Rationality and Observed Behavior*

Francesco Cerigioni[1], Fabrizio Germano[1], Pedro Rey-Biel[2], and Peio Zuazo-Garin[3]

[1]Universitat Pompeu Fabra and Barcelona GSE
[2]Universitat Ramon Llull, ESADE
[3]Universitat Pompeu Fabra

September 20, 2018

Abstract

Despite its relevance for understanding behavior in strategic interactions, very little consensus has been reached on how to identify subjects' levels of higher order rationality. To address this, we propose a novel class of two-player games with incomplete information, the *e-ring games*, that builds on the ring games of Kneeland (2015) and the email game of Rubinstein (1989). We check within subject consistency of identified rationality levels across standard games and find that our games yield rationality levels exhibiting the highest correlation both with the most frequent classification at the individual level, and with an independent measure of cognitive ability.
(JEL C70, C72, C91, D01, D80)

Keywords: Rationality, Higher-order Rationality, Revealed Rationality.

1 Introduction

In any social interaction among boundedly rational agents, optimal behavior depends on the beliefs about whether the others are rational, about whether the others believe the others are rational and so on. Therefore, it is crucial to obtain a reliable method to identify the empirical distribution of individuals' levels of rationality and higher order rationality. For example, in order to understand price formation in financial markets one needs to know the

*Corresponding author: francesco.cerigioni@upf.edu. We acknowledge the financial support from grant ECO2017-89240-P (AEI/FEDER, UE), the Spanish Ministry of Economy and Competitiveness (ECO2015-63679-P and the Severo Ochoa Programme for Centres of Excellence in R&D (SEV-2015-0563)),the Programa Ramon y Cajal, Gobierno Vasco (POS-2016-2-0003 and IT568-13), the ERC Programme (ERC 579424), Universitat Pompeu Fabra, ESADE Universidad Ramon Llull and Barcelona GSE. We thank Doug Bernheim, Itzhak Gilboa, Yoram Halevy, Nagore Iriberri, Antonio Penta, Ariel Rubinstein, Marciano Siniscalchi and Charles Sprenger for very insightful comments that contributed to the development of this work.

empirical distribution of the levels of higher order of rationality among traders, given the strategic nature of the setting. In institutional design, whether a proposed mechanism is empirically efficient depends on the participants' depth of reasoning. Similar issues arise in a variety of strategic contexts such as negotiation and conflict, monetary policy, oligopolistic competition and voting.

In this paper, (a) we introduce a new class of two-player games of incomplete information, which we call *e-ring games*, where players receive private messages that determine their payoffs, while generating a one-to-one correspondence between actions and higher order rationality. This minimizes potential framing effects and allows us to reliably estimate the empirical distribution of individuals' higher order rationality levels in an experimental setting; (b) we use our class of games to find the distribution of rationality levels in subjects participating in an experiment and, for the first time, we compare it at the individual level with other standard methods used in the literature to identify higher orders of rationality.

The two main approaches for identifying higher-order beliefs 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, when abandoning the rational framework, through the estimation of level-k behavioral models using choice data and, instead of eliciting beliefs, relying on structural assumptions on first-order beliefs, that is, behavior of players of level 0 (Costa-Gomes, Crawford and Iriberri, 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 identification tool for rationality levels 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 without neither needing to elicit beliefs nor making structural assumptions about them. 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.[1]

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

[1]There is a 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 also uses a choice based inference but cannot make a subject play at each level of the hierarchy of beliefs.

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 to the n-th opponent (see solid arrows in figure 1). In ring games, strategic uncertainty is always resolved before the emergence of beliefs about others' beliefs concerning one own actions (see dashed arrow in figure 1). 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.[2]

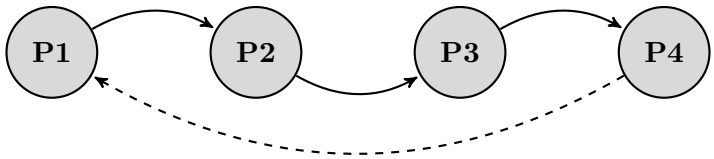

Figure 1: Ring games: Arrows denote payoff dependence. Dashed arrow denotes the presence of a dominant strategy.

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 (1989), generating a natural one-to-one correspondence between messages and higher-order beliefs. In our games, strategic uncertainty is not resolved before the emergence of beliefs about others' beliefs concerning one own actions (see figure 2). However, in order to identify subjects in a reliable manner, our games have interim payoff matrices that vary with the number of messages received.[3] This allows us to associate different payoff matrices and hence actions to different levels of higher-order beliefs.

[2]Kneeland (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."

[3]This 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.

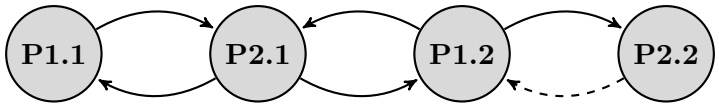

Figure 2: E-ring games: P$n.k$ means that player n received k messages.

More specifically, in the *e-ring games* each player's own payoffs depend not only on the actions chosen by every player, but also on the number of messages that player received. That is, a player's own payoffs when she has received k messages are different from those faced when receiving $k+1$ 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.

Importantly, we do not rely on the exclusion restriction assumption made in Kneeland (2015) for our identification of players' orders of rationality.[4] Instead, we rely on a *revealed rationality approach* that assigns players the maximal level of higher-order rationality consistent with the choices made as in Lim and Xiong (2016).[5] However, they propose a class of two-player games, the chain games, that still potentially suffer 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 than the ones in Lim and Xiong (2016), they substantially weaken the chances that agents are forced to think about higher levels in the hierarchy of beliefs by the game, and, moreover, they have an

[4]The 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.

[5]See also Tan and Werlang (1988) and Brandenburger, Danieli and Friendenberg (2017).

inductive step that should be less apparent, at least for subjects of orders of rationality less or equal to 4.

In our experiment all subjects play each of the following four types of games: eight of our e-ring games, eight ring games as in Kneeland (2015), two simple two-player 4×4 dominance solvable games and three different versions of the beauty contest game presented in Nagel (1995).[6] To be more specific, subjects play versions of the beauty contest game where the average of all subjects' responses is multiplied by 1/3, 2/3 and finally, in the *p-beauty game*, by an unspecified number (p) strictly between 0 and 1. Using the above mentioned revealed rationality approach, we classify subjects within each class of games into five levels. An action is categorized as R0 if it is never a best response, as R1 if it is a best response to some belief, as R2 if it a best response to the belief that the opponent is playing an R1 action, and so on. A subject is hence classified as Rk if all her actions are Rk actions and at least one is not R$k+1$. Table 1 shows the empirical distribution.[7]

	$R0$	$R1$	$R2$	$R3$	$R4$
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: Proportion of subjects classified by orders of rationality, by game

First, we find levels of higher-order rationality that are very game dependent at the individual and at the aggregate level. While this paper is the first, to the best of our knowledge, that experimentally checks for persistence of rationality classifications across games at the

[6] In 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.

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

individual level, similar results have been found in the k-level literature (Georganas, Healy and Weber, 2015; Alaoui and Penta, 2016; Cooper, Fatas, Morales and Qi, 2016). Given the many characteristics that change from one game to the other, it is not surprising to find that absolute levels of rationality vary across games. On the other hand, individual characteristics, such as cognitive abilities, remain more or less invariant across games, hence one may expect the relative ranking of individuals to exhibit more stability.[8] Nevertheless, we do not find evidence in favor of such conclusion. This might be due to the fact that our identification method, that is the revealed rationality approach, is too strict. For this reason, we use a different method that classifies individuals with the most frequent rationality level they have been identified with. Once we check for the correlation between the different games and this new classification, our game outperforms the others by a significant margin. This finding might lead us to the conclusion that *e-ring games* are successful in identifying the relative ranking of individuals once we take into account possible statistical noise. If that was the case, we should find evidence that our ranking is also more correlated than others with independent measures of cognitive ability. Indeed, we find that the ranking identified with *e-ring games* is the most correlated with the ranking of the subjects based on the results of the standardized test used for the admittance to the university.

The previous results support the theoretical advantages of *e-ring games* that are further validated by evidence of the inductive step and framing effects in *ring games*. In fact, our second finding is that only 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. A third finding relies on 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. Finally, it is interesting to notice that the classification obtained from the beauty contest varies a lot depending on the value of p, with the 2/3 beauty contest being the noisier.

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.

[8]There is also a growing literature distinguishing players cognitive bounds and actual behavior due to beliefs. For example, a subject's behavior might be consistent with low levels of rationality but this might be due to her beliefs about other's rationality more than actual limitations in her cognitive capacity. See Alaoui and Penta (2018b) and Friedenberg, Kets and Kneeland (2018) for a theoretical and empirical discussion of the problem.

2 E-Ring Games

We introduce a new class of two-player games of incomplete information which we refer to as *e-ring games*. Before defining the general class of games, we solve a simple example, used in the experiment, and show how it can identify higher-order rationality up to level 4.

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 $\Theta = \{(0,0),(0,1),(1,1)\}$ 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 they receive and their payoffs depend only on the number of messages they receive. Note also that player 2 either has the same number or one more message than player 1 and each player either gets 0 or 1 messages. To compute the payoffs of the opponent, players can compute the number of messages received by their opponent as follows. Player 1 with 0 messages knows her opponent has either 0 or 1 messages, each event with equal probability; player 1 with 1 message knows for sure the other player also has 1 message. Similarly, player 2 with 0 messages, knows for sure his opponent also has 0 messages; while player 2 with 1 message knows his opponent has either 0 or 1 messages, each event with equal probability.

Consider the following payoff matrices of the players, 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 she receives θ_1 messages, and $u_2(\theta_2)$ the payoffs of player 2 when he receives θ_2 messages.

$u_1(\theta_1 = 0) =$

	a	b	c	d
A	80	60	80	40
B	200	20	140	120
C	120	140	180	160
D	160	120	100	140

$u_2(\theta_2 = 0) =$

	a	b	c	d
A	80	160	180	200
B	40	140	80	100
C	60	100	140	20
D	80	120	100	160

$u_1(\theta_1 = 1) =$

	a	b	c	d
A	60	80	40	80
B	80	20	20	40
C	160	120	180	100
D	40	60	80	60

$u_2(\theta_2 = 1) =$

	a	b	c	d
A	180	20	100	200
B	120	40	140	20
C	160	80	200	180
D	140	60	160	80

The above payoff structure has a unique interim rationalizable action for all players and number of messages. Player 1 with 1 message (payoff matrix $u_1(\theta_1 = 1)$) has a strictly dominant action C. Player 2 with 1 message (payoff matrix $u_2(\theta_2 = 1)$), seeing this and the fact that player 1 with 0 messages has A as strictly dominated action, (and knowing that he faces player 1 with $\theta_1 = 0, \theta_1 = 1$ with equal probability), has a unique strict best reply c. Player 1 with 0 messages (payoff matrix $u_1(\theta_1 = 0)$), given the above and seeing that

player 2 with 0 messages has a as a strictly dominated action, (and again knowing that she faces player 2 with $\theta_2 = 0, \theta_2 = 1$ with equal probability), has a unique strict best reply C. Finally, player 2 with 0 messages (payoff matrix $u_2(\theta_2 = 0)$), knowing that for sure he faces player 1 with 0 messages and that she plays C as unique best reply, also has a unique strict best reply c. Thus $((C, C); (c, c))$ is the unique rationalizable strategy profile.

In the experiment each subject plays in all four different contingencies twice.[9] Thus, following the revealed rationality approach, a subject playing always C or c, depending on the role, would be identified as R4. On the other hand, a subject playing always C or c but playing either b or d when playing as player 2 with 0 messages would be identified as R3. Similarly, a subject playing C as player 1 with 1 message and c as player 2 with 1 message, while playing B, D as a player 1 or b, c, d as player 2 with 0 messages, would be identified as R2. Furthermore, a subject playing C as player 1 with 1 message while playing a or d as player 2 with 1 message and B, C, D or b, c, d as either player 1 or 2 with 0 messages, would be identified as R1. Finally, a subject playing any of the dominated actions would be identified as R0. □

Just like the email game of Rubinstein (1989), our game also provides a natural one-to-one correspondence between messages and higher-order beliefs that in turn allows us to associate actions to different levels of rationality in a context of asymmetric information. An important difference with the email game, however, is that our game has 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.[10] The addition of asymmetric information is crucial to resolve what Kneeland defines as a *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 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, over two states of the world.

[9] For the sake of brevity in this example we explain our identification strategy *as if* subjects switch roles. In the experiment we achieve this by reassigning player 1's matrix with 1 message to player 2 with 0 messages while reallocating the other matrices to maintain the dominant solvability structure.

[10] This 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.

General definition. The following is a general definition of an *e-ring game of depth m*, where m is the maximum number of messages received. Notice that to make the explanation clearer we let 1 be the minimum number of messages received in the general definition. This allows to test up to $2m$ levels of higher-order rationality with a game of depth m. The basic structure is such that a player of rationality level $1, 3, \ldots, 2m-1$, has to play action a as player 2 (column player) of type $t_2 = 1, 2, \ldots, m$, respectively, and similarly, a player of rationality level $2, 4, \ldots, 2m$ has to play action A as player 1 (row player) of type $t_1 = 1, 2, \ldots, m$, 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 (E-ring game) *An* e-ring game of depth m *is a Bayesian game* $\langle I, \Theta, (A_i, u_i, T_i, \pi_i)_{i \in I} \rangle$, *where: (i)* $I = \{1, 2\}$ *is the set of players, (ii)* $\Theta = \{(\theta_1, \theta_2) \in \{1, \ldots, m\}^2 \,|\, \theta_2 - \theta_1 \in \{0, 1\}\}$ *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)* $u_i : A_1 \times A_2 \times \Theta \to \mathbb{R}$ *are utility functions given by following matrices:*

$$u_1(\theta) = \begin{cases} \begin{array}{c|cccc} & a & b & c & d \\ \hline A & 2 & 1 & 2 & 2 \\ B & 0 & 0 & 0 & 0 \\ C & 1 & 3 & 0 & 1 \\ D & 1 & 2 & 1 & 0 \end{array} & \text{if } \theta_1 \text{ odd} \\[2ex] \begin{array}{c|cccc} & a & b & c & d \\ \hline A & 2 & 2 & 1 & 2 \\ B & 1 & 0 & 3 & 1 \\ C & 0 & 0 & 0 & 0 \\ D & 0 & 1 & 2 & 1 \end{array} & \text{if } \theta_1 \text{ even} \end{cases}$$

$$u_2(\theta) = \begin{cases} \begin{array}{c|cccc} & a & b & c & d \\ \hline A & 1 & 0 & 0 & 0 \\ B & 1 & 0 & 0 & 0 \\ C & 1 & 0 & 0 & 0 \\ D & 1 & 0 & 0 & 0 \end{array} & \text{if } \theta_2 = 1 \\[2ex] \begin{array}{c|cccc} & a & b & c & d \\ \hline A & 2 & 0 & 0 & 1 \\ B & 2 & 0 & 1 & 0 \\ C & 1 & 0 & 3 & 2 \\ D & 2 & 0 & 1 & 1 \end{array} & \text{if } \theta_2 \text{ even} \\[2ex] \begin{array}{c|cccc} & a & b & c & d \\ \hline A & 2 & 0 & 0 & 1 \\ B & 2 & 1 & 0 & 0 \\ C & 1 & 3 & 0 & 2 \\ D & 2 & 1 & 0 & 1 \end{array} & \text{if } \theta_2 \neq 1, \text{ odd} \end{cases}$$

where the numbers 0, 1, 2, 3 $\subset \mathbb{R}$ *denote equivalence classes from which arbitrary utilities can be chosen, subject to* $u \in 0$, $v \in 1$, $x \in 2$, *and* $y \in 3$, *always implies* $u < v < x < y$ *in any given matrix of a given player and for every action of the opponent, (v)* $T_i = \{1, \cdots, m\}$

are the sets of types, and (vi) $\pi_i : T_i \to \Delta(T_{-i} \times \Theta)$ *are belief maps given by:*

$$\pi_1(t_1)[(t_2, (\theta_2; \theta_1))] = \begin{cases} 1/2 & \text{if } t_1 = \theta_1 < m \text{ and } t_2 = \theta_2 = \theta_1, \\ 1/2 & \text{if } t_1 = \theta_1 < m \text{ and } t_2 = \theta_2 = \theta_1 + 1, \\ 1 & \text{if } t_1 = \theta_1 = m \text{ and } t_2 = \theta_2 = \theta_1, \\ 0 & \text{otherwise,} \end{cases}$$

and:

$$\pi_2(t_2)[(t_1, (\theta_1; \theta_2))] = \begin{cases} 1/2 & \text{if } t_2 = \theta_2 < m \text{ and } t_1 = \theta_1 = \theta_2 - 1, \\ 1/2 & \text{if } t_2 = \theta_2 < m \text{ and } t_1 = \theta_1 = \theta_2, \\ 1 & \text{if } t_2 = \theta_2 = m \text{ and } t_1 = \theta_1 = \theta_2, \\ 0 & \text{otherwise,} \end{cases}$$

The information structure of the game 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 action to play A or a for all types of player 1 and 2 respectively.[11]

3 Experimental Design

The experiment consisted of four tasks and a non-incentivized questionnaire. In the first task, subjects chose an action in a pair of two player 4×4 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, respectively. 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

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

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 remained constant (non-permuted) across decisions, while either varying the player's position (ring game) or the number of messages received (e-ring game), or whether the actions in such matrices were reshuffled (permuted); and (*iii*) whether the 1/3 version of the beauty contest game was played before or after the 2/3 one. 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 games with benchmark games used in the literature (ring games, dominance solvable games such as our 4×4 games and the p-beauty game) to empirically classify individuals according to the revealed rationality approach.

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 since there is 1.2% probability of being identified as R1-R4 while playing randomly in either games.[12]

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

[12]Of the 801 possible rational action profiles, 720 would be identified as R1 (89.8%), 72 as R2 (8.9%), 8 as R3 (0.9%) and 1 as R4 (0.1%).

each of the tasks. A majority of subjects (71%) passed all these tests. Their explicitly written rationale to their actions also shows that they understood the experimental instructions. Participants received no feedback after playing each of the games nor after finishing each of the tasks and we monitored that subjects would not jump from one task to the other unless instructed. 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 tasks and 9 minutes for the final task. The two experimental sessions lasted around 75 minutes each.

We provided high monetary incentives for 10 randomly selected participants, instead of paying all subjects a lower amount of money.[13] One of the twenty decisions 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 euro value of their payoff in the selected game. Average payments for these selected participants were €174.

4 Experimental Results

Aggregate inconsistency. Figure 3 reports the proportion of subjects classified as level Rk, for all but the p-beauty contest, irrespective of the order of the tasks.[14] More than 80% of the subjects were classified by a level of rationality between $R1$ and $R4$. 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.[15] Finally, for the ring games, we

[13] See Alaoui and Penta (2018a) for a theoretical justification of this design choice that should give higher incentives to achieve higher levels in the hierarchy of beliefs.

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

[15] Notice 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 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.

observe a steep decrease in the frequencies of subjects classified as $R2$ and $R3$ while there is an 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.

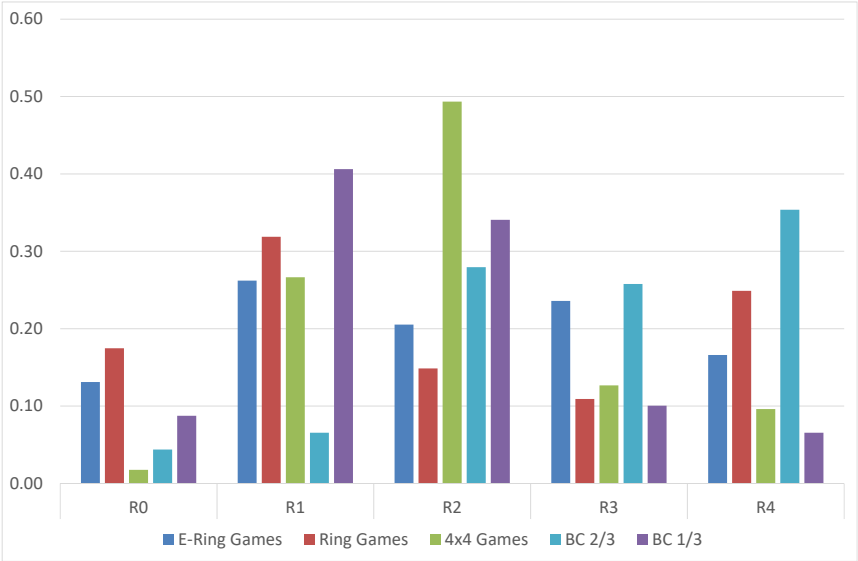

Figure 3: Classification by order of rationality, by game (`ALL SUBJECTS`)

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). It may be due to the non-permuted versions leading 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 (Oppenheimer (2008)).

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 test. The distribution of the Rk levels, conditional

on passing the test, is not qualitatively different from the unconditional case.[16]

Individual inconsistency. 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 games. When allowing individuals to be classified within two adjacent levels of rationality, we obtain that 14% of the subjects are within two levels, distributed as follows:

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

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

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

No class of games is more stringent than the others in terms of identification of an individual lower bound of rationality, that is, no class of games assigns a level of rationality to subjects that is consistently lower than the ones assigned by the other classes. Without taking into account the individuals identified as level 0, the e-ring games identify a lower bound for 26% of the population, the ring games and the 4×4 games for 30%, the 2/3-BC game for 5% and the 1/3-BC game for 45%.

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 classes of games, it might be the case that when we look at pairs of individuals, one is always ranked equal or higher across all classes. In this sense we find that among all possible pairs of subjects only 29% are classified with a consistent relative ranking across all classes of 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.[17]

An additional method to measure the stability of the relative ranking across classes of games is to check the correlation between the distributions obtained in the different classes. Table 2 shows that the correlation of the Rk levels between pairs of classes of games is also weak. Between classes of 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. Interestingly, the e-ring games perform slightly better than the others in that it has higher correlations with all other

[16] Another 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.

[17] In 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%.

games.[18]

	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 distributions of levels of rationality between classes of games.

Individual consistency. The fact that the e-ring games are more strictly related to other games than the other way around, suggests that this class of games is actually capturing some kind of underlying relative ranking of the individuals. To look for the existence of this relation, one would need to check the correlation of the classifications obtained in the different classes of games with a distribution clean from statistical noise, e.g. mistakes, idiosyncratic game characteristics, etc. A robust way of identifying this distribution, within the revealed rationality approach, is to classify each individual in the rationality level she has been most frequently identified with across all classes of games.[19] Once we do that the obtained distribution is the following:

$$R0 : 9.1\% \quad R1 : 37.0\% \quad R2 : 34.6\% \quad R3 : 6.7\% \quad R4 : 12.5\%$$

Using this distribution we calculate the aforementioned correlations. For the beauty contest, for each individual, we take the minimum level of rationality she has been identified with to avoid the noise created by the 2/3 version. The results are as follows:

$$\text{E-ring: } 0.63 \quad \text{Ring: } 0.50 \quad 4 \times 4\text{: } 0.35 \quad \text{BC: } 0.46$$

[18] When considering the p-BC game with unspecified p the correlations are as follows:

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

[19] When for an individual the most frequent classification is not unique, following the revealed rationality approach, we take the minimum. In the case that individual behavior is particularly noisy, that is an individual has been classified differently in each class of games, we do not include the data in the calculation of the correlations even if they do not qualitatively change the results. This happens 21 times out of 229.

The e-ring games outperform the other classes of games by a significant margin. This leads us to conclude that *e-ring games* succeed in identifying the relative ranking of individuals once we take into account possible statistical noise. Another piece of evidence pointing towards this direction is the higher correlation between the distribution of rationality levels in the e-ring games with an independent measure of cognitive ability than in any of the other classes of games. Indeed, the correlation between *e-ring games* and the ranking of the subjects based on the results of the standardized test used for the admittance to the university is 0.24, while the same correlation for the ring games is 0.12, for the 4×4 games is 0.06 and finally for the beauty contests is 0.05.

5 Concluding Remarks

We presented a new class of games that allows the identification of rationality levels from observed patterns of behavior. Combining the communication mechanism of the email game (Rubinstein, 1989) with the circular nature of the ring games (Kneeland, 2015) we obtain the simplest theoretical structure to avoid framing subjects into hierarchical thinking.

The experiment shows that aggregate and individual consistency of rationality levels across games is very weak. 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. Nevertheless, our class of games outperforms the others when looking for the relative ranking of individuals in the population.

This could provide a robust way of finding the ranking of individuals in settings in which this is crucial like auctions, price formation in financial markets, oligopolistic competition and so on.

References

Alaoui, Larbi and Antonio Penta (2016). Endogenous depth of reasoning. *The Review of Economic Studies*, **83**(4), 1297–1333.

Alaoui, Larbi and Antonio Penta (2018a). Cost-benefit analysis in reasoning. *Working Paper*.

Alaoui, Larbi and Antonio Penta (2018b). Reasoning about others' reasoning. *Working Paper*.

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.

Cooper, David J, Enrique Fatas, Antonio J Morales and Shi Qi (2016). Consistent depth of reasoning in level-k models. *Working Paper*.

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.

Friedenberg, Amanda, Willemien Kets and Terri Kneeland (2018). Bounded reasoning: Cognition or rationality? *Working Paper*.

Georganas, Sotiris, Paul J Healy and Roberto A Weber (2015). On the persistence of strategic sophistication. *Journal of Economic Theory*, **159**(PA), 369–400.

Healy, Paul J (2011). Epistemic foundations for the failure of nash equilibrium. *Working Paper*.

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 multiperson 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 uncertanty, and coordination failure: Evidence from repeated dominance solvable games. *Games and Economic Behavior*, **38**, 156–175.