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[4]

[1]Universitat Pompeu Fabra
[2]Barcelona Graduate School of Economics
[3]Universitat Ramon Llull, ESADE
[4]ICEF–HSE University

September 6, 2020

Abstract

Identifying higher orders of rationality is crucial to the understanding of strategic behavior. Nonetheless, the identification of a subject's actual order of rationality from observed behavior in games remains highly elusive. Games may significantly impact and hence invalidate the identified order. To tackle this fundamental problem, we introduce an axiomatic approach that singles out a new class of games, the *e-ring games*. We then present results from a within subject experiment comparing individuals' classification across e-ring games and standard games previously used in the literature. The results show that satisfying the axioms introduced significantly reduces errors and contributes towards a more reliable identification.

(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 for granting permission to run our experimental sessions. We thank Doug Bernheim, Tilman Börgers, Simone Galperti, Itzhak Gilboa, Yoram Halevy, Nagore Iriberri, Rosemarie Nagel, Antonio Penta, Ariel Rubinstein, Marciano Siniscalchi, Charles Sprenger and seminar participants at several institutions for very insightful comments that contributed to the development of this work.

1 Introduction

Bounded rationality have progressively permeated through most fields of economics in recent years, and many areas are incorporating theories of bounded hierarchical reasoning to enhance the realism and robustness of their models, including macroeconomic policy (Angeletos and Lian, 2016), mechanism design (Crawford, 2016; Börgers and Li, 2019; De Clippel, Saran and Serrano, 2019) and jury selection (Van der Linden, 2018), among many others. In monetary policy, for instance, understanding whether individuals are capable of forming higher order beliefs, whether such beliefs are bounded and, if so, what the actual bounds are, can lead the same interest rate policy to have rather different economic implications (García-Schmidt and Woodford, 2019).

In general, the applicability and predictive power of theoretical models that assume bounded hierarchical reasoning are strengthened by the ability to correctly identify the bounds in the population of interest. Indeed, in any interaction between rational agents, optimal behavior depends on the distribution of beliefs about whether others are rational, about whether others believe others are rational, and so on. Typically, the identification of the relevant reasoning bounds is crucial not only to guide economic modeling, but also for making empirical predictions of economic outcomes.[1]

Among the multitude of identification methods employed, the central and most reliable one has been to use the choices of experimental subjects in one or more games to identify the highest possible order of rationality consistent with their choices.[2] Crucially, given that such choices are made in a particular game, the structure of such game can influence and possibly bias the identification exercise. This paper introduces an axiomatic approach to study this problem.

In fact, for such identification method to be externally valid, it seems natural to require that the games used should ensure that (1) whenever a subject's choices are consistent with order k of rationality, those choices should be the outcome of a hierarchical reasoning process with k steps, and (2) the associated k steps have not been *induced* by the game (or games) being played, meaning that the structure of the games used does not in itself lead subjects to forming the *right* hierarchy of beliefs.

To better motivate these two requirements, consider the 2/3 beauty contest game with choices from $\{0, 1, \ldots, 100\}$. A strategy consistent with a certain order of rationality k may

[1] This paper focuses on the identification of orders of rationality, but the analysis is also applicable to any model of cognitive hierarchies, including level-k reasoning; see Germano, Weinstein and Zuazo-Garin (2020) for a recent discussion of various models.

[2] See 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); Georganas, Healy and Weber (2015); Kneeland (2015) among many others. Part of this literature has complemented choice based methods with other methodologies such as eye-tracking or search patterns recorded on computer interfaces. As discussed in Kneeland (2015), these methods are not always fully reliable, they may be difficult to implement in certain contexts and they may influence the way subjects choose in such games.

be chosen for a variety of reasons, raising the possibility of an identification mistake. For example, choosing 17 in such a beauty contest game is consistent with being of order $k = 4$. Nevertheless, given that we do not observe a subject's reasoning process, we cannot exclude that such a choice has been taken for other reasons such as the number corresponding to the subject's birthdate or her most preferred number. However, 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 being of order k *only if* her behavior at each step is consistent with such a classification, within the same game (that is, consistent at k and at each step from 1 to $k-1$). This motivates requirement (1) above, which contributes to reducing the probability of identification mistakes while maintaining a choice-based approach.[3] At the same time, games that allow to test for behavior at each step of the hierarchy might themselves induce, or *frame*, subjects into thinking hierarchically. Subjects might be pushed into making choices that are of higher order k, or simply into thinking hierarchically, because the structure of the games makes the iterated elimination steps, and hence the associated hierarchy of beliefs, apparent, thereby making the identification flawed. This motivates requirement (2) above, capturing a novel and intuitive notion of framing, which we formally define in the paper.

This paper addresses the highlighted identification problem both theoretically and empirically. On the theoretical side, we use for the first time an axiomatic approach to reduce the potential misidentification of higher orders of rationality. We propose two intuitive properties that pin down a unique and novel class of games we refer to as the *e-ring games*. To the best of our knowledge, none of the games previously used in the literature fulfill both requirements. On the empirical side, we test the validity of the two axioms proposed by comparing behavior across the most prominent classes of games used to identify levels of rationality and the *e-ring games*.

Regarding the theoretical contribution, we formalize the two requirements as follows. The first one, *lower order consistency*, ensures that individual behavior can be tested at *each step* of the hierarchy of beliefs, as in requirement (1) above. This property is satisfied by the *ring games* used in Kneeland (2015) and introduced by Cubitt and Sugden (1994).[4] The second property, *absence of framing*, formalizes requirement (2) above by imposing that the payoff structure of the game should be such that each level of the hierarchy of beliefs has multiple payoff interdependencies, and not just ones with lower levels. The property is formulated using the

[3] Applied to the 2/3 beauty contest example, ideally, for a subject that chooses 17 ($k = 4$), we would like to observe the same subject choose first a number between 44–66 ($k = 1$), second a number between 29–43 ($k = 2$), then a number between 19–28 ($k = 3$), all in the same game, which is impossible. Another possibility would be to make a subject play a series of (necessarily) different dominance solvable games. The problem with this is that players' beliefs and hence behavior may change because the games are changing, thereby invalidating the excercise.

[4] The *ring(-network) games* used in Kneeland (2015) are finite dominance solvable static games, 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 has a single strictly dominant action that allows to initiate the dominance solvability procedure.

language of graphs and guarantees that the payoff dependencies of the game do not correspond exactly with the "natural" hierarchy of beliefs, hence enabling players to contemplate alternative hierarchies.[5]

Surprisingly enough, lower order consistency and absence of framing greatly narrow down the set of available games. We show that the simplest class of games satisfying both properties, and identifying up to four levels of rationality—the empirically relevant ones—is a specification of a new class of games we present here, the *e-ring games*.[6] An *e-ring game* is a static game with private values, where the incompleteness of information is structured by means of messages automatically sent back and forth between players as in the email game of Rubinstein (1989). This information structure generates a natural one-to-one correspondence between messages and higher-order beliefs.

Our empirical contribution consists in testing experimentally the validity of lower order consistency and absence of framing. We carry out an experiment where all subjects play games from each of the following four classes: 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).[7] We are left with games which satisfy both properties (e-ring games), one of them (ring games) and none of them (4×4 and the beauty contest games) and allowing us to test whether satisfying the properties is effective empirically in addressing the theoretical concerns raised above.

By observing subjects' choices in these games and following the *revealed rationality approach* we categorize 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 thus classified as Rk if all her actions are Rk and at least one is not $Rk{+}1$. That is, we assign players the maximal level of higher-order rationality consistent with the choices made (Tan and Werlang, 1988; Lim and Xiong, 2016; Brandenburger, Danieli and Friendenberg, 2017).

The experiment supports our theoretical approach. Indeed, we find evidence that the properties proposed are relevant in the following sense: (1) games that violate lower order consistency, and hence do not test for consistent choices at steps 1 to k, tend to overestimate the distribution of types for levels 2 or higher, compared to the ones that satisfy lower order consistency; (2) the ring games, which satisfy lower order consistency but not absence of framing, appear to

[5] In what follows, we denote by *natural* hierarchy of beliefs the one that corresponds to the order of elimination of dominated strategies implied by the game.

[6] See Section 3.2 for a justification of why simplicity of the games may be a desirable property in empirical applications.

[7] 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 where all subjects have to specify how they would play for any p in the interval.

frame subjects into hierarchical reasoning. In fact, we find that the distribution of types for levels 2 or higher is biased towards the maximum level 4, as compared to the e-ring games that satisfy both properties. Moreover, we find an order effect, whereby subjects having played the ring games before the e-ring games tend to be identified with higher orders in the e-ring games than when the e-ring games are played before the ring games.

The remainder of the paper is organized as follows. Section 2 presents the desirable properties a class of games should satisfy to reliably identify higher orders of rationality, Section 3 introduces the e-ring games and shows that they are characterized by the properties presented in Section 2. Section 4 describes the experimental design and the experimental results. Section 5 concludes. The Online Appendix contains an English translation of the experimental instructions and the payoff matrices of all games used in the experiment.

2 Identification of Rationality Bounds

Despite the importance of understanding the extent to which agents engage in hierarchical reasoning, no consensus has been reached about which games to use to identify individuals' higher order rationality bounds.

In this section, we introduce a novel take on this problem by adopting an axiomatic approach. We proceed as follows. Section 2.1 recalls standard game-theoretic tools that formalize the notion of rationality bound as well as some basic notions concerning its identification. As it is standard in the literature, our identification method relies on the choices observed in a given game to identify the bounds hence making the features of the game crucial for their reliability. Using these basic notions, Section 2.2 proposes two axioms, *lower order consistency* and *absence of framing*, that, by shaping the structure of the game, strengthen the validity of the identification.

2.1 Rationality Bounds and the Structure of Games

Games and Higher-Order Rationality

A *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* $\pi_i : T_i \to \Delta(T_{-i})$.[8] A *conjecture* for player i is a probability function $\mu_i \in \Delta(T_{-i} \times A_{-i})$ that represents player i's subjective beliefs about her opponents' types and actions. Conjecture μ_i is *admissible* for type t_i if its marginal on T_{-i}

[8] 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 $T_{-i} \times A_{-i}$. Clearly, if $|T_i| = 1$ for every player i, then the game has complete information.

coincides with $\pi_i(t_i)$, and it represents *belief* in event $E \subseteq T_{-i} \times A_{-i}$ if it assigns probability one to E. 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:

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

With these ingredients it is easy to formalize the idea of iterated elimination of strictly dominated actions which, as discussed in the following paragraph, enables us to rely on observed behavior to identify the rationality bounds:[9]

- Action a_i is (*1st order*) *rational* for type t_i if a_i is a best response to some admissible, arbitrary conjecture for t_i. This is equivalent to action a_i not being strictly dominated for t_i. Thus, a subject choosing a strictly dominated action cannot be classified as (1st order) rational.

- For order $k \geq 2$, proceeding recursively, action a_i is *k-th order rational* for type t_i if a_i is a best response to some admissible conjecture for type t_i that represents belief in opponents playing $(k-1)$-th order rational actions.[10] Similarly as above, this is equivalent to action a_i surviving k rounds of iterated elimination of strictly dominated actions for t_i. Hence, a subject choosing an action that does not survive k rounds of iterated elimination cannot be classified as k-th order rational.

Finally, a game $\mathcal{G}$ is *dominance solvable* if the iterated elimination of strictly dominated actions eventually yields a unique action for every type; that is, if there exists some $k \geq 1$ such that for every player i and every type t_i only one action is k-th order rational. The conceptual link between higher-order belief in rationality and the iterated elimination of strictly dominated actions is what allows for the identification of a subject's higher-order rationality *bounds* by observing her behavior in the underlying game.

Identification of Bounds

The procedure, based on inference via observed choices, consists of the following three steps: (1) a game is fixed, (2) a subject is asked to make choices in the role of every type of each player, and (3) this subject's rationality bound is identified as the *highest* $k \geq 0$ such that *every*

[9] To keep our results easily comparable, here we follow the terminology in Kneeland (2015) and employ the expression 'k-th order rational' instead of 'k-th order *rationalizable*'. Nevertheless, 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.

[10] That is, such that $\mu_i[\{(t_{-i}, a_{-i}) : a_j \text{ is } (k-1)\text{-th order rational for } t_j \text{ for every } j \neq i\}] = 1$. For the link between higher-order rationality and iterated elimination see Pearce (1984) and Tan and Werlang (1988).

choice of the subject is k-th order rational and at least one is *not* $(k+1)$, if such k exists, and ∞ otherwise.[11] This idea is formalized in the following definition and discussed in more detail further below.

Definition 1 (Revealed Rationality Bound) *Let $\mathcal{G}$ be a game. Then:*

(*i*) *A* player type *is a pair (i, t_i), where i is a player and t_i is a type for player i. We denote the set of all player types in the game by $X_{\mathcal{G}} := \bigcup_{i \in I} (\{i\} \times T_i)$.*

(*ii*) *A* choice vector *is an indexed list $(a_x)_{x \in X_{\mathcal{G}}}$ that ascribes an action $a_x \in A_i$ to each player type $x = (i, t_i)$.*

(*iii*) *The* revealed rationality bound *k that corresponds to choice vector $(a_x)_{x \in X_{\mathcal{G}}}$ is given by:*

$$k = \min \left\{ \max \left\{ k' \geq 0 : a_x \text{ is not } (k'+1)\text{-th order rational for } t_i \right\} : x = (i, t_i) \in X_{\mathcal{G}} \right\},$$

if k is a well-defined integer, and $k = \infty$ otherwise.

Thus the set of *player types* $X_{\mathcal{G}}$ represents the different roles that a subject can be asked to play in a game. Notice that, for games with complete information, the set of player types coincides with the set of players of the game. The *choice vector* $(a_x)_{x \in X_{\mathcal{G}}}$ is a description of the actions the subject is observed to have played in each of the roles.

The intuition behind the *revealed rationality bound k* has already been hinted at. If a subject is observed to choose an action that survives k but *not* $(k+1)$ rounds of iterated elimination of strictly dominated actions, then her rationality bound cannot be higher than k. In the opposite case, a subject consistently choosing at the highest possible level that the game allows for, cannot be excluded to have unbounded hierarchical reasoning, that is, $k = \infty$.

Building on these notions, the following definitions will be convenient for discussing the axioms in Section 2.2 and the characterization result in Section 3.2.

Definition 2 (Testing for Bounds) *Let $\mathcal{G}$ be a game. Then, we say that:*

(*i*) *Player type $x = (i, t_i)$* can test for bound *$k \geq 1$ if every action of player i except one fails to be k-th order rational for t_i, and in such case we denote x by x_k.*

[11] An alternative identification method is the one of *exclusion restriction* used in Kneeland (2015). 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 it rules out that subjects 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).

(*ii*) *Game $\mathcal{G}$* can test for bound $k \geq 1$ *if $X_{\mathcal{G}}$ contains some player type x_k that can test for bound k.*

The interpretation of a player type x_k as a *test* to discard whether bound k is reached is straightforward: A subject who fails to choose a k-th order rational action for this type fails the test and her bound is concluded to be strictly below k.[12] In this sense, a game that includes x_k can test for bound k.

Structure of Games

To better visualize the player types and their payoff dependencies resulting from the payoff structure of the game, we introduce some basic notions from the language of graphs. Player types are represented as nodes and payoff dependencies are represented as directed links. A path in a given graph can be seen as mapping a hierarchy of beliefs.

Definition 3 (Graph of a Game) *Let $\mathcal{G}$ be a game. The* graph *of $\mathcal{G}$ consists of the pair $(X_{\mathcal{G}}, L_{\mathcal{G}})$, where the player types $X_{\mathcal{G}}$ are nodes, and $L_{\mathcal{G}}$ is the set of* directed links*, i.e., pairs of nodes $(x, x') \in X_{\mathcal{G}} \times X_{\mathcal{G}}$ such that the following two conditions hold:*

(*i*) *x has no strictly dominant action.*

(*ii*) *x's expected payoff depends on the actions of x'.*

Definition 4 (Path) *Let $\mathcal{G}$ be a game with graph $(X_{\mathcal{G}}, L_{\mathcal{G}})$. A* path *is a finite sequence of nodes $(x^{(1)}, x^{(2)}, \ldots, x^{(n)})$, where the following two conditions hold:*

(*i*) *$(x^{(\ell)}, x^{(\ell+1)}) \in L_{\mathcal{G}}$ for every $\ell = 1, \ldots, n-1$.*

(*ii*) *All the nodes except possibly $x^{(1)}$ and $x^{(n)}$ are pairwise distinct.*

Thus, the *graph* of a game $\mathcal{G}$ summarizes the first-order payoff dependencies in the game. The existence of a directed link from player type x to player type x' ($(x, x') \in L_{\mathcal{G}}$) represents the fact that a rational type x should try to anticipate the choice by x' when evaluating what is optimal for her to do. If x has a strictly dominant action, this is of course not the case, and neither is it if the choice of x' does not affect the expected payoff of x. Higher-order payoff dependencies are captured by *paths*, which represent different possible hierarchies that the first player type in the path ($x^{(1)}$) can conceive when thinking strategically.

Let us now use an example to review these concepts.

Example: Bimatrix Games vs Ring Games. Consider the following two-player bimatrix game where the left matrix describes the payoffs of Player 1 (who has strategies A, B or C) while the right matrix describes the payoff of Player 2 (who has strategies a, b or c).

[12] There is a unique k-th order rational action for each type in order to minimize the probability that a subject playing randomly accidentally chooses such an action.

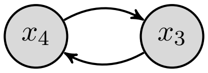
	a	b	c
A	80	20	140
B	60	160	20
C	100	200	40

Player 1

	A	B	C
a	120	20	200
b	20	40	60
c	100	120	80

Player 2

There are two player types that correspond to the two players of the game. Using Definition 2, Player 2 can test for bound 3, since a is the only strategy surviving three rounds of elimination of strictly dominated strategies. That is, Player 2 is x_3 in our notation. Player 1 can test up to bound 4 since C is the only strategy surviving four rounds of elimination of dominated strategies. That is, Player 1 is x_4. Notice that there are no further player types. In fact, even if the game is dominance solvable (with solution (C, a)), it cannot have a player type to test for each bound in the same game. This is a general problem with bimatrix games. This means that a subject capable of forming only first order beliefs, playing as Player 1 and choosing randomly over A and C, has high chances of being classified as if capable of forming higher order beliefs by playing C for example. The structure of the game is captured by the following graph.



The previous problem does not arise in the following game, which is a four-player *ring game* as used in Kneeland (2015).

	d	e	f
a	80	200	120
b	0	80	160
c	180	120	60

Player 1

	g	h	i
d	140	180	40
e	200	80	140
f	0	160	180

Player 2

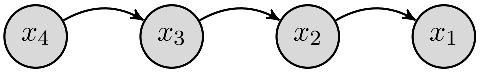
	j	k	l
g	200	140	80
h	160	20	180
i	0	160	160

Player 3

	a	b	c
j	120	160	140
k	80	120	100
l	60	100	80

Player 4

A defining feature of such a four-player ring game is that Player k's payoffs depend only on her own choice and on that of Player $k + 1$, up to Player 4 whose payoffs depend on her own choice and on that of Player 1. Notice that each player corresponds to a different player type, hence allowing the game in the example to also test up to bound 4. In fact, following our notation, we have Player 4 being player type x_1, Player 3 being x_2, Player 2 being x_3 and finally Player 1 being x_4. That is, ring games allow to test for the whole hierarchy of beliefs *within* the same game. In fact, it is enough to make a subject play in each role to test for each bound, *ceteris paribus*. A subject incapable of forming beliefs of order higher than the first, for example, would eventually make a *mistake* when playing in the role of Player 2 or 1. Nevertheless, the very structure of the game that corresponds to the iterative reasoning necessary to solve the game, might make the hierarchy of beliefs more evident to a (rational) subject that would otherwise be incapable of constructing one. This problem becomes more apparent from the graph of the game that, for each player type x_k, admits a unique path of length $k-1$.



The above discussion further underscores the importance of putting conditions on the structure of the game in order to adequately identify subjects' higher order reasoning bounds. We address this formally in the next section.

2.2 The Two Main Axioms

We here introduce our two main axioms, *lower order consistency* and *absence of framing.*

Lower Order Consistency

The example in the previous section shows that while some games that can test for a bound $k > 1$ can also test for all lower bounds $\ell = 1, \dots, k-1$, others are more limited and can only test for a subset of these bounds. In particular, some classes of games often used for identification, such as bimatrix games or p-beauty contest games, fail to test for all bounds. In such games, a subject who chooses randomly or non-rationally is likely to be wrongly interpreted as possessing a high rationality bound, as previously hinted.

At a conceptual level, a game that can test for bound k and, when doing so, can also test for bounds $\ell = 1, \dots, k-1$, makes it harder for subjects with bound strictly below k to pass all these tests, but it should not affect the behavior of a subject with bound k or above. In fact, a subject whose rationality bound is k or above should be expected to pass *all* the tests that correspond to bounds $\ell = 1, \dots, k$. By contrast, a subject whose rationality bound is $k' < k$ should be expected to fail *at least one* of the tests for bounds $\ell = k' + 1, \dots, k$.

Consequently, the explicit verification of every step of the reasoning hierarchy imposes additional challenges *only* to relatively unsophisticated subjects, but remains innocuous for sophisticated ones. Implementing such a verification significantly reduces the risk of overestimating a subject's rationality bound and, as a result, seems an obvious requirement if the identification is expected to have any external validity. Our first axiom formalizes this intuition.

Property 1 (Lower Order Consistency) *Game $\mathcal{G}$ is* lower order consistent *if whenever it can test for bound $k \geq 2$ it also can also test for bounds $\ell = 1, \dots, k-1$.*

Lower order consistency formalizes a property that has been implicitly used in the literature on identification of rationality bounds (see Kneeland, 2015, or Lim and Xiong, 2016). In terms of the graph of the game this property implies that the structure of a game testing up to bound k, has to contain the nodes x_1 to x_k. The following simple observation shows that, in addition to its intuitive appeal, lower order consistency also pins down a rather narrow class of games.

Lemma 1 *Let $\mathcal{G}$ be a game that satisfies lower order consistency. Then, for any $k \geq 1$,*

(i) *If $\mathcal{G}$ can test for bound k, then it has at least k distinct player types, of which one has a strictly dominant action.*

(ii) *If $\mathcal{G}$ can test for bound k but not for bound $k+1$, then it is dominance solvable in exactly k rounds.*

Proof. Part (i) follows from the definition of lower order consistency. To see this notice that, if $\mathcal{G}$ can test for bound k, it can also test for bounds $\ell = 1, \ldots, k-1$, and hence $X_{\mathcal{G}}$ contains player type x_ℓ for each $\ell = 1, \ldots, k$, where, by definition, x_1 has a strictly dominant action. Part (ii) follows from Definition 2 and from the definition of lower order consistency. ■

Absence of Framing

Mounting evidence from behavioral economics shows that individual behavior can be influenced by the context in which decisions are taken. Applied to the identification of rationality bounds, this suggests that the game employed may shape the actual reasoning process and *frame* the subjects (i.e., influence their reasoning process) in a way that *induces* the form of hierarchical thinking that is the object of the identification. Obviously, such a phenomenon would compromise the external validity of the identification by giving rise to the following two issues. First, subjects who would not normally engage in hierarchical thinking may be induced to do so by the game. Second, subjects with some order of hierarchical thinking may be induced to think in higher orders. To further illustrate this specific notion of framing, consider the following two situations:

$G1$. *A ring-like game with three players.* The game is dominance solvable. Player 1's payoffs only depend on her own choices, Player 2's payoffs depend on her choices and those of Player 1, and Player 3's payoffs depend on her own and those of Player 2. Furthermore, for each $k = 1, 2, 3$, Player k has a unique k-order rational action, so that each Player k can be identified with player type x_k. Figure 1 illustrates the graph of game $G1$. Notice that Player 3's second order belief has *only* one possible ordering 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.

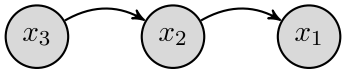


Figure 1: Game $G1$.

$G2$. *A variation of* $G1$. Take $G1$ and introduce an additional action $\bar{a}_3$ for Player 3 satisfying the following features: (*i*) $\bar{a}_3$ is strictly dominated for Player 3, (*ii*) Player 1's payoffs are independent of $\bar{a}_3$, and (*iii*) if Player 3 chooses $\bar{a}_3$ Player 2's worst possible option is to play her unique 2-nd order rational action of $G1$, independent of Player 1's choice. Obviously, the game is still dominance solvable, and, for each $k = 1, 2, 3$, Player k has a unique k-th order rational action and can thus be identified with player type x_k. However, the payoff dependency becomes slightly (though minimally) more intricate, as depicted in Figure 2. Notice that now Player 3's second order belief has *two* possible orderings that are consistent with the payoff dependencies represented in the graph: (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.

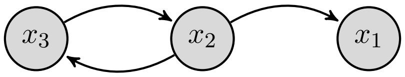


Figure 2: Game $G2$.

The comparison between the two scenarios is insightful. The multiplicity of orderings that can be used to construct the belief hierarchy in $G2$ leaves it open to the subject which hierarchy, if any, to follow. By contrast, the absence of multiplicity in the payoff dependency of $G1$ *frames* subjects to reason hierarchically.[13] Notice that this concern gains particular salience if the game is assumed to satisfy lower order consistency. In fact, the property requires the existence of a different player type to test for each bound. This can influence the subjects' reasoning process by exposing the belief hierarchy that represents the *inductive structure* of the game (i.e., the exact ordering of iterated elimination that solves the game).[14]

[13] 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 opponents' 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 having difficulties in forming these joint conjectures are categorized *as if* they were able to form them.

[14] The intuition is well conveyed in Kneeland (2015), whose ring games provide a major step forward towards the identification of rationality bounds by implicitly requiring lower order consistency: "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

To minimize this notion of framing, a game used to identify higher orders of rationality should allow each player type to be able to construct belief hierarchies about other players types' behavior that are alternative to the one associated with the inductive structure of the game. This can be achieved by enriching the payoff dependencies so that, for player types that test for bound 2 and above, payoff dependencies do not only refer to player types that test for lower bounds. Given this, it is easy to formalize a minimum requirement of the payoff dependencies of the game which prevents making the inductive structure of the game immediately apparent.

Property 2 (Absence of Framing) *A lower order consistent game $\mathcal{G}$ that can test for bound $k \geq 2$ is* framing-free *if there exist:*

(i) *For any $\ell = 2, \ldots, k$, two distinct paths of length $\ell - 1$ that start at x_ℓ.*

(ii) *For any $\ell = 3, \ldots, k$, two distinct paths of length $\ell - 2$ that start at x_ℓ.*

Let us provide some further intuition for the axiom. First, requiring $\mathcal{G}$ to be lower order consistent ensures that, if the game contains a player type x_k, then it also contains player types $x_1, \ldots, x_{k-1}$. Second, condition (i) says that a player type x_ℓ that tests for bound ℓ is considered to be *framed* if no distinct paths of length $\ell - 1$ that start at x_ℓ exist (by definition, one always exists). The interpretation is simple and visually intuitive in Figure 1. There, the payoff dependencies allow for a single path of length 1 departing from x_2, making it immediately apparent for x_2 that it is x_1 the type whose choice she cares about. This implies that a subject not capable of forming a hierarchy of beliefs might be helped by the structure of the game to behave *as if* she could. On the contrary, the presence of two distinct paths departing from x_2 in the graph in Figure 2, one towards a player who has no strictly dominant action, makes the inductive structure of the game less apparent.

The same intuition, visually represented in the left graph in Figure 3, where x_4 is interpreted as being partially framed, explains why we also require condition (i) for player types that test for bounds above 2. Here, the fact that there are not two distinct strategic paths of length 3 departing from x_4 results in the inductive structure of the game being easily recognizable for x_4, if she excludes herself from the belief hierarchy.

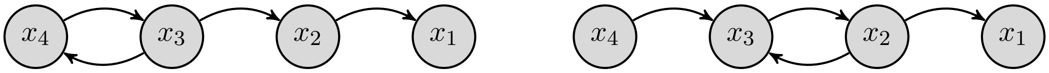


Figure 3: Games with some framing.

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."

In addition to condition (*i*), condition (*ii*) is also required for types that test for bounds 3 or above, in order to avoid situations such as the one in the right graph in Figure 3, where, again x_4 would be considered to be partially framed. The reason is that the fact that there is a unique path of length 2 that departs from x_4 results in the necessity of her first-order beliefs only pertaining player type x_3 immediately apparent to x_4.[15] This implies that a subject that can form up to second order beliefs, when playing as player type x_4, would immediately see the inductive structure of the game, hence behaving *as if* capable of forming third order beliefs. Finally, Figure 4 displays two different games in which no player type is framed. In the next section, we show that the game on the left is a particular specification of the new class of games to be introduced.

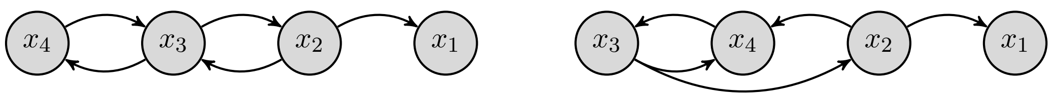


Figure 4: Games that are framing-free.

3 From the Axioms to the Structure of Games

In this section, we first present a new class of games, the *e-ring games*, and show that the simplest dominance solvable specification of this class is characterized lower order consistency and absence of framing.

3.1 E-Ring Games

An *e-ring game* is a two-player static game with private values in which players automatically receive a finite number of messages, and where each player's own payoffs depend on the number of messages that the player received as well as on the actions chosen by both players. Nature chooses the number of messages received by each player, whereby player 2 either has the same number of messages as player 1 or she has one more message than player 1.

The following example illustrates an e-ring game with three actions that is similar to the ones used in our experiments. It also shows that such specification can test for bounds up to 4.

Example: E-Ring Game of Depth 4. There are two players, Player 1 (the sender) who chooses rows, and Player 2 (the receiver) who chooses columns. Each player is initially informed about the number of messages she receives, and the payoffs depend only on the number of

[15] The requirement in Property 2 could be made more stringent and ask for every x_ℓ to have two distinct paths of length j for every $j = 1, \dots \ell{-}1$. It is important to emphasize that, while this may be a theoretically compelling alternative, it would be unnecessarily demanding for the characterization result we present in Proposition 1; it would not add any extra bite to the already narrow taxonomy we obtain there.

messages a 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 figure out the payoffs of the opponent, players can compute the number of messages received by the 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.

Player 1

	a	b	c
A	80	60	80
B	200	120	140
C	120	100	180

$t_1 = 1$

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

$t_1 = 2$

Player 2

	A	B	C
a	80	40	60
b	160	140	100
c	180	80	140

$t_2 = 1$

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

$t_2 = 2$

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 1 message (payoff matrix $u_1(1)$) 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, given the above and seeing that Player 2 with 1 message 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.

Now, we show that this particular game can test up to bound 4. According to Definition 1 we have the set of player types $X_{\mathcal{G}} = \{(1, 1), (1, 2), (2, 1), (2, 2)\}$ so that the payoff matrix that corresponds to each player type (i, t_i) is $u_i(t_i)$. Moreover, notice that according to Definition 2, we have that $x_1 = (1, 2)$, $x_2 = (2, 2)$, $x_3 = (1, 1)$, and $x_4 = (2, 1)$, where each player type x_k is the unique one used to test for rationality bound k. This is easy to see: C is the only rational action for $(1, 2)$, c is the only 2nd order rational action for $(2, 2)$, C is the only 3rd order rational action for $(1, 1)$ and c is the only 4th order rational action for $(2, 1)$. Thus the revealed rationality method yields the classification given in Table 1.

	x_1	x_2	x_3	x_4
$R4$	C	c	C	c
$R3$	C	c	C	b
$R2$	C	c	B	b, c
$R1$	C	b	B, C	b, c

Table 1: Choice vectors and revealed rationality bounds.

In addition, a subject playing a dominated action would be classified as $R0$ (A or B in the case of x_1, and A or a in the case of the remaining player types).[16] □

The next definition formalizes the general class of *e-ring games.*

Definition 5 (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:*

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

Notice that the type structure of the e-ring games builds on the communication structure of the email games of Rubinstein (1989) with two important differences. First, in the email games players can receive any arbitrary number of messages, and, second, they face the same 2 × 2 payoff matrices for essentially any number of messages received. To further clarify the relation

[16] In this example, we explain our identification strategy *as if* subjects switched roles. In the experiment detailed in Section 4, 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.

between types in an e-ring game, consider player i who has received ℓ messages. By Definition 5, this player's type is $t_i = \ell$ and the payoff she obtains from action profile (a_1, a_2) is given by $u_i(\ell, a_1, a_2)$. However, Player i is uncertain about the number of messages received by the other player and hence also about the latter's type and payoff function. In particular, Player 1 of type $t_1 = \ell$ knows that, with probability p_1, Player 2 is of type $t_2 = \ell$ and that, with probability $1-p_1$, Player 2 is of type $t_2 = \ell+1$ (with the exception of type $t_1 = k/2$, who knows that Player 2 is of type $t_2 = k/2$ for sure). Similarly, Player 2 of type $t_2 = \ell$ knows that, with probability p_2, Player 1 is of type $t_1 = \ell - 1$ and that, with probability $1 - p_2$, Player 1 is of type $t_1 = \ell$ (with the exception of type $t_2 = 1$, who knows that Player 1 is of type $t_1 = 1$ for sure).

3.2 Simplest Lower Order Consistent and Framing-Free Games

The theoretical appeal of lower order consistency and absence of framing for the identification of rationality bounds has been discussed in Section 2. We now turn to the question of the implementation of both axioms, and, more specifically, to the characterization of games satisfying the axioms and the complexity they entail. As previously mentioned, standard games used in the literature such as bimatrix games and the beauty contest games do not satisfy, in general, lower order consistency. This creates potential overestimation concerns related to subjects choosing randomly (or not following any hierarchical reasoning procedure) being easily identified as having high rationality bounds. A different family of benchmark games, the ring games, while satisfying lower consistency, do not satisfy absence of framing. Again, this causes potential overestimation concerns, related this time to the identification procedure altering the object of identification itself due to framing.

Besides our two axioms, and attending to simplicity of implementability, we will focus on the characterization of games satisfying two additional requirements. First, we restrict our attention to games that can test for bound up to 4, since from the experimental literature the first four levels seem to be the empirically relevant ones. Second, we order games following *minimality* criteria according to which, all things equal, we favor the lowest possible number of players, player types, actions per player and directed links. Minimality may be important for resources-saving empirical implementations but also to minimize the complexity of the game to avoid as much as possible artificially generating noise in the data.

The following proposition shows that the only games enhancing the external validity of the identification (according to lower order consistency and absence of framing) that satisfy the aforementioned simplicity criteria are dominance solvable e-ring games with depth 4 and 2 actions per player:

Proposition 1 *Let $\mathcal{G}$ be a game. Then, $\mathcal{G}$ is minimal within the class of games that can test for bound 4 and satisfy lower order consistency and absence of framing if and only if $\mathcal{G}$ is a dominance solvable e-ring game of depth* 4 *with 2 actions per player.*

Proof. The 'if' part is immediate (simply notice that such e-ring games have a graph as the one depicted on the left of Figure 4 and are clearly minimal) so we focus on the 'only if' one. Lemma 1 implies that $\mathcal{G}$ is dominance solvable, contains player types x_1, x_2, x_3 and x_4 and has a set of links containing (x_4, x_3), (x_3, x_2) and (x_2, x_1). Also, by definition, there is no link starting from x_1. Minimality allows for excluding the presence of further player types and ensures that there are only two players, so that x_1 and x_3 must belong to one player and x_2 and x_4, to the other—the directed links whose existence we previously concluded precludes any other configuration. This excludes the presence of links (x_3, x_1), (x_2, x_4) and (x_4, x_2). Given this, absence of framing implies the presence of links (x_3, x_4) and (x_2, x_3). Finally, minimality excludes the presence of link (x_4, x_1). We are thus left with the graph depicted on the left of Figure 4, which corresponds to the graph of a dominance solvable e-ring game of depth 4. ■

The next section shows how these games were implemented in the experiment to test the properties proposed.

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 distribution of orders of rationality identified by the e-ring games with the ones identified by 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 and hence test the importance of the properties proposed. 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 2.2, 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.

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.[17] 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 classified as a rational type by random chance since there is 1.2% probability of being identified as $R1$-$R4$ while playing randomly in either game.[18]

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.

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). The importance of this decision is twofold. First, very sophisticated subjects should be *less* influenced by the structure of the game in their reasoning process, hence making the test of the axioms stricter. Second, if such a particular pool of subjects showed bounds in their hierarchical reasoning, then this would cast a stronger doubt on the underlying assumption in economic modeling that individuals are unbounded in their reasoning process.

[17] In the implementation we decided to have 4 actions for each player type in both classes of games for the following two reasons. The first one is that with only two actions per player type in the e-ring games, the unique action of level l for each player type x_l would be risk dominant, thus bringing new potential concerns in the identification. This means that at least three actions were needed. The second one is that, to avoid assuming that the subjects maximize expected utility in the e-ring games, we needed to have strict dominance to test for each bound, hence making it necessary to have at least one dominated action for each player type. However, to ensure comparability of the choice data, given that the ring games have three undominated actions for each player type, we added a strictly dominated action to the ring games and an undominated to e-ring games. Thus, all games have 4 actions.

[18] 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%).

All undergraduate engineering students from the school were sent an email message announcing two experimental sessions and they were confirmed on a first-come first-served basis. 229 students participated. No subject participated in more than one session. Subjects made all decisions using a booklet including all instructions stapled 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.[19]

Participants received no feedback, neither after playing each of the games nor after finishing each of the tasks, and were monitored such that they would not move from one task to another 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 110 minutes each.

We provided high monetary incentives for 10 randomly selected participants, instead of paying all subjects a lower amount of money.[20] 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.

4.3 Experimental Results

We start with the revealed rationality approach, whereby the choices made by the individual in a given class of games determine an upper bound for the level of higher-order rationality of that individual. The key question we address is whether that upper bound is also a good lower bound for the level of rationality of the individual. We claim that both Property 1 and Property 2, each contribute in different ways towards reducing the gap between the upper and

[19] 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.

[20] 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.

the lower bound. Next we provide some evidence in favor of such a claim.

Experimental Evidence for Property 1 (Lower Order Consistency). Games that satisfy lower order consistency and identify an individual as being of level $k \geq 2$ ensure that such individual makes choices that are consistent with level k also in decisions that test for levels $\ell = 1, \dots, k-1$ within the same game. In other words, games satisfying lower order consistency allow for the application of the revealed rationality principle at each step of the hierarchy of beliefs from level 1 up to level k within the same game.

To check that the requirement has bite, we compare the classification of individuals' levels of rationality obtained using the e-ring and ring games, which do satisfy lower order consistency (*LOC* games), with the levels obtained with the 4×4 and the beauty contest games that do not satisfy lower order consistency (non-*LOC* games). To see that the 4×4 and the two beauty contest games are not as good at accurately identifying higher order levels $R2$, $R3$, and $R4$ when identifying individuals as such, we look at two tests.

Test 1.1. First, we take the identification of subjects as being $R0$ or $R1$ by *LOC* games as valid, and look at how many of these subjects are *misclassified* as being $R2$, $R3$, or $R4$ by the non-LOC games. Second, we take the identification of subjects as being $R0$ or $R1$ by non-*LOC* games as valid, and look at how many of these subjects are *misclassified* as being $R2$, $R3$, or $R4$ by the *LOC* games.

Consider first all subjects that are identified as being of level $R0$ or $R1$ in the e-ring and ring games (52 subjects). The share of these subjects that are also identified as being of levels $R2$, $R3$ or $R4$ in the 4×4 and in the two beauty contest games are as follows (where the numbers in parenthesis give the shares out of all the 73 subjects that have been revealed as being of level at most $R0$ or $R1$ at least twice in the e-ring and ring games):

$$4 \times 4\text{: } 63.5\%\,(57.5\%) \quad \text{2/3-BC: } 84.6\%\,(89.0\%) \quad \text{1/3-BC: } 38.5\%(43.8\%).$$

Next, for comparison, consider all subjects that are identified as being of level $R0$ or $R1$ in at least two games of the 4×4 and the two beauty contest games (48 subjects).[21] We calculate the share of these subjects who are also identified as being of levels $R2$, $R3$ or $R4$ in e-ring or ring games. This leads to the following shares (the numbers in parenthesis give the shares out of all subjects that have been revealed being of level at most $R0$ or $R1$ at least twice, that is in at least two decisions among all relevant decisions in the 4×4 and the two beauty contest games (50 subjects)):

$$\text{E-ring games: } 35.4\%\,(38.0\%) \quad \text{Ring games: } 43.8\%\,(44.9\%).$$

[21] We do not consider all three games because the number of subjects satisfying this very strict condition is too small to make statistically significant comparisons.

The numbers show that for individuals that have been classified as not having higher order beliefs, the non-LOC games, with the exception 1/3-BC games, are significantly more likely to misclassify those individuals as having higher order beliefs, than the e-ring and ring games that are LOC games.

Test 1.2. First, we take subjects identified as being $R2$, $R3$ or $R4$ in each of the non-LOC games, and look at how many of these subjects are classified as being $R0$ or $R1$ by the LOC games. Second, we take subjects identified as being $R2$, $R3$ or $R4$ in each of the LOC games, and look at how many of these subjects are classified as being $R0$ or $R1$ by the non-LOC games.

Consider all subjects that are identified as being of level $R2$, $R3$ or $R4$ in the 4×4 and in the two beauty contest games (respectively, 164, 207 and 118 subjects). For each of these three populations separately, we calculate the share of individuals who are also identified as being of level $R0$ or $R1$ in the e-ring and ring games. We obtain the following shares (where the numbers in parenthesis give the shares of subjects that have been revealed as being of level at most $R0$ or $R1$ at least twice in the e-ring and ring games):

$$4 \times 4\text{: } 20.1\%\,(25.6\%) \quad \text{2/3-BC: } 21.3\%\,(31.4\%) \quad \text{1/3-BC: } 16.9\%\,(27.1\%).$$

Next, for comparison, consider all subjects that are identified as being of level $R2$, $R3$ or $R4$ in the e-ring game and then in the ring games (respectively, 139 and 116 subjects). For each of these two populations separately, we calculate the share of individuals who are also identified as being of level $R0$ or $R1$ in at least two games of the 4×4 and the two beauty contest games. We obtain the following shares (where the numbers in parenthesis give the shares out of all subjects revealed as being of level at most $R0$ or $R1$ at least twice in the 4×4 and the two beauty contest games):

$$\text{E-ring games: } 12.2\%\,(13.7\%) \quad \text{Ring games: } 18.1\%\,(15.8\%).$$

The numbers show that for individuals that have been classified as having higher order beliefs by non-LOC games, there are significantly more that are then classified as not having higher order beliefs by LOC games than the other way around. Again, the 1/3-BC seems to be an exception.

In both tests we observe that games that do not satisfy Property 1 tend to be nosier in the identification of higher orders of rationality, potentially reducing the external validity of the identified distribution.

Experimental Evidence for Property 2 (Absence of Framing). Next, we build on the established empirical relevance of Property 1 to check the importance of requiring absence of framing (Property 2). Again, we consider two tests.

Test 2.1. We consider all subjects that are identified as being of level $R0$ or $R1$ in at

	$R4$	$R3$	$R2$
E-ring game	8.3% (10.0%)	16.7% (18.0%)	35.4% (38.0%)
Ring game	20.8% (22.0%)	25.0% (26.0%)	43.8% (44.0%)

Table 2: Cumulative distribution of higher-order rationality levels for e-ring and ring games for subjects identified as being of level $R0$ or $R1$ in at least two games of the 4×4 and the two beauty contest games (48 subjects) (in parenthesis for subjects revealed as being of level at most $R0$ or $R1$ at least twice in the 4×4 and the two beauty contest games (50 subjects)).

least two games of the 4×4 and the two beauty contest games (48 subjects) (or alternatively, revealed as being of level at most $R0$ or $R1$ at least twice in the 4×4 and the two beauty contest games). Such 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 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 2 presents the cumulative distribution function of the rationality levels as classified by the e-ring and ring 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 2, the distribution of levels identified by the ring games first order stochastically dominates the one identified by the e-ring games (significant at the 1% level using the Kolmogorov-Smirnov test in both cases).

Test 2.2. We find further evidence of the relevance of Property 2 when comparing treatments in which the ring games and the e-ring games were presented in different orders to subjects, we find generally higher levels of rationality in the e-ring games when they are played after having played the ring games (126 subjects), than when played in the opposite order (103 subjects). We find the average identified level by the e-ring game increases by 9.8%. Also, the Kolmogorov-Smirnov test is significant at the 1% level.[22]

Both tests suggest that Property 2 reduces misclassification of subjects as satisfying higher order rationality.

Finally, we report the empirical correlation between the orders of rationality identified by the various games and the results of the standardized tests used for admittance to university in Spain. We find that it is highest for the e-ring games among all the classes of games used. We view this as potential further evidence that e-ring games, as the only games satisfying both

[22] When looking at the levels identified by the ring games, we find slightly lower levels of rationality when the ring games are played after having played the e-ring games, than when they are played beforehand. The average level identified by the ring game decreases by 3.5% when the ring games are played after the e-ring games.

properties, are less noisy in identifying higher order rationality. These correlations are:

E-ring games: 0.24 Ring games: 0.12 4×4: 0.06 2/3-BC: 0.16 1/3-BC: 0.08.

Notice that while, statistically, e-ring games may tendentially outperform the 4×4 and the beauty contest 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 as they both have 8 choices.

5 Conclusion

The identification of a reliable distribution of orders of rationality in the population is a crucial prerequisite for predicting behavior in many applications, including price formation and oligopolistic competition, mechanism and institutional design or monetary policy. This identification is a highly problematic exercise. A fundamental issue, addressed here for the first time, is that standard games used so far do not allow for the observation of behavior at the different steps of the hierarchy of beliefs and, when they do, they might frame individuals into thinking in levels, thereby compromising the very exercise.

This paper tackles this apparent contradiction by taking an axiomatic approach. Two intuitive properties are introduced that, at a practical level, narrow down the class of games that can be used for identification to a single class: the *e-ring games*. The empirical evidence presented suggests that both properties are relevant in reducing the misidentification. As a result, e-ring games might constitute a useful starting point for the study of higher order rationality.

The introduction of the axiomatic approach in this literature might be important per se, by enabling a more transparent discussion of what features of the game enhance the external validity of the identification. On one hand, the axioms make explicit what is required of the games to be used. On the other, by being stated explicitly, the axioms can be tested, discarded, and alternatives can be thought of, thus pushing the discussion in the literature forward in a more structured way.

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.*

Angeletos, George-Marios and Chen Lian (2016). Incomplete information in macroeconomics: Accommodating frictions in coordination. In John Taylor and Harald Uhlig (editors), *Handbook of Macroeconomics, Vol. 2*, pp. 1065–1240. Elsevier.

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.

Börgers, Tilman and Jiangtao Li (2019). Strategically simple mechanisms. *Econometrica*, **87**(6), 2003–2035.

Brandenburger, Adam, Alex Danieli and Amanda Friendenberg (2017). How many levels do players reason? an observational challenge and solution. *Mimeo.*

Burchardi, Konrad B. and Stefan P. Penczynski (2014). Out of your mind: Eliciting individual reasoning in one shot games. *Games and Economic Behavior*, **84**, 39–57.

Costa-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.

Crawford, Vincent P. (2016). Efficient machanisms for level-k bilateral trading. *Mimeo.*

Cubitt, Robin and Robert Sugden (1994). Rationally justifiable play and the theory of non-cooperative games. *Economic Journal*, **104**(425), 798–803.

De Clippel, Geoffroy, Rene Saran and Roberto Serrano (2019). Level-k mechanism design. *Review of Economic Studies*, **86**(3), 1207–1227.

Dekel, Eddie, Drew Fudenberg and Stephen Morris (2007). Interim correlated rationalizability. *Theoretical Economics*, **2**, 15–40.

García-Schmidt, Mariana and Michael Woodford (2019). Are low interest rates deflationary? a paradox of perfect-foresight analysis. *American Economic Review*, **109**(1), 86–120.

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

Germano, Fabrizio, Jonathan Weinstein and Peio Zuazo-Garin (2020). Uncertain rationality, depth of reasoning and robustness in games with incomplete information. *Theoretical Economics*, **15**(1), 89–122.

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.

Pearce, David G. (1984). Rationalizable strategic behavior and the problem of perfection. *Econometrica*, **52**, 1029–1050.

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 der Linden, martin (2018). Bounded rationality and the choice of jury selection. *Journal of Law and Economics*, **61**(711–738).

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.