chooses the ***Team's Action***.

When you click the "Ready" button, you will start the first round of **Part I** of the experiment.

Start Part II

[A different game was played here. Instructions omitted.]

B.2. Classification

The self-contained classification instructions were given to the RAs. Part 2 was handed out once the classification of Part 1 was finished.

Part 1

In the following we will describe the classification process for the analysis of our experiment. We, Konrad and Stefan, assume that you are familiar with the level-k model as it has been introduced by Nagel (1995) or represented by Camerer et al. (2004). However, in order to clarify potential questions of terminology, the appendix on page 55 reproduces the main features of the model in the terminology used in this document.

The classification proceeds in two steps, Part 1 and Part 2. You are now provided by us with the transcripts for Part 1. The transcripts differ in the amount of information about the decisions taken. Only in Part 2 will you see the choices of the players that were made.

After your individual classification of each part, you will meet with your co-classifier to reconcile your classification. In this process, try to agree on common classifications if possible and note them in the third sheet. If an agreement is not possible and you keep your initial individual classification, simply note nothing in the third sheet. After you finished this process for Part 1, you will hand in the three sheets and we will provide you with the material for Part 2. If you have questions about the procedure at any point, simply write an email to us and we will clarify any point in a mail to both of you.

For Part 1, follow the instructions of this booklet now. Read them entirely to get an overview and then start the classification. Please read the messages of each player in Period 1 and note for each player the minimum level of reasoning, the level-0 belief mean and the population distribution. Below you find detailed instructions for classifying each player. Please limit yourself to making inferences only from what can clearly be derived from the message stated, i.e. do not try to think about what the player *might have thought*.

IMPORTANT: When you think that the information does not clearly lend itself to any inference, simply do not note any classification. Consequently, do not note anything if no statement has been made! It seems that this time the statements are less clear than in the first round. Please note only those classifications for which you are certain. Make use of the comments space if you are not certain but still want to indicate a feature of the reasoning. Similarly, please comment if the statement exhibits some argument that does not fit the level-k model as we present it here.

Levels: Minimum lower bounds

For the minimum lower bound on the level of reasoning, you should ask yourself: "What is the minimum level of reasoning that this statement clearly exhibits?" Once noted, you should be able to say to yourself: "It seems impossible that the players' level of reasoning is below this number!"

Here we ask you to be very cautious with the classification, not giving away high levels easily. ***Please only write down the highest lower-bound for which you are absolutely certain!*** In Part 2 you will be asked to classify the maximum lower bounds of the level of reasoning. This will be the time to be generous with the interpretation of the statements.

Level 0 The player does not exhibit any strategic reasoning whatsoever. Different versions of this might be randomly chosen numbers, misunderstanding of the game structure or giving other non-strategic 'reasons' for picking a number, e.g. taste. Important is that no best-responding to the others' play occurs. There could be considerations of what others might play, but without best responding to it. Examples[38]: "Let's use 50. This is the average between 0 and 100." "It's random, so let's guess something." "My favourite number is 74."

Level 1 This player best responds to something (calculates 'two thirds'). However, he does not realise that others will be strategic as well. Example: "They will all go for a number of about 50–55. So we should do something like 35."

Level 2 This player not only best responds (calculating 'two thirds'), but also realises that other players best respond as well. At level 2 the question about the extent of strategic reasoning of other players can come up. In the theory, this is reflected as a population belief on levels 0 and 1. Example: "Thinking that others play 60, everybody will play 40. So, we should be more clever and play two thirds of 40." "Some will play just play 90, while others will think and play 60 in response. We should therefore play somewhere between 60 and 40."

Level 3 This player realises that others could be level 2 and reacts by best responding to this as well. Put differently, he realises that others realise that others best respond as well. As for all players above level 1, the extent of strategic reasoning by others is important for level 3 reasoners as well. In addition, they have to ask themselves how the level-2 players think about the distribution of level 1 and level 0 players.

Level 4, 5, ... The process continues to higher levels. More levels of best responses and higher orders of beliefs become relevant.

Level 0 belief mean

If the comment hints toward a value of the mean of the level-0 distribution, then indicate this value as level-0 mean. Remember, the level-0 mean is the starting point of the reasoning. Players of positive level start best responding on this number. Please note a number as level-0 mean only when this number is not logically derived through level reasoning or the like. If an interval is indicated, please note the average of the lower and upper bound. For example, 'I think the others play around 50–60.' can be noted as a mean of 55. If only a qualitative statement is made about the level-0 mean, try to quantify it if possible. Otherwise, please write a short comment that indicates what is written down. Similarly, if a distribution is specified, please comment precisely on the relevant passage.

The literature usually assumes a mean of 50. Be reminded that this is only a common assumption which should not influence your considerations at this point.

Population belief

The population belief distribution $g_k(h)$ of a player of level k gives the fraction in the population he expects to be of level h. By definition, for level 0 players, the population belief is irrelevant. Level 1 players are defined as believing that all others are level 0. Hence, differences in the population belief distribution can only show up for players who are level 2 or higher. Therefore, we do not expect any statement from you for reasoners below level 2. But even if the level is 2 or higher, be reminded that at points where you think the information does not lend itself to any inference, simply do not note any classification.

We want to distinguish two sorts of population beliefs, distinguished by the degenerateness of the population distribution.

Degenerate Under a degenerate population belief, a player believes that *all* other players reason exactly one level below themselves. A level 5 player believes everybody else to be level 4. Higher order beliefs are also degenerate. So a level 3 player would think that all others (who are believed to be level 2 players) will believe that all others are level 1 and so on. Example: "Thinking that others play 60, everybody will play 40. So, we should be more clever and play two thirds of 40."[39]

Non-degenerate A non-degenerate population belief gives non-zero probability to more than one lower level. In such a case, a level 2 player believes that both level 0 and level 1 subjects are in the game with positive probability. Higher order beliefs could be either degenerate or non-degenerate. An example would be: "Some will play just play 90, while others will think and play 60 in response. We should therefore play somewhere between 60 and 40." (In the unlikely case, that the population belief is of the 'degenerate' type but some higher order beliefs are not, please make a note.)[40]

Model and terminology

The level-k model of bounded rationality assumes that players only think through a certain number (k) of best responses. The model has four main ingredients:

[38] All examples have been made up for illustrative purposes.

[39] Nagel (1995) proposed such a population belief.

[40] Camerer et al. (2004) proposed a truncated Poisson distribution as population belief distribution.

Population distribution This distribution on $\mathbb{N}_0$ reflects the proportion of types with a certain level k.

Level-0 distribution By definition, a level-0 player does not best respond. Hence, his actions are random to the game and distributed over the action space, which in our case is $\mathcal{A} = \{\{0\}, \{1\}, \{2\}, \ldots, \{99\}, \{100\}\}$.

Level-0 belief In the model, players with $k > 0$ best respond to what they believe the level-0 players play. Their level-0 belief might not be consistent with the level-0 distribution. For best responding, all that matters of the level-0 belief is the mean, which lies in $[0, 100]$. It is frequently assumed that the level-0 distribution and the level-0 belief are consistent, but for the classification this is irrelevant.

Population belief Players do not expect other players to be of the same or a higher level of reasoning. For a level-k player, the population belief is therefore defined on the set of levels strictly below k. It follows that level-0 players have no defined belief, level-1 players have a trivial belief with full probability mass on $\{0\}$, level-2 players have a well defined belief on $\{\{0\}, \{1\}\}$. From level 3 higher order beliefs are relevant as level-3 players have to form a belief about level-2's beliefs.

Part 2

The classification proceeds in two steps, Part 1 and Part 2. You are now provided by us with the transcripts for Part 2. You can now see the choices of the players that were made.

Please consider the information on each player in Period 1 and note for each player the maximum lower bound of level of reasoning and whether the equilibrium has been identified and whether dominance reasoning has been applied in the excel sheet provided. Below you find detailed instructions for classifying each player. Please limit yourself to making inferences only from what can clearly be derived from the message and the action data, i.e. do not try to think about what the player might have thought.

Be reminded that when you think that the information does not clearly lend itself to any inference, simply do not note any classification. Consequently, do not note anything if no statement has been made! It seems that this time the statements are less clear than in the first round. Please note only those classifications for which you are certain. Make use of the comments space if you are not certain but still want to indicate a feature of the reasoning. Similarly, please comment if the statement exhibits some argument that does not fit the level-k model as we present it here.

Levels: Upper bounds

The upper bounds should give the maximum level of reasoning that could be interpreted into the statement. Therefore, you should ask yourself: "What is the highest level of reasoning that can be underlying this statement?" Once noted, you should be able to say: "Although maybe not clearly communicated, this statement could be an expression of this level. If the player reasoned higher than this number, this was not expressed in the statement!"

Please refer to the level characterisations in Part 1 of the instructions.

Type: Equilibrium identification

With this dummy, you indicate whether the player realised that the unique equilibrium is 0. For this he has to *mention* the equilibrium action 0. It is not enough to describe a process of downward convergence. The equilibrium might be mentioned anywhere in the statement, so it is irrelevant whether he stops reasoning when he found the equilibrium strategies or whether finds further arguments not to play 0. People will not necessarily use the word 'equilibrium', but they might describe that 'theoretically everybody should play 0' or that 'the process will be going down to 0'.

Set the dummy to 0 if the equilibrium was not identified and to 1 if it was.

Type: Dominance reasoning

With this dummy, you indicate whether the reasoning applied the concept of dominance for the explanation. This is defined as involving iterative deletion of dominated strategies and randomly playing one of the remaining actions or best responding to a distribution over the partner's remaining action space. People will not necessarily use the word 'dominance', but they might describe that 'playing above 66 makes no sense'. Note that 'everybody plays on average 50 so I should not play higher than 34' is *not* a dominance reasoning, because it rules out strategies based on a distribution on the full action space. Dominance reasoning rules out first and plays a best response then.

Set the dummy to 1 if dominance was used in the argument and to 0 if it was not.

References

Agranov, M., Caplin, A., Tergiman, C., 2013. Naive play and the process of choice in guessing games. Working paper. California Institute of Technology.

Arad, A., Rubinstein, A., 2012. The 11–20 money request game: Evaluating the upper bound of k-level reasoning. Amer. Econ. Rev. 102 (7).

Bosch-Domènech, A., Montalvo, J.G., Nagel, R., Satorra, A., 2002. One, two, (three), infinity, ...: Newspaper and lab beauty-contest experiments. Amer. Econ. Rev. 92 (5), 1687–1701.

Bosch-Domènech, A., Montalvo, J., Nagel, R., Satorra, A., 2010. A finite mixture analysis of beauty-contest data using generalized beta distributions. Exper. Econ. 13 (4), 461–475.

Breitmoser, Y., 2012. Strategic reasoning in p-beauty contests. Games Econ. Behav. 75 (2), 555–569.
Camerer, C., 2003. Behavioral Game Theory. Princeton University Press.
Camerer, C., Johnson, E.J., Sen, S., Rymon, T., 1993. Cognition and framing in sequential bargaining for gains and losses. In: Ken Binmore, A.K., Tani, P. (Eds.), Frontiers of Game Theory. MIT Press, pp. 27–48.
Camerer, C.F., Ho, T.-H., Chong, J.-K., 2004. A cognitive hierarchy model of games. Quart. J. Econ. 119 (3), 861–898.
Cooper, D.J., Kagel, J.H., 2005. Are two heads better than one? Team versus individual play in signaling games. Amer. Econ. Rev. 95 (3), 477–509.
Costa-Gomes, M.A., Crawford, V.P., 2006. Cognition and behavior in two-person guessing games: An experimental study. Amer. Econ. Rev. 96 (5), 1737–1768.
Costa-Gomes, M.A., Crawford, V.P., Broseta, B., 2001. Cognition and behavior in normal-form games: An experimental study. Econometrica 69 (5), 1193–1235.
Crawford, V.P., Iriberri, N., 2007a. Fatal attraction: Focality, Naïveté and sophistication in experimental 'hide and seek' games. Amer. Econ. Rev. 97 (5), 1731–1750.
Crawford, V.P., Iriberri, N., 2007b. Level-k auctions: Can a nonequilibrium model of strategic thinking explain the winner's curse and overbidding in private-value auctions?. Econometrica 75 (6), 1721–1770.
Crawford, V.P., Gneezy, U., Rottenstreich, Y., 2008. The power of focal points is limited: Even minute payoff asymmetry may yields large coordination failures. Amer. Econ. Rev. 98 (4), 1443–1458.
Crawford, V.P., Costa-Gomes, M.A., Iriberri, N., 2013. Structural models of nonequilibrium strategic thinking: Theory, evidence, and applications. J. Econ. Lit. 51 (1), 5–62.
Ericsson, K., 2002. Towards a procedure for eliciting verbal expression of non-verbal experience without reactivity: interpreting the verbal overshadowing effect within the theoretical framework for protocol analysis. Appl. Cogn. Psychol. 16 (8), 981–987.
Fischbacher, U., 2007. z-Tree: Zurich toolbox for ready-made economic experiments. Exper. Econ. 10 (2), 171–178.
Goeree, J.K., Holt, C.A., 2004. A model of noisy introspection. Games Econ. Behav. 46 (2), 365–382.
Ho, T.-H., Camerer, C., Weigelt, K., 1998. Iterated dominance and iterated best response in experimental 'p-beauty contests'. Amer. Econ. Rev. 88 (4), 947–969.
McKelvey, R.D., Palfrey, T.R., 1995. Quantal response equilibria for normal form games. Games Econ. Behav. 10 (1), 6–38.
Nagel, R., 1994. Reasoning and learning in guessing games and ultimatum games with incomplete information: An experimental investigation. PhD thesis. Universität Bonn.
Nagel, R., 1995. Unraveling in guessing games: An experimental study. Amer. Econ. Rev. 85 (5), 1313–1326.
Penczynski, S.P., 2012. Persuasion: An experimental study of team decision making. Working paper. University of Mannheim.
Rydval, O., Ortmann, A., Ostatnicky, M., 2009. Three very simple games and what it takes to solve them. J. Econ. Behav. Organ. 72 (1), 589–601.
Stahl, D.O., Wilson, P.W., 1995. On players' models of other players: Theory and experimental evidence. Games Econ. Behav. 10 (1), 218–254.