Higher Orders of Rationality and the Structure of Games*

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

[1]Universitat Pompeu Fabra
[2]Barcelona Graduate School of Economics
[3]Universitat Ramon Llull, ESADE

August 10, 2019

Abstract

Identifying individual levels of rationality is crucial to modeling strategic interaction and understanding behavior in games. Nevertheless, there is no consensus on how to best identify levels of higher order rationality, and the identification of an empirical distribution remains highly elusive. In particular, the games used for the task can have a huge impact on the identified distribution. To tackle this fundamental problem, this paper introduces an axiomatic approach that singles out a simple class of games that minimizes the probability of misidentification errors. It then shows that the axioms are empirically meaningful in a within subject experiment that compares the distribution of orders of rationality across different games, including standard games from the literature. The games singled out by the axioms exhibit the highest correlation both with the distribution of the most frequent rationality level a subject has been classified with and with an independent measure of cognitive ability. Finally, there is no evidence in our sample of within subject consistency of identified rationality levels across games.

(JEL C70, C72, C91, D01, D80)

Keywords: Rationality, Higher-Order Rationality, Revealed Rationality, Levels of Thinking.

*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)), Programa Ramon y Cajal, Gobierno Vasco (POS-2016-2-0003 and IT568-13), the ERC Programme (ERC 579424), Caixa Bank, Universitat Pompeu Fabra, ESADE Universidad Ramon Llull and Barcelona GSE. We are particularly grateful to UC3M and the team of FET Open IBSEN (H2020 RIA 662725) for their help in recruiting subjects and in running some experimental sessions. We thank Doug Bernheim, Simone Galperti, Itzhak Gilboa, Yoram Halevy, Nagore Iriberri, Rosemarie Nagel, Antonio Penta, Ariel Rubinstein, Marciano Siniscalchi and Charles Sprenger for very insightful comments that contributed to the development of this work.

1 Introduction

In any interaction among 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. If agents are bounded in their reasoning process about others' rationality, it becomes crucial to obtain a reliable method to identify the empirical distribution of individuals' orders of rationality. For example, to understand price formation in financial markets one needs to know the empirical distribution of the levels of higher order rationality among traders. In institutional design, whether a proposed school-matching procedure is empirically efficient depends on the participants' depth of reasoning. Similar issues arise in a variety of strategic contexts, such as monetary policy, negotiation and conflict, oligopolistic competition, and voting.

This crucial empirical problem has given rise to an extended literature that uses different identification methods to find the empirical distribution of individual levels of hierarchical thinking (Beard and Beil, 1994; Schotter, Weigelt and Wilson, 1994; Nagel, 1995; Costa-Gomes, Crawford and Broseta, 2001; Van Huyck, Wildenthal and Battalio, 2002; Costa-Gomes and Weizsacker, 2008; Rey-Biel, 2009; Healy, 2011; Costa-Gomes, Crawford and Iriberri, 2013; Burchardi and Penczynski, 2014). While there is no consensus on which method is best, there is agreement on the fact that higher orders of rationality are difficult to identify. The problem, however, might be deeper than the choice of the best identification method. In fact, the fundamental issue is that standard games do not allow for the observation of behavior at the different levels of the hierarchy of beliefs and, if they do, they might induce hierarchical thinking due to their structure.

This conflict is easily visible in dominance solvable games. When a subject plays a strategy consistent with having beliefs of a certain level k there is a high probability of making an identification mistake. Given it is not possible to observe behavior at the different steps of the reasoning process, we cannot exclude that it is other belief systems or decision processes which lead to the observed choice. On the other hand, if the structure of the game allows for the observation of behavior at the different steps of the hierarchy of beliefs, the subject would be classified as having beliefs of level k only if her behavior at each step of the ladder is consistent with such a classification, keeping other things constant.[1] This would consistently reduce the probability of identification mistakes but such games might frame subjects into thinking hierarchically. Due to framing, subjects might be classified into higher categories hence making the identified empirical distribution flawed. To the best of our knowledge, there is still no solution to this fundamental problem.

This paper provides an answer to the above concerns. The contribution is theoretical and

[1]Another possibility would be to make a subject play a series of dominance solvable games but given each game is different from the others, there are too many factors out of the control of the researcher that might determine changes in individual behavior.

empirical. Theoretically, we use for the first time an axiomatic approach to define a new class of games. We propose two properties that reduce potential misidentification of higher order rationality levels and allow to reliably estimate their distribution in experimental settings. Empirically, we test the validity of the proposed properties and then the existence of an underlying distribution of higher orders of rationality.

Regarding the theoretical contribution, we propose the following two properties that address the concerns raised. First, the game should allow for reliable, choice based, inference. This can be achieved via a structure that allows for the test of individual behavior at *each step* of the hierarchy of beliefs. We refer to this property of games as *lower order consistency*. It was first incorporated in Kneeland (2015) which used *ring(-network) games*, introduced by Cubitt and Sugden (1994).[2] Second, the payoff dependencies of the game should not correspond exactly with the hierarchical belief "structure," otherwise it might lead subjects who would not form higher-order beliefs to play as if they had them. In particular, it should not frame players with low levels of rationality to behave as if they had higher ones. Hence, the payoff structure of the game should be such that each level of the hierarchy of beliefs has payoff interdependencies not just with lower levels. This way players can form different hierarchies of beliefs. We refer to games satisfying this property as *framing-free*.

Somewhat surprisingly, these two natural properties are sufficient to pin down a unique class of games. Indeed we show that the simplest class of games satisfying lower order consistency and absence of framing, and identifying four levels of rationality—the empirically relevant ones—is a specification of a new class of games we present here, the *e-ring games*. An *e-ring game* is a normal-form game with incomplete information, where the latter is structured by means of messages that automatically 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.

Regarding the empirical contribution, we study the validity of e-ring games as an identification tool. We first test the properties we propose. Then, we compare the distribution identified by the e-ring games against the distributions obtained with standard games used in the literature. We do this to verify the existence of an underlying distribution of types across games and to see which of the games achieves the closest identification. 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).[3]

[2]The *ring games* used in Kneeland (2015) are finite dominance solvable games in normal form, where player 1's payoffs depend on player 1's and player 2's actions; player 2's payoffs depend on player 2's and player 3's actions and so on, until player k, whose payoffs depend on player k's and player 1's actions, but who has a single strictly dominant action that allows to initiate dominance solvability procedure.

[3]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 contest game*, by an unspecified number (p) strictly between 0 and 1 and assumed to be commonly known.

We use the *revealed rationality approach* to classify subjects within each class of games into five levels depending on the actions they choose. 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$k+1$. That is, we assign players the maximal level of higher-order rationality consistent with the choices made (as in Lim and Xiong (2016), Tan and Werlang (1988) and Brandenburger, Danieli and Friendenberg (2017)). We do not use the exclusion restriction assumption made in Kneeland (2015), which maintains that *subjects satisfying lower-order rationality do not respond to changes in higher-order payoffs*, even if the e-ring games would allow for such an approach. The reason is twofold. First, the 4×4 games and the beauty contest games do not allow the use of such a restriction, hence rendering the comparison across games difficult to interpret. Second, there exist influential theoretical frameworks in which subjects satisfying lower-order rationality do respond to changes in higher-order payoffs.[4]

The experiment supports our theoretical approach. First, we find evidence that the properties proposed are relevant. In particular, we find that games that violate lower order consistency, and hence do not have a one to one correspondence with the beliefs structure, tend to misidentify the distribution of types for levels 2 or higher. Second, we find framing effects in the *ring games*, which are the only other class of games that satisfy lower order consistency, but are not framing free. In fact, we find evidence that subjects that do not behave as if they had higher orders of rationality in other games, are classified as if they had them in ring games.

We next look at the existence of an underlying distribution of types in the population and we study which class of games, if any, seems to better identify such distribution. We find no clear evidence of the existence of a stable distribution. The results show that our games are more reliable than the others for the identification of the distribution of higher order of rationality once we take into account possible mistakes within the revealed rationality approach. Specifically, once we classify individuals with the most frequent rationality level they have been identified with and we check for the correlation between the different games and this classification, our game outperforms the others by a significant margin. This finding might suggest that e-ring games successfully identify 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.

Furthermore, the data show that the depth of higher-order rationality is very game dependent at the individual and at the aggregate level. While this paper is the first, to the best of

[4]See Section 4.3 for further details.

our knowledge, that experimentally checks for persistence of rationality classifications across games at the 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; Ellingsen, Östling and Wengström, 2018). Given that many characteristics change from one game to the other, it is not surprising to find that absolute levels of rationality vary across games.[5] Nevertheless, this evidence seems to suggest that assuming the existence of a stable absolute ranking of individuals in the population lacks empirical support.

The remainder of the paper is organized as follows. In the next section, we describe our class of games. In Section 3, we present the desirable properties a class of games should have to reliably identify higher orders of rationality and then we show that e-ring games are the only class of games that satisfy such properties. In Section 4, we present the experimental design along with the other classes of games we used and the experimental results. Section 5 concludes. The Online Appendix contains an English translation of the experimental instructions and the payoff matrices for all games used in the experiment.

2 E-Ring Games

An *e-ring game* is a two player incomplete information game in normal form in which players automatically send and receive messages and each player's own payoffs depend not only on the actions chosen by the players but also on the number of messages that player received. That is, a player's own payoffs when she has received ℓ messages are different from those faced when receiving $\ell + 1$ messages. There is a maximal number of messages that any player can receive with an otherwise email game-like communication structure (Rubinstein, 1989). The payoff of a sender with ℓ messages depends on the actions of a receiver with ℓ or $\ell + 1$ messages, whose payoffs in turn depend on actions of a sender with $\ell - 1$ or ℓ messages and ℓ or $\ell + 1$ messages respectively, and so on. This allows us to associate different payoff matrices and hence actions to different levels of higher-order beliefs. 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. As will be clear in Section 3, ℓ messages are needed for each of the two players to test for up to 2ℓ levels of higher-order rationality. This is a major difference with the email game of Rubinstein (1989), where players face the same 2×2 payoff matrices for almost all the number of messages received. The next definition formalizes these features of the e-ring game.

Definition 1 (E-Ring Game) *An* e-ring game of depth k *(even) is a list* $\mathcal{G} = \langle T_i, A_i, u_i, \pi_i \rangle_{i=1,2}$, *where, for each player* i*:*

[5]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), Friedenberg, Kets and Kneeland (2018) and Germano, Weinstein and Zuazo-Garin (2019) for theoretical and empirical discussions.

1. $T_i = \{1, 2, \ldots, k/2\}$ *is a set of types.*

2. A_i *is a finite set of actions.*

3. $u_i : T_i \times A_1 \times A_2 \to \mathbb{R}$ *is a payoff function.*

4. $\pi_i : T_i \to \Delta(T_{-i})$ *is a belief-map such that, for fixed* $p_1, p_2 \in (0, 1)$,

$$\pi_1(t_1)[t_2] = \begin{cases} p_1 & \text{if } t_2 = t_1 \\ 1 - p_1 & \text{if } t_2 = t_1 + 1 \end{cases} \qquad \pi_2(t_2)[t_1] = \begin{cases} p_2 & \text{if } t_1 = t_2 - 1 \\ 1 - p_2 & \text{if } t_1 = t_2 \end{cases}$$

for $1 \leq t_1 < k/2$ *and* $1 < t_2 \leq k/2$, *and otherwise* $\pi_1(k/2)[k/2] = 1$ *and* $\pi_2(1)[1] = 1$.

To see that this borrows from the communication structure presented in Rubinstein (1989), consider player i who has received ℓ messages. Then, by Definition 1, we say that such a player is of type $t_i = \ell$. Player i knows that the payoff she obtains from each action profile is given by the map $u_i(t_i, \cdot)$. However, player i is uncertain about the number of messages received by the other player and, therefore, also about the latter's type and payoff function. In particular, player 1 knows with probability p_1 that player 2 is of type $t_2 = \ell$, and with probability $1 - p_1$, that she is of type $t_2 = \ell + 1$ (with the exception of type $t_1 = k/2$, who knows that player 2 is of type $k/2$ as well); similarly, player 2 knows with probability p_2 that player 1 is of type $t_1 = \ell - 1$, and with probability $1 - p_2$, that she is of type $t_1 = \ell$ (with the exception of type $t_2 = 1$, who knows that player 1 is of type 1).

The next example of a dominant solvable e-ring game illustrates the computation of equilbria using the message structure explained above.

E-ring game of depth 4. There are two players, row (player 1) and column (player 2). Each player is initially informed about the number of messages she receives, and the payoffs depend only on the number of messages the player receives as well as on the actions chosen by both players. Each player either gets 1 or 2 messages, whereby player 2 either has the same number or one more message than player 1. 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 her opponent has either 1 or 2 messages, each event with equal probability ($p_1 = 1/2$); player 1 with 2 messages knows for sure the other player also has 2 messages. Similarly, player 2 with 1 message knows for sure that her opponent also has 1 message; while player 2 with 2 messages knows her opponent has either 1 or 2 messages, each event with equal probability ($p_2 = 1/2$).

Consider the following payoff matrices, where, respectively, A, B, C are the actions of player 1 and a, b, c the actions of player 2, and where $u_1(t_1)$ are the payoffs of player 1 when she receives t_1 messages, and $u_2(t_2)$ the payoffs of player 2 when she receives t_2 messages.

$u_1(1)$	a	b	c
A	80	60	80
B	200	100	140
C	120	140	180

$u_2(1)$	a	b	c
A	80	160	180
B	40	140	80
C	60	100	140

$u_1(2)$	a	b	c
A	60	80	40
B	80	20	20
C	160	120	180

$u_2(2)$	a	b	c
A	180	20	100
B	120	40	140
C	160	80	200

The above payoff structure has a unique (interim correlated) rationalizable action for all players and number of messages. Player 1 with 2 messages (payoff matrix $u_1(2)$) has a strictly dominant action C. Player 2 with 2 messages (payoff matrix $u_2(2)$), seeing this and the fact that player 1 with 2 messages has A as strictly dominated action, (and knowing that she faces player 1 with $t_1 = 1, t_1 = 2$ with equal probability), has a unique strict best reply c. Player 1 with 1 message (payoff matrix $u_1(1)$), given the above and seeing that player 2 with 2 messages has a as a strictly dominated action, (and again knowing that she faces player 2 with $t_2 = 1$, $t_2 = 2$ with equal probability), has a unique strict best reply C. Finally, player 2 with 1 message (payoff matrix $u_2(1)$), knowing that for sure she faces player 1 with 1 message 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. □

3 Identification of Orders of Rationality

We now introduce a novel take on the problem of identification of orders of rationality by adopting an axiomatic approach. Section 3.1 recalls some basic game-theoretic notions, and formalizes the notion of orders of rationality and the identification of the latter via *revealed rationality*. Since the identification is based on choice data obtained from decisions in a given game, the particular features of the implemented game can crucially affect the identification and, eventually, put its external validity into question. Section 3.2 deepens on these extrapolation concerns and presents two properties, *lower order consistency* and *absence of framing*, aimed at minimizing identification errors. Somewhat surprisingly, it turns out that lower order consistency and absence of framing *alone*, greatly narrow the space of all conceivable games: as Proposition 1 in Section 3.3 shows, if the analyst aims at the most simple experimental design that combines both properties, she is left precisely with (simply) dominance solvable e-ring games.

3.1 Preliminaries

Games and *k*-th Order Rationality

A (*Bayesian*) *game* consists of a list $\mathcal{G} := \langle T_i, A_i, u_i, \pi_i \rangle_{i \in I}$ where I is a finite set of *players*, and for each player i we have a finite set of *types* T_i, a finite set of *actions* A_i, a *utility function* $u_i : T \times A \to \mathbb{R}$, and a *belief function* $\pi_i : T_i \to \Delta(T_{-i})$. A *conjecture* for player i is a probability function $\mu_i \in \Delta(T_{-i} \times A_{-i})$. Conjecture μ_i is *admissible* for type t_i if its marginal over T_{-i} coincides with $\pi_i(t_i)$, and *believes* in event $E \subseteq T_{-i} \times A_{-i}$ if it assigns probability 1 to E.[6] The set of best responses to conjecture μ_i admissible for type t_i consists on the actions that maximize the expected utility induced by t_i and μ_i, that is:

$$\arg\max_{a_i \in A_i} \sum_{t_{-i} \in T_{-i}} \sum_{a_{-i} \in A_{-i}} \mu_i[(t_{-i}, s_{-i})] \cdot u_i((t_{-i}; t_i), (a_{-i}; a_i)).$$

Given a type, each action can be consistent with rationality, rationality and belief in rationality, or rationality and some other higher order belief in rationality. Formally, we say that action a_i is 1-*st order rational* for t_i if a_i is a best response to some admissible conjecture for type t_i. For order $k \geq 2$, proceeding recursively, we say that action a_i is *kth order rational* for type t_i if a_i is a best response to some admissible conjecture for type t_i that believes that her opponents' play $(k-1)$-th order rational strategies.[7] Finally, we say that game $\mathcal{G}$ is:

- *Dominance solvable*, if for every player i and every type t_i there exists some $k \geq 1$ such that only one action is k-th order rational for type t_i.
- *Simply* dominance solvable, if it is dominance solvable and, in addition, for every order $k \geq 1$ there exists some type t_i that has a unique k-th order rational action. As discussed in Remark 1, this strengthening of dominance solvability is convenient for minimizing identification errors.

Identification and Classification of Orders of Rationality

The identification in this paper relies on *revealed rationality* and can be summarized as follows: a subject is asked to make a choice in the role of every type of every player in the game, and her order of rationality is estimated as the lowest $k \geq 0$ for which every choice is k-th order rational

[6] The notation is standard. By $T := \prod_{i \in I} T_i$ and $A := \prod_{i \in I} A_i$ we denote the set of type and action *profiles*, respectively, and for each player i we write $T_{-i} := \prod_{j \neq i} T_j$ and $A_{-i} := \prod_{j \neq i} A_j$. $\Delta(T_{-i} \times A_{-i})$ denotes the set of probability functions on $T_{-i} \times A_{-i}$.

[7] That is, such that $\mu_i[\{(t_{-i}, a_{-i}) : a_j \text{ is } (k-1)\text{-th order rational for } t_j \text{ for every } j \neq i\}] = 1$. To keep our results easily comparable, here we follow the terminology in Kneeland (2015), but notice that an action a_i is k-th order rational for type t_i if and only if it is k-th order *interim correlated rationalizable* for t_i, as defined by Dekel, Fudenberg and Morris (2007) and Battigalli, Di Tillio, Grillo and Penta (2011). For further details about the solution concept, the reader is referred to these two papers.

at its corresponding role.[8] A subject whose choice vector is identified as being of order $k \geq 0$ is *classified* as being of order Rk. More formally, let $X := \bigcup_{i \in I} T_i$ denote the set of *player types* in game $\mathcal{G}$,[9] that is, the set of all possible roles of the subject. The set of observable *choice vectors* is:

$$\mathcal{C} := \{(a_i, t_i) \,|\, a_i \in A_i, t_i \in T_i, i \in I\}.$$

Then, for a given game $\mathcal{G}$, revealed rationality provides a well-defined mapping between the set of observable choice vectors and the possible orders of rationality:

Definition 2 (Identification) *Let $\mathcal{G}$ be a game. Then, the* identification *induced by $\mathcal{G}$ is the map $\mathcal{I} : \mathcal{C} \to \mathbb{N} \cup \{0\}$ where for every choice vector $\{(a_\ell, x_\ell)\}_{x_\ell \in X}$,*

$$\mathcal{I}(\{(a_\ell, x_\ell)\}_{x_\ell \in X}) := \min \{k \geq 0 \,|\, a_\ell \text{ is } k\text{-th order rational for every } x_\ell \in X\}.$$

$\mathcal{I}$ identifies *order $k \geq 1$ if game $\mathcal{G}$ has a unique player type x_k all of whose actions except one fail to be k-th order rational and, in such case, we say that player type x_k is used to identify order k.*

E-ring game of depth 4 (example continued). Consider again the game discussed at the end of Section 2, which we here reproduce for convenience.

$u_1(1)$	a	b	c
A	80	60	80
B	200	100	140
C	120	140	180

$u_2(1)$	a	b	c
A	80	160	180
B	40	140	80
C	60	100	140

$u_1(2)$	a	b	c
A	60	80	40
B	80	20	20
C	160	120	180

$u_2(2)$	a	b	c
A	180	20	100
B	120	40	140
C	160	80	200

Applying the above definitions yields the following identification of orders of rationality. 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 1

[8] An alternative identification method is Kneeland's (2015) *exclusion restriction*. We do not use this method for three reasons. First, using this identification strategy would eliminate the standard classes of 4×4 and BC games, thereby severely limiting our between games comparisons. Second, a main criticism to this approach 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. Third, there are influential theoretical frameworks for which a subject satisfying lower-order rationality might respond to changes in higher order payoffs (see Alaoui and Penta (2016, 2018a,b))

[9] Notice that, for games of complete information, the set of player types coincides with the set of players of the game.

message would be identified as $R3$. Similarly, a subject playing C as player 1 with 2 messages and c as player 2 with 2 messages, while playing B, D as a player 1 or b, c, d as player 2 with 1 message, would be identified as $R2$. Furthermore, a subject playing C as player 1 with 2 messages while playing a or d as player 2 with 2 messages and B, C, D or b, c, d as either player 1 or 2 with 1 message, would be identified as $R1$. Finally, a subject playing any of the dominated actions would be identified as $R0$.[10] □

3.2 An Axiomatic Approach to Identification

Lower Order Consistency

It is natural to expect that, when choosing as a player type that only has one ℓ-th rational action a_ℓ, a subject who systematically performs a hierarchical reasoning process of order $k \geq \ell$, would opt for said action. Accordingly, a subject who follows a different decision making rule should be expected to fail to choose a_ℓ for some $\ell = 1, \ldots, k$. Hence, an identification that, when identifying order k, also identifies orders $\ell = 1, \ldots, k$ poses an additional challenge *only* to those decision makers who *do not* consistently engage in higher-order reasoning, and serves, in consequence, as an estimation error minimizing check. We formalize the requirement as follows:

Property 1 (Lower Order Consistency) *Identification $\mathcal{I}$ is* lower order consistent *if whenever it identifies order $k \geq 1$ it also identifies orders $\ell = 1, \ldots, k$.*

Unsurprisingly, lower order consistency is an implicit standard in the modern literature in identification of rationality orders—e.g., Kneeland (2015) or Lim and Xiong (2016), and is of course a property satisfied by e-ring games. However, other classes of games often employed for identification, such as bimatrix games or beauty contests, fail to satisfy this requirement.[11] The following simple observation shows that, besides its intuitive appeal, lower order consistency also pins down a rather narrow class of games:

Lemma 1 *Let $\mathcal{G}$ be a game with lower order consistent identification of orders $k \geq 1$. Then $\mathcal{G}$ has at least k distinct player types of which one has a strictly dominant action. Moreover, if $\mathcal{G}$ identifies exactly k orders, then it is dominance solvable in exactly k steps.*

Remark 1 The requirement in Definition 2 that there be just one action that is k-th order rational is made to minimize the likelihood that such actions be chosen by chance, thereby leading

[10] In this example, we explain our identification strategy *as if* subjects switched roles. In the experiment, we achieve this by reassigning player 1's matrix with 2 messages to player 2 with 1 message while reallocating the other matrices to maintain the dominant solvability structure.

[11] In bimatrix games there are, at most, two player types. Beauty contests, due to the symmetric role its players, only possess one player type. Notice that the p-beauty contest games can identify at most $k = 1$ levels, and the two-player complete information, dominance solvable games can identify at most up to $k = 2$ levels. On the other hand, a ring game with k players and an e-ring game of depth k can identify k levels of higher-order rationality.

to identification error. This implies that, if a game has a lower order consistent identification, it must be simply dominance solvable.

Absence of Framing

In principle, it could be expected that subjects' decision making rules were context-dependent, so that observing sophisticated higher order reasoning in the game employed for identification would not mean that this behavior would extend to games of different nature. Hence, in order to minimize this variant of identification error, and enhance the external validity of the identification, the game structure should not *frame* players into the *precise* hierarchical thinking that is the object of the identification. Moreover, a game that frames players into hierarchical thinking makes it easier for players that have some degree of hierarchical thinking to reach higher levels. To better illustrate this phenomenon let us discuss the following two situations:

$G1$. There are three players, 1, 2 and 3. Player 1's utility only depends on her own choices, Player 2's utility depends on her choices and those of Player 1, and Player 3's utility depends on her own and Player 2's choices. As a consequence, Player 3's second order belief has *only* one possible order that is consistent with the payoff dependency of the game: her first order belief is about Player 2's choices and her second order belief, about Player 2's first order belief about Player 1's choices.

$G2$. There are three players, 1, 2 and 3. Player 1's utility only depends on her own choices, while Player 2 and 3's utilities depend on all three players' choices. In this case, Player 3's second order belief has *two* possible orders that are consistent with the payoff dependency of this game: (1) her first order belief is about Player 2's choices and the second, about Player 2's first order belief about Player 1's choices; (2) her first order belief is about Player 2's choices and her second order beliefs, about Player 2's first order belief about Player 3's choices.

As in the case of lower order consistency above, the ambiguity in $G2$ on the possible orders that can be used to construct the belief hierarchy seems to be an obstacle *only* for subjects that do not systematically reason hierarchically. On the contrary, the simplicity of the structure in $G1$ naturally frames subjects into hierarchical reasoning.[12] To minimize this notion of framing, the game should allow each player type to be able to construct hierarchical orders of other player types that are alternative to the natural one associated with the payoff structure of the game. This can be achieved by enriching the payoff dependencies so that, for player types of order 2

[12]Of course, the distinction above deals with subjects that, unlike what the standard model of higher-order reasoning admits, do not form *joint* beliefs about their opponent's behavior and higher-order beliefs (i.e., Player 2 may have a joint belief about Players 1 and 3's behavior in $G2$). However, this is immaterial for the argument: ideally, we want to avoid that players that have difficulties in forming these joint conjectures are categorized *as if* they were able to form them.

and above, payoffs always depend on additional player types than the ones associated with the natural hierarchy of beliefs. This idea is easy to formalize in the language of graphs; to this end, let us introduce some terminology first:

Definition 3 (Link Structure) *Let $\mathcal{G}$ be a game. The* link structure *of $\mathcal{G}$ consists of a set of* directed links *$\mathcal{L} \subseteq X \times X$ where, for any pair of distinct player types z_1 and z_2, $(z_1, z_2) \in \mathcal{L}$ if and only if the following two conditions hold:*

(i) z_1's payoffs depend on the actions of z_2.

(ii) z_1 has no strictly dominant action.

A path *between z_1 and z_n is a sequence $(z_1, z_2, ..., z_n)$ such that $\{(z_i, z_{i+1})\}_{i=1}^{n-1} \subseteq \mathcal{L}$ and all the z_i's except possibly z_1 and z_n are pairwise distinct.*

With this, the absence of framing into hierarchical reasoning can be formalized as follows:

Property 2 (Absence of Framing) *A lower order identification $\mathcal{I}$ is* framing-free *if whenever it identifies order $k \geq 2$ then, for any $\ell = 2, \ldots, k$, the link structure of game $\mathcal{G}$ has at least two distinct paths that start at x_ℓ and are of length $\ell - 1$ and $\ell - 2$.*

There are two aspects to this definition. First, at any level ℓ of the hierarchy of beliefs there should be at least one path distinct from the natural one (of length $\ell - 1$) associated with the payoff structure. Second, we want to avoid that a subject of rationality level $\ell - 1$ be identified as ℓ due to framing. This implies that for each player type $\ell \geq 2$, we need at least two distinct paths of lengths $\ell - 1$ and $\ell - 2$.

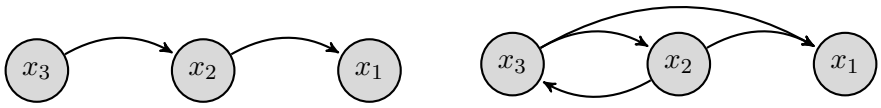

Figure 1: Games $G1$-Ring Game (left) and $G2$ (right).

To see that this formalism captures the intuition discussed at the beginning of the paragraph, Figure 1 illustrates the games $G1$ and $G2$ (where player k is identified with player type x_k). In $G1$, player type x_3 is framed as she has only one path of length $\ell = 2$ and of length $\ell = 1$. By contrast, in $G2$, player type x_3 has two paths of both length $\ell = 2$ and $\ell = 1$. Notice that $G1$ has the same link structure as a ring game, therefore such a class of games does not satisfy Property 2.

Figure 2: Game with some framing.

Figures 2 illustrates a case with less extreme framing. Here player types x_2 and x_4 are framed: player type x_2 has a single path of length 1 and player type x_4 has a single path of length 3. Player type x_3 is not framed as she has two paths of both length $\ell = 2$ and $\ell = 1$.

Finally, Figures 3 and 4 below show the link structures of two framing-free games.

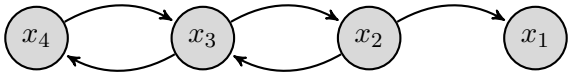

Figure 3: Framing-free game. E-Ring Game

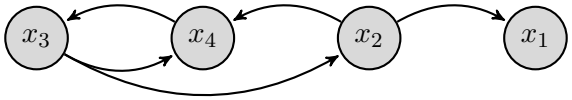

Figure 4: Framing-free game.

3.3 E-Ring Games as Minimal Class of Games Guaranteeing Lower Order Consistent and Framing-Free Identification

It turns out that, as we show in Proposition 1 below, by imposing lower order consistency and absence of framing, the class of games the analyst may employ in the identification of up to 4 levels of rationality is significantly reduced.[13] If, in addition, for the sake of simplicity of the implementation the game is required to be *minimal*, then we are left with just the e-ring games. A game is minimal if it has a minimal number of players, in order to reduce the noise in the belief formation process, and its link structure contains the lowest possible number of links, to minimize the complexity of the game.

Proposition 1 *Let $\mathcal{G}$ be a game. Then, $\mathcal{G}$ is minimal among the class of games inducing a lower order consistent and framing-free identification that identifies exactly order* 4 *if and only if $\mathcal{G}$ is a simply dominance solvable e-ring game of depth* 4.

Proof. The 'if' part is immediate so we focus on the 'only if' one. Lemma 1 implies that $\mathcal{G}$ is dominance solvable and that it has player types x_1, x_2, x_3, x_4 and a link structure containing (x_4, x_3), (x_3, x_2) and (x_2, x_1). Minimality of players implies that player types x_1 and x_3 belong to one player and player types x_2 and x_4 belong to the other. Moreover there cannot be links between player types belonging to the same player. Given this, absence of framing implies that the link structure also contains (x_3, x_4) and (x_2, x_3). The only other link that can be added is (x_4, x_1), but is excluded by minimality of links. We are thus left with the link structure of Figure 3. ■

[13] We focus on 4 orders of rationality since the experimental literature agrees that these are empirically the most relevant ones.

4 Experiment

4.1 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 standard 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. 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) commonly known 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.

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 contest games) to empirically classify individuals according to the revealed rationality approach. We chose these classes of games as they are the ones most frequently used in the literature for the identification of the empirical distribution of higher orders of rationality. Moreover, they are particularly convenient to test the empirical validity of the two axioms proposed in Section 3.2, since the 4×4 dominance solvable games and the beauty contest games do not satisfy lower order consistency, while the ring games satisfy lower order consistency but not absence of framing.

We designed eight 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 Online Appendix.

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.[14]

4.2 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 and compliance with our instructions.

Instructions were read aloud and included examples of the payoff consequences of several actions in each of the tasks. Participants answered a demanding comprehension test prior to each of the tasks. A majority of subjects (71%) answered all 13 questions correctly. We made sure that all remaining issues were clarified before proceeding to the actual experiment.[15] 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.[16] 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

[14] 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%).

[15] Although our analysis uses the full sample of participants, results are robust to using the subsample of subjects who made no mistakes in the tests.

[16] 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.

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.

4.3 Experimental Results

Empirical Relevance of the Properties. The first empirical question is whether the desirability of the proposed properties is actually justified by the data. To test lower order consistency (Property 1), we check whether the probability of classifying subjects in higher categories is higher in those games that do not satisfy it.

To this end, we compare the proportion of subjects identified as having higher order beliefs in rationality ($R2$, $R3$ or $R4$) in 4×4 games and in the two beauty contest games but that are identified as not having such beliefs (and thus being classified as $R0$ or $R1$) in both ring games and e-ring games, and vice versa.

Below we report the proportions of subjects (out of the total population) identified as $R2$-$R4$ in 4×4 and beauty contest games, and whose highest rationality level identified in e-ring and ring games is $R0$ or $R1$:

$$4 \times 4\text{: } 0.14 \quad 2/3\text{-BC: } 0.19 \quad 1/3\text{-BC: } 0.09.$$

When we calculate the proportions of subjects identified as $R0$ or $R1$ in 4×4 and beauty contest games, and who are identified as $R2$-$R4$ in e-ring and ring games, we obtain the following:

$$4 \times 4\text{: } 0.04 \quad 2/3\text{-BC: } 0.02 \quad 1/3\text{-BC: } 0.12.$$

The data goes in the direction we should expect if the proposed Property 1 was relevant. That is, the first proportions should be greater than the second ones. This means that 4×4 and 2/3-BC games are clearly worse in identifying higher order beliefs. The data does not exclude that the 1/3-BC game may do reasonably well at identifying higher order rationality.

By contrast, when we invert the games, we find that the proportions of subjects identified as $R2$-$R4$ in ring and e-ring games, and who are identified as $R0$ or $R1$ in 4×4 and beauty contest games are:

$$\text{E-ring games: } 0.01 \quad \text{Ring games: } 0.01,$$

whereas, the proportions of subjects identified as $R0$ or $R1$ in ring and e-ring games, and who are identified as $R2$-$R4$ in 4×4 and beauty contest games are:

$$\text{E-ring games: } 0.11 \quad \text{Ring games: } 0.14.$$

Again, the data goes in the expected direction.

Next, we build on the established empirical relevance of Property 1 to understand the importance of requiring absence of framing (Property 2). For this, we look at the distribution of levels of rationality obtained in ring games and e-ring games by those individuals whose maximum levels of rationality identified in the other games is not higher than 1 (i.e., $R0$ or $R1$ in 4×4 games and BC games). We do this, given that, if Property 1 holds, 4×4 games and BC games are good tests for levels $R0$ or $R1$. Individuals that do not show higher order beliefs in any of these games have a higher probability of not having been misidentified. We focus on this particular population of 36 subjects because the strongest effects of framing (from the e-ring and ring games), if present, should be highlighted within a population that shows otherwise no evidence of higher order beliefs.

Table 1 presents the cumulative distribution function of the rationality levels as classified by the e-ring and ring games among participants who are identified as $R0$ or $R1$ in the 4×4 and 1/3-BC games. We find that the ring games consistently classify subjects in higher categories than the e-ring games. In fact, as is clear from Table 1, the distribution of levels identified by the ring games first order stochastically dominates the one identified by the e-ring games.[17]

	$R4$	$R3$	$R2$	$R1$	$R0$
E-ring game	0.0%	8.3%	33.3%	75.0%	100.0%
Ring game	16.7%	22.2%	38.9%	75.0%	100.0%

Table 1: Cumulative distribution of rationality levels for e-ring and ring games for individuals identified no more than $R1$ in the 4×4 and 1/3-BC games.

We find further evidence of the relevance of Property 2. Figure 5 reports the proportion of subjects classified as level Rk, for all but the p-BC game with unspecified p, irrespective of the order of the tasks.[18] First, when looking at the distribution of the classification of subjects according to the ring game, we 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. Second, 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). We interpret this evidence as highlighting two things. First, the properties are addressing potentially serious flaws of the different methods. Second, they capture the observed problems in the right direction, given that the e-ring games, that are the only

[17]Requiring participants to be identified as $R0$ or $R1$ in all three games (4×4, 2/3-BC and 1/3-BC games) leaves us with only 8 participants, while maintaining the stochastic dominance of the cumulative distribution of the ring games over the e-ring games.

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

games satisfying both properties, seem to be less affected by these potential misidentification problems.

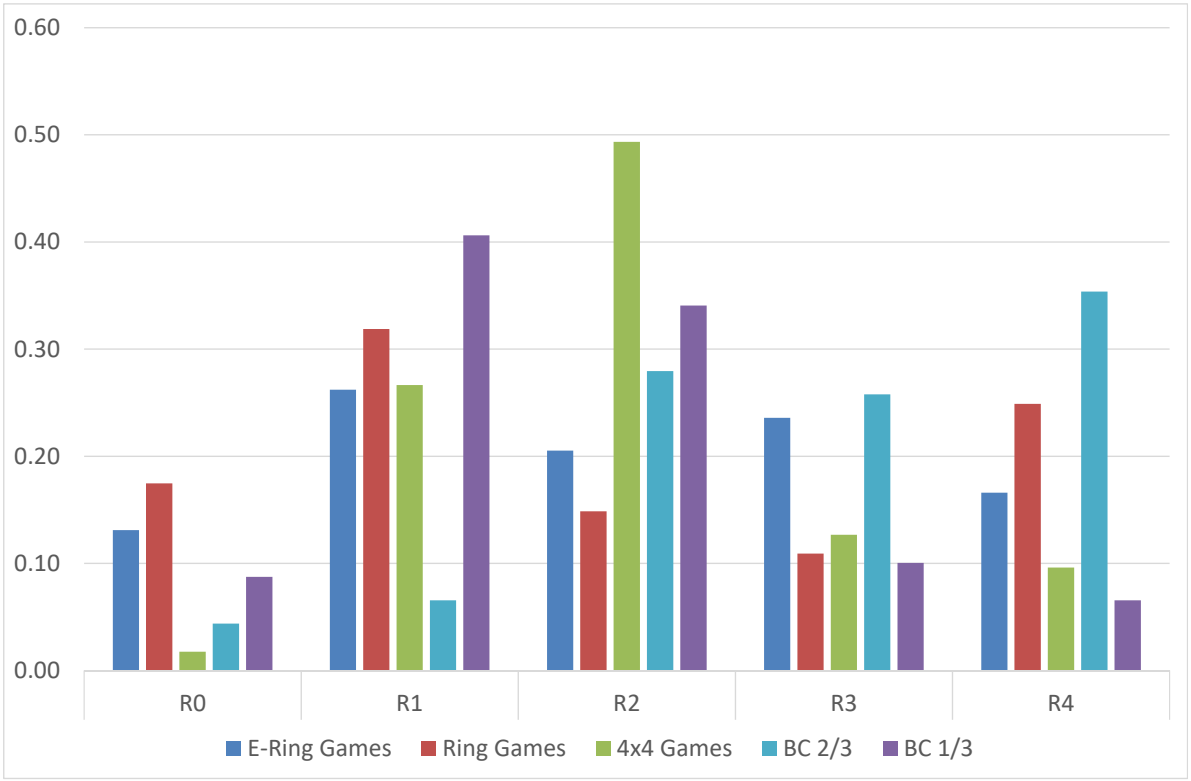

Figure 5: Classification by order of rationality, by game.

Existence of an Underlying Distribution. Having studied the empirical relevance of Properties 1 and 2, we now address a further potential problem with this kind of exercise. The main aim of the literature is to identify the distribution of types in the population. This assumes that, either individuals truly behave as described by the model (*as is approach*) or that their behavior can be described by the model (*as if approach*). In both instances it is fundamental to understand whether there is a stable distribution of types in the population.

We first look at the aggregate results. As shown in Figure 5, 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. All this evidence seems to suggest that there is no stable distribution and assuming its existence might lead to serious mistakes. 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.[19]

[19]Notice that in the 2/3-version of the BC game, numbers below 30 and 20 are already classified as, respec-

We also find additional treatment effects. First, 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)). Second, 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 7-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.[20] The distribution of the Rk levels, conditional on passing the test, is not qualitatively different from the unconditional case.[21] These findings suggest that the identified rationality level might be not only game dependent but also path dependent.

The high degree of variation of the classification across games is confirmed 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 $R0$, 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

tively, $R3$ and $R4$. When looking at the reasoning processes reported in the 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.

[20]Notice however that our test is much more demanding (7 vs 3 questions). Hence, statistically, we should expect to find a difference.

[21]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.

1/3-BC game for 45%. However, this result is mainly driven by subjects classified as $R1$ and is therefore not very informative.

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 increases to 30% if we exclude beauty contest games and is 49% if we exclude e-ring, ring and 4×4 games.[22]

An additional method to measure the stability of the relative ranking across classes of games is to check the correlation between the distributions obtained for 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 exhibits higher correlations than the other games.[23]

	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.

The fact that the e-ring games exhibit higher correlations than all the other classes, 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

[22]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%.

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

E-ring: 0.29 Ring: 0.14 4×4: 0.01 2/3-BC: 0.37 1/3-BC: 0.53.

statistical noise (e.g. mistakes, idiosyncratic game characteristics, etc.) that might increase the probability of identification errors. A robust way of identifying this distribution, within the revealed rationality approach, is to classify each individual by the rationality level she has been most frequently identified with across all classes of games.[24] 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 games, 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:

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

The e-ring games outperform the other classes of games by a significant margin.[25] This leads us to conclude that e-ring games are rather successful in identifying the relative ranking of individuals once we take into account possible statistical noise.

A final piece of evidence pointing in the same direction is the higher correlation between the distribution of rationality levels in the e-ring games with an independent measure of cognitive ability than for any of the other classes of games. Indeed, the correlation between the rationality levels as identified by the different classes of games and the ranking of the subjects based on the results of the standardized test used for the admittance to university is as follows:

$$\text{E-ring: } 0.24 \quad \text{Ring: } 0.12 \quad 4 \times 4\text{: } 0.06 \quad \text{BC: } 0.05.$$

To conclude, while we do not find evidence of a stable absolute ranking of individuals, that is, allowing for no mistakes and no variation across games, we do find some kind of stable relative ranking of individuals when relaxing these constraints. In particular, it seems that the e-ring games do relatively well at identifying this kind of ranking.

5 Concluding Remarks

The identification of a reliable distribution of orders of rationality in the population is a crucial exercise for many applications. As argued in the introduction, it can have an impact on the

[24] 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.

[25] Notice that while, statistically, e-ring games should outperform BC and 4×4 games due to the higher number of choices and hence the higher informative content of the classification, there should be no difference between e-ring games and ring games in terms of informativeness of the classification.

understanding of many economic problems like price formation, institutional design, monetary policy, oligopolistic competition, among many others. The main problem in this kind of exercise is that usual games do not allow for the observation of behavior at the different steps of the hierarchy of beliefs and, if they do, they might frame individuals into thinking in levels, thereby invalidating the very exercise.

This paper tackles this unresolved problem by taking an axiomatic approach. We propose two intuitive axioms that, at a practical level, narrow the class of games valid for identification down to a single class: *e-ring games*. The empirical evidence presented suggests that both properties are relevant and that the new class of games manages to significantly reduce the likelihood of identification errors and to more consistently identify the relative ranking of subjects. Nevertheless, the data cannot confirm the existence of a stable and game-independent distribution of rationality types in the population. 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 flexible or game-dependent approach.

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*.

Battigalli, Pierpaolo, Alfredo Di Tillio, Eduardo Grillo and Antonio Penta (2011). Interactive epistemology and solution concepts for games with asymmetric information. *The B.E. Journal of Theoretical Economics*, **11**, Article 6.

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.

Ellingsen, Tore, Robert Östling and Erik Wengström (2018). How does communication affect beliefs in one-shot games with complete information? *Games and Economic Behavior*, **107**, 153–181.

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.

Germano, Fabrizio, Jonathan Weinstein and Peio Zuazo-Garin (2019). Uncertain rationality, depth of reasoning and robustness in games with incomplete information. *Theoretical Economics, forthcoming*.

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.