

Rationality and Observed Behavior

Francesco Cerigioni¹, Fabrizio Germano¹, Pedro Rey-Biel², and Peio Zuazo-Garin³

¹Universitat Pompeu Fabra and Barcelona GSE

²Universitat Ramon Llull, ESADE

³Universidad del País Vasco

July 12, 2018

ABSTRACT

Rationality and higher-order rationality are both central assumptions implicit in most of economic theory. Yet, there is no consensus on what is the most reliable methodology to find the empirical distribution of higher orders of rationality. We build on previous methodologies to empirically identify higher orders of rationality, by proposing a novel class of games with incomplete information, the *e-ring games*, that combines aspects of the seminal ring games used by Kneeland (2015), with an information and communication structure related to the email game of Rubinstein (1989). This allows us to overcome some of the shortcomings present in the literature and, for the first time, to check within subject consistency of identified rationality levels across games and methods. We find that it is feasible to use our e-ring games to classify levels of rationality of a highly skilled pool of experimental subjects. Moreover, we show that for the same subject, higher-order rationality is generally very variable across games. In fact, not only there is no individual classified in the same rationality level across all games, but also when allowing for much weaker consistency requirements, we find little evidence of within subject persistence. This casts doubts on using the standard concepts of rationality and higher order rationality as fixed behavioral benchmark in games and points toward taking a more game dependent approach.

(JEL C70, C91, D01, D83)

KEYWORDS: Rationality, Higher-Order Reasoning, Contingent Thinking, Strategic Uncertainty

1 Introduction

Central to economic theory is the assumption that individuals' actions not only satisfy rationality but also higher-order rationality, that is, besides being rational, individuals' actions should also be consistent with beliefs that their opponents are rational, with beliefs that their opponents have beliefs that their opponents are rational, and so on. An important part of the literature has focused on testing this assumption through experiments without consensus on which method is best. The

two main approaches for identifying higher-order rationality have consisted in direct elicitation of subjects’ first-order beliefs and matching these with choice data (Costa-Gomes and Weizsacker, 2008; Healy, 2011), as well as through the estimation of level- k behavioral models using choice data and, instead of eliciting beliefs, relying on often ad hoc structural assumptions on first-order beliefs (Costa-Gomes, Crawford and Iriberry, 2013; Burchardi and Penczynski, 2014). The direct elicitation method quickly runs into practical difficulties of identifying belief orders above the lowest ones, while the second method depends critically on the structural assumptions on beliefs. More recently, a major step forward towards an agreed methodology was made by Kneeland (2015), who used a class of games, *ring-network games*, first introduced by Cubitt and Sugden (1994), to achieve reliable choice based inference. In this class, players take part in a series of two-player normal form games in which they are matched with each other in an ordered sequence that associates each player with certain level of hierarchical thinking.

We argue that her innovative approach has three shortcomings: (i) the ordered opponent structure may actually frame players into thinking in levels, (ii) it allows for a relatively straightforward inductive step that subjects may use to identify the action that is consistent with the maximal (or n ’th) level of higher-order rationality within the given n -player game structure and (iii) it requires n players to test for n levels of higher-order rationality. The last point is obvious. To see the other two, notice that the opponent structure of the ring games is in one-to-one correspondence with the hierarchical belief structure necessary to play actions consistent with higher-order rationality. To understand how to play the game a subject is forced to form a hierarchy of beliefs. In particular, once a player has looked at the strategic situation of her opponent, it is just a minor inductive step to see that there is a repetition of the same situation for that opponent and for subsequent opponents all the way up to the opponent n steps up. This greatly facilitates and possibly even encourages higher-order reasoning, and may lead, not only to an overestimation of actions complying with higher-order rationality, but, moreover, may potentially result in the highest possible level (n) being particularly overrepresented, due to the ease of the inductive step.¹

In this paper, (a) we propose a new class of two-player games, which we call *e-ring games*, that overcomes the shortcomings mentioned above—at least for the (empirically relevant) lower-order rationality levels—and (b) we experimentally test the feasibility of our class of games and compare it at the individual level with other standard methods used in the literature to identify higher orders of rationality.

The key difference between our games and the ring games used in Kneeland (2015), besides having just two players, is that we add an additional layer of incomplete information. This makes less obvious not only the one-to-one correspondence between opponent structure and higher-order reasoning but also the inductive step. Importantly, the incompleteness of information is structured by means of messages that go back and forth between players as in the email game of Rubinstein

¹Kneeland (2015) already remarks the first two points and argues that it may in part be due to the weaker identification assumptions required by her method. She also points out that the ring games “may make iterative reasoning more natural.”

(1989), generating a natural one-to-one correspondence between messages and higher-order beliefs. However, in order to identify subjects in a reliable manner, our games have interim payoff matrices that vary with the number of messages received.² This allows us to associate different payoff matrices and hence actions to different levels of higher-order beliefs. At the same time, we do not rely on the exclusion restriction assumption made in Kneeland (2015) for our identification of players' orders of rationality.³ Instead, we rely on a revealed rationality approach that assigns players the maximal level of higher-order rationality consistent with the choices made (Tan and Werlang, 1988; Lim and Xiong, 2016; Brandenburger, Danieli and Friendenberg, 2017).

More specifically, in the *e-ring games* two players, a sender and a receiver, each need to choose one of four possible actions. The key assumption is that each player's own payoffs depend not only on the actions chosen by both players, but also on the number of messages that player received. That is, a player's own payoffs associated with the four possible actions when he has received k messages are different from the ones associated with $k + 1$ messages. There is a maximal number m of messages that any player can receive with an otherwise email game-like communication structure. The payoff of a sender with k messages depends on the actions of a receiver with k or $k + 1$ messages, whose payoffs in turn depend on actions of a sender with $k - 1$ or k messages and k or $k + 1$ messages respectively, and so on. The fact that messages are finite puts a natural limit to the number of levels that can be identified, as well as to the complexity of the game. In general, m messages are needed for each of the two players to test for up to $2m$ levels of higher-order rationality.

As mentioned above, we experimentally test the feasibility of our new type of games and compare it with other classic games used for the identification of higher orders of rationality (dominance solvable normal form games, beauty contests, *ring games*). To the best of our knowledge this is the first attempt in comparing different identification methodologies at the individual level. Moreover, we also check for the persistence of the identified rationality level within subject across games. Given the complexity of the type of games used in our experiment, we chose as our subject pool a group of highly qualified undergraduate engineering students from a top school in Madrid (Spain), with intense mathematical training, which should potentially allow for generally higher levels of rationality, at least in the e-ring and the ring games. Such subject pool also facilitates the identification of the inductive step.

In our experiment subjects play four types of games, namely, our e-ring games, the ring games presented in Kneeland (2015), two simple two-player 4×4 dominance solvable games and different

²This is a major difference with the email game of Rubinstein (1989), where given the fact that players face the same 2×2 payoff matrices for almost all messages received, they can choose simple cut-off strategies for when to switch strategy that can easily misidentify subjects' actual levels of rationality.

³The exclusion restriction assumption maintains that *subjects satisfying lower-order rationality do not respond to changes in higher-order payoffs*. It is a key methodological assumption pursued in Kneeland (2015). A main criticism is that subjects in general may change strategies even when not responding to changes in the payoffs of high-order opponents. To test this, Lim and Xiong (2016) have subjects play the ring games of Kneeland (2015) multiple times (as well as other games), and find up to 77% *non-compliance* with the assumption in the ring games, meaning that 77% of the experimental subjects chose different actions at least once.

versions of the beauty contest game presented in Nagel (1995). In particular, our subjects play a version of the beauty contest game where the average of all subjects' responses is multiplied by $1/3$ and another one in which it is multiplied by $2/3$. Finally, subjects also play a version of the beauty contest game, that we call *p-beauty game*, that uses the strategy method to obtain subjects' planned strategy with respect to any possible number (p) between 0 and 1 (excluded) which could multiply the average of subjects' responses. For e-ring games and ring games, we use sets of 8 different games with different payoffs in order to be able to classify subjects into orders of rationality.

Our first finding is that it is feasible to implement the e-ring game in a laboratory setting with a highly qualified subject pool, since a majority of subjects (76%) pass the demanding comprehension test, and their explicitly written rationale to their actions show that they understand the experimental instructions. More importantly, we find that the levels of higher-order rationality are very game dependent, at the individual and at the aggregate level.⁴ Our third finding is that the ring games show a significant downward jump in the number of subjects in the rationality classes after the first one, e.g. R1, and a significant upward jump in the highest possible rationality class. This may be evidence for the ease of inductive step effect for the ring games. By contrast, for the e-ring games, the distribution of types is rather evenly distributed across rationality classes. A fourth finding, uses treatment variation in the order of the tasks to show that when ring games are played first, subjects tend to be classified in a higher category on subsequent games which we believe is evidence for the framing effect for this class of ring games. Finally, it is interesting to notice that the classification obtained from the beauty contest varies a lot depending on the value of p . This may be due to the fact that identification of rationality levels from observed behavior in beauty contest games relies heavily on the structural assumptions on beliefs thus pointing out the necessity of a methodology such the one presented in Kneeland (2015) and in this paper.⁵

Related literature. Besides the approach of Kneeland (2015) and the related experimental literature on iterated dominance solvable games (Beard and Beil, 1994; Schotter, Weigelt and Wilson, 1994; Nagel, 1995; Costa-Gomes, Crawford and Broseta, 2001; Van Huyck, Wildenthal and Battalio, 2002; Rey-Biel, 2009) which are both very close to our own, two main methods for identifying higher-order rationality have consisted in direct elicitation of subjects' first-order beliefs and matching these with choice data (Costa-Gomes and Weizsacker, 2008; Healy, 2011) and through the estimation of level- k behavioral models using choice data and relying on often ad hoc structural assumptions on first-order beliefs (Costa-Gomes, Crawford and Iriberri, 2013; Burchardi and Penczynski, 2014). As stated in the introduction, the direct elicitation method quickly runs into practical difficulties of identifying orders above the lowest ones, while the second method depends critically on the structural assumptions on beliefs.

⁴Notice that the p -beauty game classification is based on the reasoning expressed in the comments section of the questionnaire.

⁵In particular, looking at the distribution in the p -beauty contest it is clear that many players that have no beliefs of order higher than 1 are misclassified in different categories due to the high variance their answers show regarding their beliefs about the behavior at order 0.

	$R0$	$R1$	$R2$	$R3$	$R4$
E-ring	0.13	0.22	0.21	0.25	0.20
Ring	0.17	0.31	0.16	0.12	0.25
4×4	0.02	0.27	0.47	0.14	0.10
2/3-BC	0.04	0.07	0.28	0.26	0.35
1/3-BC	0.09	0.41	0.34	0.10	0.07
p -BC	0.22	0.45	0.21	0.05	0.07

Table 1: Subjects classified by orders of rationality, by game

Also related is the paper of Lim and Xiong (2016). It takes the exclusion restriction assumption of Kneeland (2015) as its starting point and shows that it is often violated in experiments, including in the ring games of Kneeland (2015). Then it proposes an alternative class of games, the *chain games*, which are two-player bimatrix games that incorporate a ring-game like payoff structure between the payoffs of the two players. To identify higher-order rationality they rely on a simpler, revealed rationality approach that directly identifies behavior that does not contradict k -order rationality with k -order rationality. Thus they invoke a weaker *upper-bound assumption* instead of the exclusion restriction assumption. Our approach has in common with the one of Lim and Xiong (2016) that we also propose a class of two-player games and also replace the exclusion restriction assumption with the upper-bound assumption. However, we believe that the chain games proposed, still suffer potentially from a framing effect as well as an easy inductive step. Although we do not test the chain games, they have an ordered structure very similar to that of the ring games of Kneeland (2015) that forces subjects to think about each step of the hierarchy of beliefs. By contrast, our games involve uncertain states of the world. Subjects, at the interim stage, therefore face an opponent in different contingent situations, each of which in turn faces as opponent, the original subject, who from the point of view of that opponent considers as possible contingencies that the original subject already discards. While our games are admittedly more complex, they substantially weaken the chances that agents are framed by the game, and, moreover, they have an inductive step that should be less apparent, at least for subjects of orders of rationality less or equal to 4.

Finally, while this paper is the first, to the best of our knowledge, that experimentally compares different identification methods and the persistence of their classification at the individual level, we are not the first ones in testing whether within subject classifications are consistent across games. In fact, Georganas, Healy and Weber (2015) classify subjects in their experiment using k -level theory in two different class of games and then check the consistency of the classification

within subject and across games. While similar in spirit, the papers are very different. First of all, we compare for the first time different identification methods, while Georganas, Healy and Weber (2015) simply study the persistence of subjects' classification across two very specific class of games. Secondly, the identification methods used are very different. In Georganas, Healy and Weber (2015) subjects are classified using a maximum likelihood estimation while in this paper, as explained before, we use the revealed rationality approach. As a third and final point, here we test the standard rationality model while Georganas, Healy and Weber (2015) test the level-k model that, while similar, is based on very different assumptions.

The paper is organized as follows. In the next section, we describe our class of games. In Section 3, we present the experimental design while the experimental results can be found in Section 4. Section 5 concludes. The appendices contain a theoretical description of the class of games used in the paper as well as an English translation of the experimental instructions and the payoff matrices for all games used in the experiment.

2 E-Ring Games

We introduce a new class of two-player games of incomplete information, we refer to as *e-ring games*. Before defining the general class of games, we give a simple example that can identify higher-order rationality up to a level of 4. This is the type of game used in all our experimental sessions, although to make it simpler for subjects, in the experiment they either received one message or none.

Example (E-Ring Game of Depth 2). There are two players, a row player (player 1) and a column player (player 2), and three states of the world $\Theta = \{(1, 1), (1, 2), (2, 2)\}$ that have equal prior probability, where $\theta = (\theta_1, \theta_2) \in \Theta$ denotes the state, where player 1 receives θ_1 messages and player 2 receives θ_2 messages. Each player is initially informed about the number of messages he has received and his payoffs depend only on the number of messages he received. Note also that player 2 either has the same number or one more message than player 1 and each player either gets 1 or 2 messages. To compute the payoffs of the opponent, players can compute the number of messages received by their opponent as follows. Player 1 with 1 message knows his opponent has either 1 or 2 messages, each event with equal probability; player 1 with 2 messages knows for sure the other player also has two messages. Similarly, player 2 with 1 message, knows for sure his opponent also has 1 message; while player 2 with 2 messages knows his opponent has either 1 or 2 messages, each event with equal probability. Consider the following “stylized” payoff matrices of the players,⁶ where A, B, C, D are the actions of player 1 (row player) and a, b, c, d are the actions of player 2 (column player), and where $u_1(\theta_1)$ are the payoffs of player 1 when he receives θ_1 messages, and $u_2(\theta_2)$ the payoffs of player 2 when he receives θ_2 messages.

⁶The numbers 0, 1, 2, 3 can be thought of as minimal payoffs or also as sets of possible payoffs that can be used to allow more variation in payoffs, under the condition that if $u \in 0$, $v \in 1$, $x \in 2$, and $y \in 3$, then it must also be true that $u < v < x < y$ in any given matrix.

	<i>a</i>	<i>b</i>	<i>c</i>	<i>d</i>
<i>A</i>	2	1	2	2
<i>B</i>	0	0	0	0
<i>C</i>	1	3	0	1
<i>D</i>	1	2	1	0

 $u_1(\theta_1 = 1) =$

	<i>a</i>	<i>b</i>	<i>c</i>	<i>d</i>
<i>A</i>	2	2	1	2
<i>B</i>	1	0	3	1
<i>C</i>	0	0	0	0
<i>D</i>	0	1	2	1

 $u_1(\theta_1 = 2) =$

	<i>a</i>	<i>b</i>	<i>c</i>	<i>d</i>
<i>A</i>	1	0	0	0
<i>B</i>	1	0	0	0
<i>C</i>	1	0	0	0
<i>D</i>	1	0	0	0

 $u_2(\theta_2 = 1) =$

	<i>a</i>	<i>b</i>	<i>c</i>	<i>d</i>
<i>A</i>	2	0	0	1
<i>B</i>	2	0	1	0
<i>C</i>	1	0	3	2
<i>D</i>	2	0	1	1

 $u_2(\theta_2 = 2) =$

The above payoff structure has a unique interim rationalizable action for all players and player types. Player 2 with 1 message (payoff matrix $u_2(\theta_2 = 1)$) has a strictly dominant action to play a . Player 1 with 1 message (payoff matrix $u_1(\theta_1 = 1)$), seeing this and the fact that player 2 with 2 messages has b as strictly dominated action, (and knowing that he faces player 2 with $\theta_2 = 1, \theta_2 = 2$ with equal probability), has a unique strict best reply to play A . Player 2 with 2 messages (payoff matrix $u_2(\theta_2 = 2)$), given the above and seeing that player 1 with 2 messages has C as a strictly dominated action, (and again knowing that she faces player 1 with $\theta_1 = 1, \theta_1 = 2$ with equal probability), has a unique strict best reply to play a . Finally, player 1 with 2 messages (payoff matrix $u_1(\theta_1 = 2)$, knowing that he faces player 2 for sure with 2 messages and that she plays a as unique best reply, also has a unique strict best reply to play A . Thus $((A, A); (a, a))$ is the unique rationalizable strategy profile.⁷ □

Just like the email games of Rubinstein (1989), our games also provide a natural one-to-one correspondence between messages and higher-order beliefs that in turn allows us to associate actions to different levels of beliefs in a context of asymmetric information. An important difference with the email game, however, is that our games have interim payoff matrices that change with the number of messages received, making it much more difficult for a subject of, say level 1, to qualify as being of level 3 or 4.⁸ Notice that the addition of asymmetry of information is crucial to resolve what Kneeland defines as *catch-22* of this kind of identification exercise:

“A particularly salient effect of ring games (relative to standard normal form games) is that they may make iterative reasoning more natural. This might happen if the ring game highlights the higher-order dependencies between the players or if it induces backward induction

⁷In the experiment, player 1 by playing with 0 and 1 messages can be tested for R2 and R4 respectively while player 2 would be tested for R1 and R3. Notice that by making the matrix of player 2 with 0 message the matrix of player one with 1 message and changing the order of the other matrices accordingly one can test for the other levels of rationality for each player.

⁸This is a major difference with the email game of Rubinstein (1989), where given the fact that players face the same 2×2 payoff matrices for almost all messages received, they can choose simple cut-off strategies for when to switch strategy that can easily misidentify subjects' actual levels of rationality if one were to use the game with such objective in mind.

reasoning because of the presentation of the game. Here we face a catch-22: we must depart from typical games to achieve reliable choice based inference, but doing so unavoidably raises concerns of this sort.”

That is, to avoid highlighting the higher-order dependencies between the players, we add the minimal amount of informational asymmetry, namely due to uncertainty over at most two possible states of the world.

To make the whole reasoning clearer, the following is a general definition of an *e-ring game of depth m* that in principle allows to test up to $2m$ levels of higher-order rationality. The basic structure is such that a player of level $1, 3, \dots, 2m - 1$, has to play action a as player 2 (column player) of type $t_2 = 1, 2, \dots, m$, respectively, and similarly, a player of level $2, 4, \dots, 2m$ has to play action A as player 1 (row player) of type $t_1 = 1, 2, \dots, m$, respectively. Slightly modifying the payoff structure, generates a game where odd levels are played by corresponding types of player 1 and even levels by types of player 2.

Definition 1 *A two-player e-ring game of depth m is a list $\mathcal{G}_{ERG} = \langle I, \Theta, (A_i, u_i)_{i \in I}; \mathcal{B} \rangle$, where: (i) $I = \{1, 2\}$ is the set of players, (ii) $\Theta = \{(\theta_1, \theta_2) \in \{1, \dots, m\}^2 \mid \theta_2 - \theta_1 \in \{0, 1\}\}$ is the set of payoff states, (iii) $A_1 = \{A, B, C, D\}$, $A_2 = \{a, b, c, d\}$, are the sets of actions, (iv) $u_i : A_1 \times A_2 \times \Theta \rightarrow \mathbb{R}$ are utility functions given by following matrices:*

$$u_1(\theta) = \left\{ \begin{array}{l} \begin{array}{c} \begin{array}{c|c|c|c} a & b & c & d \\ \hline A & \mathbf{2} & \mathbf{1} & \mathbf{2} & \mathbf{2} \\ \hline B & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \hline C & \mathbf{1} & \mathbf{3} & \mathbf{0} & \mathbf{1} \\ \hline D & \mathbf{1} & \mathbf{2} & \mathbf{1} & \mathbf{0} \end{array} \\ \text{if } \theta_1 \text{ odd} \\ \\ \begin{array}{c|c|c|c} a & b & c & d \\ \hline A & \mathbf{2} & \mathbf{2} & \mathbf{1} & \mathbf{2} \\ \hline B & \mathbf{1} & \mathbf{0} & \mathbf{3} & \mathbf{1} \\ \hline C & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \hline D & \mathbf{0} & \mathbf{1} & \mathbf{2} & \mathbf{1} \end{array} \\ \text{if } \theta_1 \text{ even} \end{array} \right. \quad u_2(\theta) = \left\{ \begin{array}{l} \begin{array}{c|c|c|c} a & b & c & d \\ \hline A & \mathbf{1} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \hline B & \mathbf{1} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \hline C & \mathbf{1} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \hline D & \mathbf{1} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{array} \\ \text{if } \theta_2 = 1 \\ \\ \begin{array}{c|c|c|c} a & b & c & d \\ \hline A & \mathbf{2} & \mathbf{0} & \mathbf{0} & \mathbf{1} \\ \hline B & \mathbf{2} & \mathbf{0} & \mathbf{1} & \mathbf{0} \\ \hline C & \mathbf{1} & \mathbf{0} & \mathbf{3} & \mathbf{2} \\ \hline D & \mathbf{2} & \mathbf{0} & \mathbf{1} & \mathbf{1} \end{array} \\ \text{if } \theta_2 \text{ even} \\ \\ \begin{array}{c|c|c|c} a & b & c & d \\ \hline A & \mathbf{2} & \mathbf{0} & \mathbf{0} & \mathbf{1} \\ \hline B & \mathbf{2} & \mathbf{1} & \mathbf{0} & \mathbf{0} \\ \hline C & \mathbf{1} & \mathbf{3} & \mathbf{0} & \mathbf{2} \\ \hline D & \mathbf{2} & \mathbf{1} & \mathbf{0} & \mathbf{1} \end{array} \\ \text{if } \theta_2 \neq 1, \text{ odd} \end{array} \right. \quad (1)$$

where the bold numbers $\mathbf{0}, \mathbf{1}, \mathbf{2}, \mathbf{3} \in \mathbb{R}$ denote equivalence classes from which arbitrary utilities can be chosen, subject to $u \in \mathbf{0}$, $v \in \mathbf{1}$, $x \in \mathbf{2}$, and $y \in \mathbf{3}$, always implies $u < v < x < y$ in any

given matrix; and (v) $\mathcal{I} = \langle S_i, q_i \rangle_{i \in I}$ is the information structure, where: (v.i) $S_i = \{0, 1, \dots, m\}$ are the signal sets, and (v.ii) $q_i : S_i \rightarrow \Delta(S_{-i} \times \Theta)$ are belief maps given by:

$$q_1(s_1)[(s_2, (\theta_2; \theta_1))] = \begin{cases} 1/2 & \text{if } s_1 = \theta_1 < m \text{ and } s_2 = \theta_2 = \theta_1, \\ 1/2 & \text{if } s_1 = \theta_1 < m \text{ and } s_2 = \theta_2 = \theta_1 + 1, \\ 1 & \text{if } s_1 = \theta_1 = m \text{ and } s_2 = \theta_2 = \theta_1, \\ 0 & \text{otherwise,} \end{cases}$$

and:

$$q_2(s_2)[(s_1, (\theta_1; \theta_2))] = \begin{cases} 1/2 & \text{if } s_2 = \theta_2 < m \text{ and } s_1 = \theta_1 = \theta_2 - 1, \\ 1/2 & \text{if } s_2 = \theta_2 < m \text{ and } s_1 = \theta_1 = \theta_2, \\ 1 & \text{if } s_2 = \theta_2 = m \text{ and } s_1 = \theta_1 = \theta_2, \\ 0 & \text{otherwise,} \end{cases}$$

The information structure is such, if $\theta = (\theta_1, \theta_2) \in \Theta$ is the state of the world, then player 1 is informed about θ_1 and believes that $\theta_2 = \theta_1$ or $\theta_2 = \theta_1 + 1$ with equal probability; unless $\theta_1 = m$ in which case he knows for sure that $\theta_2 = m$. Analogously, player 2 is informed about θ_2 and believes that $\theta_1 = \theta_2$ or $\theta_1 = \theta_2 - 1$, also with equal probability; unless $\theta_2 = 1$ in which case he knows for sure that $\theta_1 = 1$. It can be checked easily that the game has as unique rationalizable action⁹ to play A or a for all types of player 1 and 2 respectively.

3 Experimental Design

The experiment consisted of four tasks and a non-incentivized questionnaire. In the first task, subjects chose an action in two, two player 4x4 dominance solvable games. In each of the subsequent two tasks, subjects chose actions in a set of eight ring games and eight e-ring games as described in the previous section. The set of eight ring games and the set of eight e-ring games were presented in different random orders to each of the subjects. In the final task, subjects were presented with the beauty contest game as in Nagel (1995) and had to choose a number for two different versions of the game (one where the average of all players' numbers was multiplied by 2/3 to determine the winner, and another where the average was multiplied by 1/3) and a more general version, where subjects were asked to explain a general strategy about how they would choose for any (unspecified) number p between 0 and 1 (both not included) that could be announced publicly in the beauty contest game. For this final task, subjects were told that they could either choose a number, a mathematical formula or provide any text which would show their reasoning process. We designed 8 treatments differing in three aspects: (i) whether the ring game was played before or after the e-ring game; (ii) whether the payoff matrices used in the ring and e-ring games were remained constant (non-permuted) across decisions while either varying the player's position (ring

⁹Or, more precisely, unique *interim correlated* rationalizable action; see Dekel, Fudenberg and Morris (2007) or Definition 8 in the appendix.

game) or the number of messages received (e-ring game), or whether the actions in such matrices were reshuffled; and *(iii)* whether subjects were asked to answer the 1/3 or the 2/3 versions of the beauty contest game before or after. A translation of the original Spanish instructions as well as the actual games used for each of the tasks can be found in the Supplemental Appendix.

Our experimental design intends to compare the e-ring and the ring games with those previously used in the literature (dominance solvable games such as our 4×4 games and the p -beauty game) to empirically classify individuals according to their revealed order of rationality as used in Brandenburger, Danieli and Friendenberg (2017).

In both the e-ring and the ring games, each subject can play four possible actions in each of the eight games for a total of 65,536 possible action profiles. In both the e-ring and the ring games, there are 801 action profiles that do not violate any of the predicted action profiles of types R1-R4, independently of subjects' role following the revealed rationality approach. Thus, it is unlikely for a subject to be assigned to a rational type by random chance. Hence, there is 1.2% probability of being identified as R1-R4 while playing randomly in either games.¹⁰

3.1 Laboratory Implementation

The experiment was conducted at the Engineering School of Universidad Carlos III in Madrid (Spain) in April, 2018. This particular school was selected due to being one of the most prestigious universities in the country. Accordingly, the average grade in the entrance to university exam of our pool of participants is 12 (out of 14 possible points). All undergraduate engineering students from the school were sent an email message announcing the experimental sessions and they were confirmed on a first-come first-served basis according to our sample size requirement. 229 students participated. No subject participated in more than one session. Subjects made all decisions using a booklet including all instructions in the order determined by their treatment assignment and the randomization of the order of eight ring and e-ring games, the answer sheets and a post-experimental questionnaire. Sessions were closely monitored resembling exam-like conditions in order to ensure independence across participants' responses. Instructions were read aloud and included examples of the payoff consequences of several actions in each of the tasks. Participants answered an understanding questionnaire prior to each of the tasks. Participants received no feedback after playing each of the games nor after finishing each of the tasks. Once all four tasks were completed, subjects filled up a questionnaire, which included non-incentivized questions about the reasoning process used to choose in each of the tasks, as well as questions about knowledge of game theory and demographics. Subjects were given 4 minutes to complete the first task, 20 minutes each for the second and third task and 9 minutes for the final task. The two experimental

¹⁰In the experiment, we used 4×4 versions of Kneeland's ring games in order to add a dominated action in each matrix to obtain the same misidentification probabilities between the ring and the e-ring games, given that the e-ring games require a dominated strategy in each payoff matrix. In a pilot experiment, we replicated Kneeland's experiment and used 3×3 versions of the e-ring game, and found identical qualitative results. However, for 3×3 versions we cannot assure that the probability of misidentification of the ring game and the e-ring game are the same.

sessions lasted around 75 minutes each.

Given the high cognitive load of our tasks, we decided to provide high monetary incentives for 10 randomly selected participants, instead of paying all subjects a lower amount of money. One game was randomly selected for payment at the end of the experiment for each of these 10 participants. Subjects were randomly and anonymously matched into groups of 2-players (e-ring and 4×4 games), 4-players (ring games) or all players (p -BC games) depending on the game selected, and were paid based on their choice and the choices of their group members in the selected game. Subjects received €100 plus the dollar value of their payoff in the selected game. Average payments for these selected participants were €174.

4 Experimental Results

Aggregate higher order rationality levels. Figure 1 reports the proportion of subjects classified as of level Rk , for each of the five main games, irrespective of the order of the tasks. More than 80% of the subjects were classified by a level of rationality between $R1$ and $R4$, depending on their choices in the different games.¹¹ The picture that emerges is quite clear. The classification of subjects in the different levels shows high variability across games. The e-ring games, the ring games and the 1/3-BC game have level $R1$ as their mode, whereas the 4×4 games have level $R2$, and the 2/3-BC game has level $R4$ as mode. The frequencies of Rk levels tend to decrease after $R1$ or $R2$ for the e-ring games, the 1/3-BC and the 4×4 games. In the 2/3-BC game, we find that the distribution is generally shifted towards higher levels, in particular, with high frequencies of $R2$'s, $R3$'s and $R4$'s.¹² Finally, for the ring games, we observe a steep decrease in the frequencies of subjects classified as $R2$ and $R3$ while there is an increase increase in the frequencies of $R4$'s. This may be evidence of the presence of the inductive step in the ring game, as discussed in the introduction.¹³

¹¹We leave out the p -BC game with unspecified p because of the different identification strategy used.

¹²Notice that in the 2/3-version of the BC game, numbers below 30 and 20 are already classified as, respectively, $R3$ and $R4$. When looking at the reasoning processes reported in p -BC game with unspecified p , we observe that many of the subjects reporting such low numbers, do it for idiosyncratic and nonstrategic reasons (e.g., lucky number, birthdate, age, etc.). By contrast, in the 1/3-BC game, subjects need to choose numbers below 4 and 1.2, to be classified as $R3$ and $R4$, respectively.

¹³Similarly to Kneeland (2015), we repeat subjects' classification in the ring and e-ring games allowing for one mistake. While, logically, the Rk levels tend to increase for both games, the difference in the numbers of subjects classified as $R3$ vs. $R4$ increases only for the ring game, consistent with the inductive step observation.

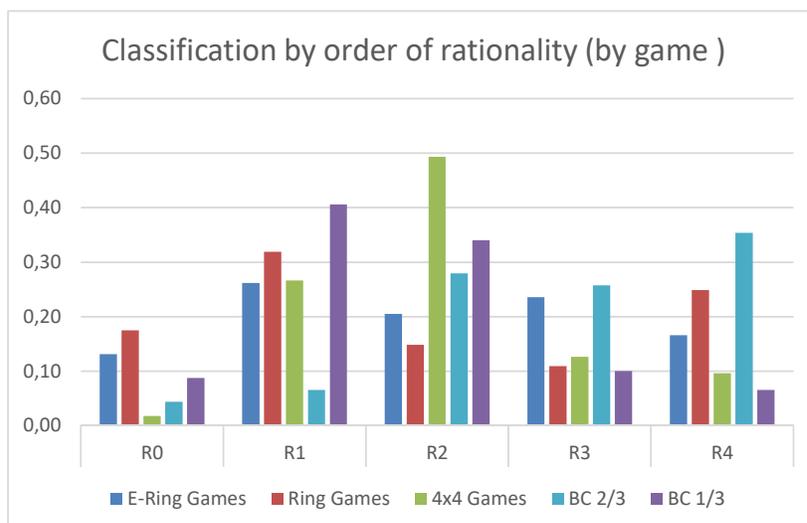


Figure 1: Classification by order of rationality, by game (ALL)

Reinforcing the latter observation, we also find indirect evidence for the framing effect. When comparing treatments in which the ring games and the e-ring games were presented in different orders to subjects, we find higher levels of rationality in the e-ring games when they are played after having played the ring games, than when played in the opposite order (Kolmogorov-Smirnov test significant at the 1% level). Moreover, when comparing treatments with permuted and non-permuted versions of the ring and e-ring games, we find higher levels of rationality in permuted versions (Kolmogorov-Smirnov test significant at the 1% level for the e-ring games and 2% for the ring games). We believe that the non-permuted versions may lead to more mechanical processes and rules of thumb, while the permuted versions may induce subjects to think harder about the games. This is in line with the literature on fluency.¹⁴

Finally, we find some evidence for cognitive depletion, namely, lower levels of rationality in the ring games when they are played after having played the e-ring games (Kolmogorov-Smirnov test significant at the 1% level). This could be due to the higher complexity of the e-ring game compared to the other games, as proven by the fact that 76% of the subjects (174 out of 229) passed the 10-question comprehension test, whereas, in the ring and the 4×4 games, respectively, 95% (218 out of 229) and 92% (211 out of 229) of the subjects passed the corresponding comprehension

¹⁴For a survey see Oppenheimer (2008). This research shows that individuals put in a disfluent situation, i.e. more difficult to process, tend not only to be more coherent and less prone to errors and biases. For example, Alter, Oppenheimer, Epley and Eyre (2007) show that individuals choosing in a fluent environment tend to do worst in the Cognitive Reflection Test and in solving logical syllogisms than individuals in a disfluent situation. In another study, Hernandez and Preston (2013) show that putting people in a disfluent environment disrupts the confirmation bias. Another example comes from the study conducted by Costa, Foucart, Arnon, Aparici and Apesteguia (2014). They show that students put in a fluent environment in which assignments are written in their mother tongue tend to make more biased decisions than children put in a disfluent one where assignments are written in a foreign language.

test. As can be seen from Figure 2 the distribution of the R_k levels, conditional on passing the test, is not visibly different from the unconditional case.¹⁵

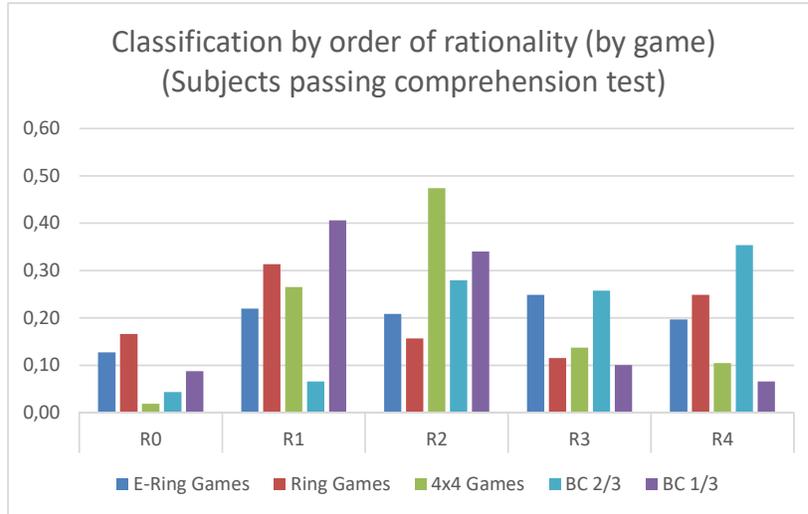


Figure 2: Classification by order of rationality, by game (COMP)

Individual higher order rationality levels. Similarly to what we find at the aggregate level, we also find a high degree of variation in the classification of individuals across games at the individual level. Out of 229 subjects, *no one* was classified at the same level of rationality across all of the five main games. When allowing individuals to be classified within two adjacent levels of rationality, we obtain that 14.4% of the subjects are within two levels, distributed as follows:

$$R0-R1: 2\% \quad R1-R2: 7\% \quad R2-R3: 4\% \quad R3-R4: 1\%$$

If we further classify individuals by the lowest level of rationality a subject has been identified with then we obtain the following distribution:

$$R0 : 32\% \quad R1 : 49\% \quad R2 : 18\% \quad R3 : 1\% \quad R4 : 0\%$$

Notice that no game outperforms the others in terms of identification of an individual lower bound of rationality, i.e., no game assigns a level of rationality to subjects that is consistently lower than the ones assigned by the other games.¹⁶

¹⁵Another treatment effect we find is that when the 1/3 version of the BC game is played after the 2/3 version, rationality levels are on average lower (Kolmogorov-Smirnov test significant at the 1% level). This effect might be due to the fact that subjects might use the numbers they said in the 2/3 version as a reference.

¹⁶Without taking into account the individuals identified as level 0, the e-ring games identify a lower bound for

Table 2 shows that the correlation of the Rk levels between pairs of games is also weak. Between games that are “more similar” (e.g., between the two BC games or between the ring and the e-ring games) it is clearly higher. Moreover, the e-ring games seem to perform slightly better than the others in that it has higher correlations with all other games.¹⁷

	E-ring	Ring	4×4	2/3-BC	1/3-BC
E-ring	1.00	0.24	0.14	0.13	0.15
Ring		1.00	0.13	0.09	0.10
4×4			1.00	0.02	0.09
2/3-BC				1.00	0.67
1/3-BC					1.00

Table 2: Correlation of the classifications by order of rationality between games

An alternative way of analyzing consistency in the data is to check the stability of the relative ranking of rationality levels across games for pairs of individuals. While the levels of rationality vary a lot across games, it might be the case that when we look at pairs of individuals, one is always ranked equal or higher across all games. In this sense we find that among all possible pairs of subjects only 29% are classified with a consistent relative ranking across all games. This number raises to 30% if we exclude beauty contest games and to 49% if we exclude e-ring, ring and 4×4 games.¹⁸

To get a clearer picture of the results at the individual level, we now allow for a weaker notion of belief in higher order rationality.

Individual higher order rationality levels with λ -beliefs. Consider subjects who are rational (at least $R1$) and assume they have probabilistic beliefs about others’ higher order rationality. Define λ_1 as the largest probability such an individual assigns to the others’ being rational, consistent with his choices, and call him λ_1 -rational. Next, define λ_2 as the largest probability a λ_1 -rational individual assigns to the others being λ_1 -rational, and call him $\lambda_2(\lambda_1)$ -rational. Finally, define λ_3

¹⁷When considering the p -BC game with unspecified p the correlations are as follows:

$$\text{E-ring: } 0.29 \quad \text{Ring: } 0.14 \quad 4 \times 4: 0.01 \quad 2/3\text{-BC: } 0.37 \quad 1/3\text{-BC: } 0.53.$$

¹⁸In the latter category of games, if we also include the version of the beauty contest with abstract p , the level of consistency goes down to 38%.

as the largest probability a λ_2 -rational individual assigns to the others being $\lambda_2(\lambda_1)$ -rational, call him $\lambda_3(\lambda_2)$ -rational. We estimated these λ 's for all subjects and across all games.¹⁹ The graph on the left in Figure 3 represents the distribution of the average individual λ_1 's in the population. The graph on the right shows the distribution of the individual coefficients of variation (CV) of the λ_1 's.

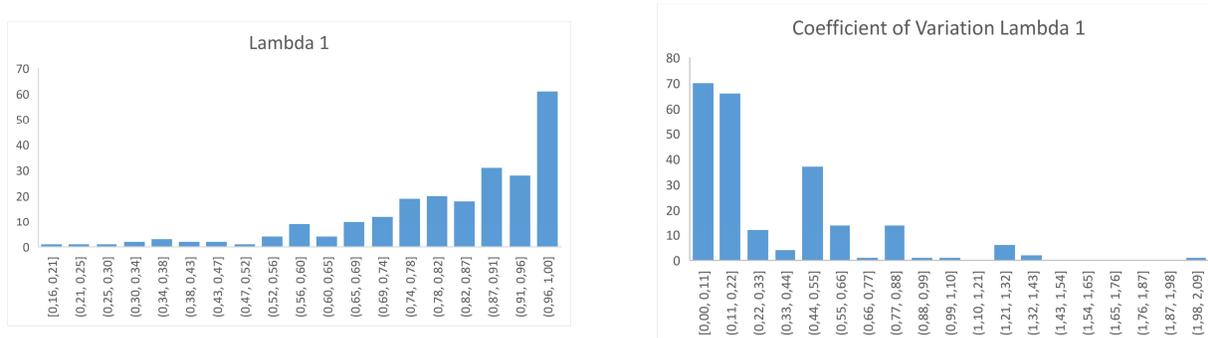


Figure 3: Distributions of average individual λ_1 's (left) and of corresponding coefficients of variation across all games (right), (ALL)

While only 19% of individuals are at least $R2$ across all games, 77% have an average $\lambda_1 \geq 0.75$ (λ_1 has mean 0.83 and median 0.88). Moreover, the graph on the right shows that the coefficients of variation are very low. In particular, 59% of the population has a CV between 0 and 0.22 (CV for λ_1 has mean 0.31 and median 0.19), suggesting that the λ_1 's are fairly stable across games. By contrast, even if they cannot be compared across individuals, given how they are defined, it is interesting to notice that λ_2 and λ_3 are lower and increasingly less stable (λ_2 has mean 0.65 and median 0.65, CV for λ_2 has mean 0.58 and median 0.55; λ_3 has mean 0.46 and median 0.45, CV for λ_3 has mean 0.80 and median 0.82).

5 Concluding Remarks

We presented a new class of games that by combining the communication structure of the email game (Rubinstein, 1989) with the circular structure of the ring games (Kneeland, 2015) presents

¹⁹See Germano, Weinstein and Zuazo-Garin (2018) or Definitions 7 and 8 in the appendix for detailed formalizations λ -rationality of different order. For example, in the game

	a	b
A	3,0	0,1
B	0,0	3,1

let p be probability of playing b , then, for a player choosing A , $\lambda_1 = 0.5$, which is the maximal p that solves $3(1 - p) \geq 3p$. Subsequent λ 's can be computed, using the same logic, as a function of the previous ones.

the simplest possible structure to avoid framing subjects into hierarchical thinking, hence allowing a cleaner identification of rationality levels from observed patterns of behavior.

A first potential concern is that our game might be too difficult to be implemented in the lab. In this sense, the experiment achieved two main things: (i) showing that the game is experimentally feasible and, for the first time, (ii) comparing the different methods used in the literature to identify higher-order rationality.

The experiment shows that the different methods proposed in the literature, including our own, do not identify individual higher-order rationality levels in a consistent manner. This casts doubts on using the standard concepts of rationality and higher order rationality as fixed behavioral benchmark in games and points toward taking a more game dependent approach.

Finally, it is worth highlighting that when allowing for weaker notions of belief in rationality, subjects behave as if they assign high probabilities to others being rational, while falling short of being identified as being $R2$. This could provide a robust starting point for applications such as mechanism design.

References

- Alter, Adam L, Daniel M Oppenheimer, Nicholas Epley and Rebecca N Eyre (2007). Overcoming intuition: Metacognitive difficulty activates analytic reasoning. *Journal of Experimental Psychology: General*, **136**(4), 569.
- Beard, T. Randolph and Richard Beil (1994). Do people rely on the self-interested maximization of others? *Management Science*, **40**, 252–262.
- Brandenburger, Adam, Alex Danieli and Amanda Friendenberg (2017). How many levels do players reason? an observational challenge and solution. *Mimeo*.
- Burchardi, Konrad B. and Stefan P. Penczynski (2014). Out of your mind: Eliciting individual reasoning in one shot games. *Games and Economic Behavior*, **84**, 39–57.
- Costa, Albert, Alice Foucart, Inbal Arnon, Melina Aparici and Jose Apesteguia (2014). “piensa” twice: On the foreign language effect in decision making. *Cognition*, **130**(2), 236–254.
- Costa-Gomes, Miguel, Vincent P. Crawford and Bruno Broseta (2001). Cognition and behavior in normal-form games: An experimental study. *Econometrica*, **69**(5), 1193–1235.
- Costa-Gomes, Miguel, Vincent P. Crawford and Nagore Iriberri (2013). Structural models of nonequilibrium strategic thinking: Theory, evidence, and applications. *Journal of Economic Literature*, **51**, 5–62.
- Costa-Gomes, Miguel and Georg Weizsacker (2008). Stated beliefs and play in normal form games. *Review of Economic Studies*, **75**, 729–762.
- Cubitt, Robin and Robert Sugden (1994). Rationally justifiable play and the theory of non-cooperative games. *Economic Journal*, **104**(425), 798–803.
- Dekel, Eddie, Drew Fudenberg and Stephen Morris (2007). Interim correlated rationalizability. *Theoretical Economics*, **2**, 15–40.
- Georganas, Sotiris, Paul J Healy and Roberto A Weber (2015). On the persistence of strategic sophistication. *Journal of Economic Theory*, **159**(PA), 369–400.
- Germano, Fabrizio, Jonathan Weinstein and Peio Zuazo-Garin (2018). Uncertain rationality, depth of reasoning and robustness in games with incomplete information. *Mimeo*.
- Healy, Paul J (2011). Epistemic foundations for the failure of nash equilibrium. *Working Paper*.
- Hernandez, Ivan and Jesse Lee Preston (2013). Disfluency disrupts the confirmation bias. *Journal of Experimental Social Psychology*, **49**(1), 178–182.

- Kneeland, Terri (2015). Identifying higher-order rationality. *Econometrica*, **83**(5), 2065–2079.
- Lim, Wooyoung and Siyang Xiong (2016). On identifying higher order rationality. *Mimeo*.
- Nagel, Rosemarie (1995). Unraveling in guessing games: An experimental study. *American Economic Review*, **85**(5), 1313–26.
- Oppenheimer, Daniel M (2008). The secret life of fluency. *Trends in Cognitive Sciences*, **12**(6), 237–241.
- Rey-Biel, Pedro (2009). Equilibrium play and best response to (stated) beliefs in normal form games. *Games and Economic Behavior*, **65**(2), 572–585.
- Rubinstein, Ariel (1989). The electronic mail game: Strategic behavior under “almost common knowledge.”. *American Economic Review*, **79**(3), 385–91.
- Schotter, Andrew, Keith Weigelt and Charles Wilson (1994). A laboratory investigation of multi-person rationality and presentation effects. *Games and Economic Behavior*, **6**, 445–468.
- Tan, Tommy C.C. and Sergio R.C. Werlang (1988). The bayesian foundations of solution concepts of games. *Journal of Economic Theory*, **45**, 370–391.
- Van Huyck, John, John Wildenthal and Ray Battalio (2002). Tacit coordination games, strategic uncertainty, and coordination failure: Evidence from repeated dominance solvable games. *Games and Economic Behavior*, **38**, 156–175.