# Higher Orders of Rationality and the Structure of Games*

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

[1]Universitat Pompeu Fabra and BSE
[2]Universitat Ramon Llull, ESADE
[3]HSE University, International College of Economics and Finance

February 4, 2022

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 since games may significantly impact and hence invalidate the identified order. To tackle this fundamental problem, we propose an axiomatic approach that formalizes some of the key difficulties in an explicit manner. We then introduce a probability space to study under which conditions the proposed axioms are necessary for efficient identification. The axioms single out a new class of games, the *e-ring games*, that we use in a within subject experiment to compare individuals' classifications with the ones obtained in standard games previously used in the literature. The results show that our theoretical approach is empirically feasible and a first step towards a more reliable identification.

**Keywords**: Rationality, Higher-Order Rationality, Revealed Rationality, Hierarchic Reasoning
*JEL Classification*: C70, C72, C91, D01, D80.

## 1 Introduction

Many fields of economics 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

*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.

and Serrano, 2019) and jury selection (Van der Linden, 2018), among many others. In macroeconomics, for instance, understanding the existence and bounds on agents' higher-order beliefs has attracted a lot of attention recently, given that they can lead to rather different economic implications (García-Schmidt and Woodford, 2019; Coibion, Gorodnichenko, Kumar and Ryngaert, 2021). 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. The empirical identification of the relevant reasoning bounds is then crucial to accurately predict strategic behavior and optimally design institutions.

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 reasoning consistent with their choices.[1] Here, we rely on this literature and focus on the *revealed rationality* method which is straightforward.[2] First, an analyst observes a subject's choices in different roles in a game. Then the analyst computes how many rounds $k$ of iterated deletion of strictly dominated actions each of these choices survives. Given this, the rationality (upper) bound the analyst estimates is equal to the *lowest* $k$ among the computed ones—no bound is estimated if every choice survives the full process of iterated deletion of strictly dominated actions. That is, 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. Players are assigned 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).Crucially, given that such choices are made in a particular game, the structure of the game can influence and possibly bias the identification exercise.[3]

This paper provides a comprehensive analysis of the impact of the properties of a game on the validity of the revealed rationality method. It highlights a tight theoretical dependency between the statistical properties of the distribution of rationality bounds in a population, and the effect of the structure of a game on the estimation. By adopting an axiomatic approach, we discuss two natural properties on the payoff dependency of a game and identify conditions on the distribution on the rationality bounds under which the two properties are necessary for a valid estimation of the rationality bounds. The experimental evidence we find supports

[1]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.

[2]Notice that we focus on rationality levels just for the importance rational beliefs have in the economic literature. The whole analysis developed in the paper can be applied to any hierarchical model of thinking, like k-levels for example.

[3]Alaoui and Penta (2016) already recognize the possibility that different payoffs can influence the depth of reasoning. Here we complement their approach by saying that the structure of the game itself can influence the reasoning process and its identification.

the theoretical result. The main purpose of the exercise though, is to introduce a common formal language to allow for a structured and explicit discussion of what are the desirable properties of games (possibly beyond our own suggestions) and what they would imply for the identification.

To start shaping the discussion, consider the following two-player bimatrix game where the left matrix describes the payoffs of Player 1 (with actions $A$, $B$, $C$) and the right matrix describes the payoff of Player 2 (with actions $a$, $b$, $c$).

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

**Player 1**

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

**Player 2**

For Player 2, action $a$ is the only one surviving three rounds of iterated deletion of strictly dominated actions and for Player 1, action $C$ is the only one surviving four rounds of deletion. As it is standard in the literature, a subject choosing actions $C$ as player 1 would be classified as reaching four orders of rationality. Similarly, a subject choosing action $a$ as player 2 would be classified as reaching no less than three orders of rationality. Nonetheless, a strategy consistent with a certain order of rationality $k$ may be chosen for a variety of reasons beyond reasoning according to a particular order of rationality, raising the possibility of an identification mistake. Given that we do not observe a subject's reasoning process, we cannot exclude that choice $C$ in the role of player 1, for example, has been taken for other reasons such as choosing the strategy labeled with the first letter of a subject's surname. A possibility then would be to make the same subject play across different games and see whether her behavior is consistent with the initial classification. This would raise a major theoretical concern though. If the structure of games influences the estimation, looking at behavior across differing games without taking this channel into account could make the estimation invalid. Ideally, we would need to identify the bounds in one externally valid game.

With this concern in mind, we propose two requirements for the structure of the games employed in the estimation: (1) that behavior at each step of the hierarchy of beliefs is observed *within* the same game, not via different games with differing depths to test for subject's bounds, and (2) that the structure imposed by (1) does not *induce* hierarchical thinking. Our first requirement then is to assure that the structure of the game allows for the observation of behavior at the different steps of the hierarchy of beliefs. In this way, the subject would be classified as being of order $k$ *only if* her behavior when playing in a role to test for level $\ell = k$ is consistent with such a classification, and that when playing in a role to test for $\ell = 1$, $\ell = 2$, all the way up to $\ell = k - 1$, the choices are also all consistent with the classification $k$, at each step of the hierarchy from 1 to $k - 1$, and, importantly, all *within the same game*.

A four-player *ring game* as used in Kneeland (2015) already satisfies this requirement:

| | $d$ | $e$ | $f$ |
|---|---|---|---|
| $a$ | 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 this game is that Player $\ell$'s payoffs depend only on her own choice and on that of Player $\ell + 1$, up to Player 4 whose payoffs depend on her own choice and on that of Player 1. In addition, the game is dominance solvable in four steps: action $j$ is strictly dominant for Player 4, which implies that action $g$ is the only one surviving two round of deletion for Player 3, action $e$ is the only one surviving three rounds for Player 2, and action $a$ is the only one surviving four rounds for Player 1. In consequence, because of the structure of the game, the conclusion that a subject who chose $a$ in the role of Player 1 has four orders of rationality is falsified if the same subject is observed not to choose actions $e$, $g$ and $j$ when in the roles of players 2, 3 and 4, respectively. The key idea is that a subject incapable of forming beliefs of order higher than the first ($R2$), would eventually make a *mistake* when playing in the role of players 1 and 2.

Notice however that 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. For example, an R2 player playing as Player 2, given the game asks her to focus on Player 3 and 4's behavior, might be pushed into behaving as if R3. In fact, the game does not allow for an alternative structure of higher-order beliefs than the one that makes the game dominance solvable. Her strategic thinking cannot take *paths* different from the one creating dominance solvability. The structure of the game maps one-to-one into the structure of higher-order beliefs that is needed to solve the game. This motivates requirement (2) above, capturing a novel and intuitive notion of framing, which we formally define in the paper.

In Section 2, we introduce a probabilistic setting that models the problem of estimating rationality bounds. In fact, the importance of the proposed requirements cannot be grasped without a formal definition of what is the objective of the estimation. We define as the estimation error the difference between the rationality level estimated from observed behavior and the *true one* generating such behavior, that is, the object of estimation. We show that the error can be decomposed in two terms. One denotes the distortion that can arise due to not observing the whole hierarchy of thinking of the subject, and so it is related to requirement (1). The other one due to the inductive structure of payoff dependencies and hence related to requirement (2).

In Section 3, we formalize requirements (1) and (2) as two separate properties of a given game. The first property, *lower-order consistency*, ensures that individual behavior can be tested at *each step* of the hierarchy of beliefs, as in requirement (1) above. The second one, *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 language of graphs and guarantees that the payoff dependencies of the game do not correspond exactly with the "natural" hierarchy of beliefs, thereby enabling players to contemplate alternative hierarchies.[4]

Proposition 1 shows that under some clearly stated conditions regarding the probabilistic framework defined above, lower-order consistency and absence of framing are *necessary* for the estimator, that is, the game used to estimate the true rationality bound, to be *efficient*. Any game not satisfying the two properties will have an estimation error that first order stochastically dominates the one obtained from a game that does.

Thus, proving the existence of a class of games that satisfies lower-order consistency and absence of framing becomes of primary importance for Proposition 1 not to be empty. It turns out, as explained in Section 4, that the two properties 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*, inspired by Kneeland (2015).[5] 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.

In Section 5 we then test the validity of the two properties proposed by comparing behavior across the most prominent classes of games used in the literature. Specifically, we identify levels of rationality for each class of games and the new *e-ring games*. Even if not all the games we consider have been used to identify levels of rationality, they are all considered standard games to assess hierarchies of beliefs, which is the more general problem this paper addresses. We focus on rationality for two main reasons. First, the absence of sophistication that is, level 0 is clearly specified and easily comparable across games. Second, the rationality assumption is still the central one in economics and thus is of crucial importance to identify the distribution of rationality bounds in the population to make better predictions as emphasized by Kneeland (2015).

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

[4]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.

[5]See Section 4 for a justification of why simplicity of the games may be a desirable property in empirical applications.

in Kneeland (2015), two simple two-player 4×4 dominance solvable games, and three different versions of the beauty contest game presented in Nagel (1995).[6] We thus compare behavior across games which satisfy both properties (e-ring games), one of them (ring games) and none of them (4×4 and the beauty contest games) allowing us to test whether satisfying the properties is effective empirically in addressing the theoretical concerns raised above.

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 frame subjects into higher 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.

Finally, Section 6 concludes. The Appendix contains the proofs of the mathematical results, and the Online Appendix contains the distribution of oreders of rationality across the different games, the English translation of the experimental instructions and the payoff matrices of all games used in the experiment.

## 2 Framework

In this section we introduce the tools that allow for the formal analysis of the revealed rationality approach. Section 2.1 recalls standard game-theoretic preliminaries that relate the iterated deletion of strictly dominated actions with higher-order beliefs in rationality. For reasons related to simplicity of experimental implementation and discussed in a later section, we allow for settings with incomplete information, but assume that the information structure (the type space of the game) is exogenously specified by the analyst. Section 2.2 describes in detail the estimation procedure, the revealed rationality method, standard in the literature. Finally, for those readers concerned with deeper conceptual and formal issues, Section 2.3 provides a rigorous formalization of the object of estimation and the error in the estimation, and introduces a notion of *efficient* estimation (based on first-order stochastic dominance). These notions of error and efficiency will provide the main criterion to evaluate the validity of a given

[6] To be more specific, subjects play three versions of the beauty contest game where the average of all subjects' responses is multiplied by 1/3 and 2/3 in the first two versions, and the third one consists in a *p-beauty contest game* where the multiplying factor is an unspecified number $p \in (0, 1)$, assumed to be commonly known, and where subjects have to specify how they would play for any $p$ in the interval. Notice also that, throughout the paper, for the analysis of beauty contest games (and only for this class of games), we consider that at each step it is weakly dominated actions that are eliminated, instead of strictly dominated ones. This reinforces our message, since with strict domination, and the possibility of choosing nonintegers, the orders of rationality explode: 0 for 100, and $\infty$ for every choice strictly smaller than 100.

game as a tool for estimation.

### 2.1 Games and Finite-Order Rationalizability

A *game* is 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 $i$, $T_i$ is a finite set of *types*, $A_i$ is a finite set of *actions*, $u_i : \prod_{j \in I}(T_j \times A_j) \to \mathbb{R}$ is a *utility function*, and $\pi_i : T_i \to \Delta(\prod_{j \neq i} T_j)$ is a *belief function*.[7] As usual, $T := \prod_{i \in I} T_i$ and $A := \prod_{i \in I} A_i$ denote the sets of type and action *profiles*, respectively, and, for each player $i$, $T_{-i} := \prod_{j \neq i} T_j$ and $A_{-i} := \prod_{j \neq i} A_j$ denote the sets of $i$'s opponents' type and action profiles, respectively. We say that game $\mathcal{G}$ has *complete* (resp. *incomplete*) information if $T$ is a singleton (resp. not a singleton).

The solution concept that captures the idea of iterated deletion of strictly dominated actions is (*interim correlated*) *rationalizability*,[8] and the link between these two notions constitutes the conceptual foundation of the identification strategy in the next section. Before proceeding formally, the iterated definition of rationalizability can be sketched as follows:[9]

- Action $a_i$ is *1st-order rationalizable* for type $t_i$ if there exists a conjecture about $i$'s opponents' types and actions (a probability function on $T_{-i} \times A_{-i}$) such that: (1) $a_i$ maximizes the resulting expected utility, and (2) the conjecture is consistent with the beliefs about $T_{-i}$ specified by $\pi_i(t_i)$. This is equivalent to $a_i$ *not* being strictly dominated for $t_i$.

- For order $k \geq 2$, action $a_i$ is *$k$th-order rationalizable* for type $t_i$ if there exists a conjecture about $i$'s opponents' types and actions such that: (1) $a_i$ maximizes the resulting expected utility, (2) the conjecture is consistent with the beliefs about $T_{-i}$ specified by $\pi_i(t_i)$, *and* (3) the conjecture *believes* opponents to play $(k-1)$th-order rationalizable actions (if a pair $(t_j, a_j)$ gets positive probability, then $a_j$ is $(k-1)$th-order rationalizable for $t_j$). This is equivalent to $a_i$ surviving $k$ rounds of iterated deletion of strictly dominated actions from $t_i$'s perspective.

Given this, an action is rationalizable for type $t_i$ if it is $k$th-order rationalizable for $t_i$, for every $k \geq 1$. Thus, formally, the set of *rationalizable* actions for type $t_i$ is defined as $R_i(t_i) := \bigcap_{k \geq 0} R_{i,k}(t_i)$, where the set of 0-*th order rationalizable* actions of type $t_i$ is simply $R_{i,0}(t_i) := A_i$ and, for each $k \geq 1$, the set of *k-th order rationalizable* actions of type $t_i$ is inductively defined

[7]Following usual conventions, for each finite set $S$, we let $\Delta(S)$ denote the set of probability measures on the power set of $S$.

[8]The duality between the iterated deletion of strictly dominated actions and rationalizability is a rather straightforward corollary of the classic Wald-Pearce Lemma. For details about the case of complete information, see Pearce (1984) and Tan and Werlang (1988); for the case of incomplete information, Dekel, Fudenberg and Morris (2007) and Battigalli, Di Tillio, Grillo and Penta (2011).

[9]It is important to stress that the evaluation is carried out at the *interim* level, that is, from the perspective of each type $t_i$. That a conjecture $\mu_i \in \Delta(T_{-i} \times A_{-i})$ is consistent with $t_i$ simply means that its marginal on $T_{-i}$ coincides with $\pi_i(t_i)$.

as follows:

$$R_{i,k}(t_i) := \left\{ a_i \in A_i \;\middle|\; \begin{array}{ll} \multicolumn{2}{l}{\text{There exists some } \mu_i \in \Delta(T_{-i} \times A_{-i}) \text{ such that:}} \\ (1) & a_i \in \arg\max_{a'_i \in A_i} \sum_{t_{-i} \in T_{-i}} \sum_{a_{-i} \in A_{-i}} \mu_i[(t_{-i}, a_{-i})] u_i((t_{-i}, t_i), (a_{-i}, a'_i)) \\ (2) & \mathrm{marg}_{T_{-i}} \mu_i = \pi_i(t_i) \\ (3) & \mu_i \left[ \left\{ (t_{-i}, a_{-i}) \in T_{-i} \times A_{-i} \;\middle|\; a_{-i} \in \prod_{j \neq i} R_{j,k-1}(t_j) \right\} \right] = 1 \end{array} \right\}.$$

Finally, let us introduce some terminology for expositional purposes. Throughout the paper, we refer to the different roles within a game as *player-types* and, as usual, we say that a game is *dominance solvable* if the iterated deletion of strictly dominated actions always yields a unique prediction:

**Definition 1** (Player-types, depth of a player-type, depth of a game). *Let $\mathcal{G}$ be a game. Then, a* player-type *in $\mathcal{G}$ is a pair $x = (i, t_i)$ where $i \in I$ and $t_i \in T_i$. Player-type $x = (i, t_i)$ has* depth $k \geq 1$ *if $R_{i,k-1}(t_i) \supsetneq R_{i,k}(t_i) = R_i(t_i)$, and that $\mathcal{G}$ has* depth $n \geq 1$ *if $n$ is the maximum depth of the player-types in $\mathcal{G}$. Let $X_{\mathcal{G}}$ denote the set of player-types in $\mathcal{G}$ and, for each $k \geq 1$, let $x_k$ denote a player-type of depth $k$.*

**Definition 2** (Dominance solvable game). *Let $\mathcal{G}$ be a game. Then, $\mathcal{G}$ is* dominance solvable *if $R_i(t_i)$ is a singleton for every $i \in I$ and every $t_i \in T_i$.*

## 2.2 Estimation of Rationality Bounds

We now formalize the estimation procedure briefly sketched in the beginning of the section. The setting consists in a probability space $(\Omega, \mathcal{F}, P)$, where $\Omega$ is a population of *subjects* and $P(E)$ represents the *probability* of each *event* $E \in \mathcal{F}$, *not necessarily known* to the analyst. For obvious reasons, throughout the paper we exclusively focus on finite $\Omega$ and $\mathcal{F} = 2^{\Omega}$. The estimation procedure, commonly referred to in the literature as the *revealed rationality approach*, can then be described in four steps:

1. The analyst fixes a game $\mathcal{G}$.

2. Each subject $\omega \in \Omega$ chooses an action in each role $x \in X_{\mathcal{G}}$. Thus, random variable $\hat{a}_x : \Omega \to A_i$ represents the choice of each subject in the role of player-type $x = (i, t_i)$. The description of choices $(\hat{a}_x)_{x \in x_{\mathcal{G}}}$ is the choice-data, or database, *observable* to the analyst.

3. If subject $\omega$ chooses an action $\hat{a}_x(\omega) \in R_{i,k}(t_i) \setminus R_{i,k+1}(t_i)$ in the role of some player-type $x = (i, t_i)$, the analyst interprets that the subject performs *less* than $k + 1$ rounds of

iterated elimination of strictly dominated actions (see the discussion in Section 2.1) and hence estimates $k$ as the rationality bound in role $x$. If $\hat{a}_x(\omega) \in R_i(t_i)$, no rationality bound is estimated for role $x$. Thus, the random variable $\hat{r}_x : \Omega \to \mathbb{N} \cup \{0, \infty\}$ where, for each $\omega \in \Omega$,

$$\hat{r}_x(\omega) := \begin{cases} \min\{\ell \geq 0 \,|\, \hat{a}_x(\omega) \notin R_{i,\ell+1}(t_i)\} & \text{if } \hat{a}_x(\omega) \notin R_i(t_i), \\ \infty & \text{otherwise,} \end{cases}$$

represents the subjects' *estimated* rationality bounds given choice-data $\hat{a}_x$.

4. Finally, the estimated rationality bound for each subject is equal to the minimum rationality bound estimated for the subject in each role $x \in X_{\mathcal{G}}$.[10] Thus, random variable $\hat{r}_{\mathcal{G}} : \Omega \to \mathbb{N} \cup \{0, \infty\}$ where, for each $\omega \in \Omega$,

$$\hat{r}_{\mathcal{G}}(\omega) := \begin{cases} \min\{\hat{r}_x(\omega) \mid x \in X_{\mathcal{G}}\} & \text{if } \hat{a}_x(\omega) \notin R_i(t_i) \text{ for some } x \in X_{\mathcal{G}}, \\ \infty & \text{otherwise,} \end{cases}$$

represents the subjects' *estimated* rationality bounds given choice data $(\hat{a}_x)_{x \in X_{\mathcal{G}}}$.

These steps show how to a employ a game $\mathcal{G}$ to easily construct an estimator $\hat{r}_{\mathcal{G}}(\omega)$ for each subject $\omega$. While intuitive, the method described is not explicit about the underlying model of the subject's behavior, and thus, lack a proper formalization of what the variable $\hat{r}_{\mathcal{G}}$ is trying to estimate. In consequence, it is impossible to assess the latter's validity as an estimator. The following section tackles this issue.

## 2.3 Formalization of the Estimation Error

### 2.3.1 Rationality Orders and Lower Rationality Bound

The revealed rationality method detailed in the previous section relies on the *as if* assumption that agents' choices can be explained by a possibly finite-order version of rationalizability entailing some coherency across player-types. Let us elaborate on this. First, for each subject $\omega$ there exists some $r_x(\omega)$ that describes the *true* rounds of deletion that $\omega$ performs in the role of player-type $x$ before choosing her action, $\hat{a}_x(\omega)$ (for simplicity, this number of rounds cannot be higher than the depth of the player-type). Second, these values $(r_x(\omega))_{x \in X_{\mathcal{G}}}$ are *coherent* in the following sense: A subject $\omega$ that performs $k$ rounds in some role, also performs the maximum possible rounds in the role of any player-type of depth lower than $k$, and at least $k$ rounds in the role of every player-type of depth higher than $k$. These two ideas are formalized, in reverse order, in the following definition:

[10] With some abuse of notation, for the sake of computation of a minimum we treat the symbol representing "no estimated bound," $\infty$, as an element larger than any $m \in \mathbb{N} \cup \{0\}$.

**Definition 3** (List of Rationality Orders)**.** *Let* $(\Omega, \mathcal{F}, P)$ *be a probability space and* $\mathcal{G}$*, a game. Then:*

(*i*) *A list of random variables* $(r_x)_{x \in X_{\mathcal{G}}}$ *is a* list of rationality orders *if, for any player-type* $x$ *of depth* $k_x$ *and any player-type* $x'$ *of depth higher than* $k_x$*, the following hold for every subject* $\omega$*,*

$$r_x(\omega) \in \{0, \dots, k_x\}, \quad \text{and} \quad r_{x'}(\omega) \geq k_x \Longleftrightarrow r_x(\omega) = k_x.$$

(*ii*) *A list of rationality orders* $(r_x)_{x \in X_{\mathcal{G}}}$ *is* consistent *with choice-data* $(\hat{a}_x)_{x \in X_{\mathcal{G}}}$ *if for any player-type* $x = (i, t_i)$*, the following holds for every subject* $\omega$*,*

$$\hat{a}_x(\omega) \in R_{i, r_x(\omega)}(t_i).$$

**Remark 1.** *For any game and choice-data* $(\hat{a}_x)_{x \in X_{\mathcal{G}}}$*, there* always exists at least one *list of rationality orders* $(r_x)_{x \in X_{\mathcal{G}}}$ *consistent with it, namely the trivial one given by constants* $r_x \equiv 0$*. Typically, every choice-data* $(\hat{a}_x)_{x \in X_{\mathcal{G}}}$ *is consistent with multiple, different lists of rationality orders.*

Finally, we turn to the object of estimation of the paper. We assume the existence of a random variable $r : \Omega \to \mathbb{N} \cup \{0, \infty\}$ that represents the *minimum* number of rounds of deletion that each subject is *ensured* to perform in the role of *every* player-type of *every* game. That is, consider a *comprehensive* collection of lists of rationality orders $\{(r_x)_{x \in \mathcal{G}} \mid \mathcal{G} \text{ is a game}\}$ that specifies the true number of rounds of deletion each subject would perform in the role of every player-type of every game. Then, for each subject $\omega$, each game $\mathcal{G}$ and each player-type $x \in X_{\mathcal{G}}$ with depth denoted by $k_x$, it must be the case that:

- If $k_x \leq r(\omega)$, then subject $\omega$ performs *exactly* $k_x$ rounds of deletion in the role of $x$.
- If $k_x > r(\omega)$, then subject $\omega$ performs *at least* $r(\omega)$ rounds of deletion in the role of $x$, such that $r_x(\omega) \geq r(\omega)$. The gap $r_x(\omega) - r(\omega)$ can be interpreted as representing the impact of the *context* on $\omega$'s reasoning process; in particular (as discussed below in Sections 3.2 and 3.3), in the role of player-types of games in which the inductive structure of dominance solvability is very transparent, $r_x(\omega)$ could be much higher than in other roles in which this structure is more opaque.

Thus, formally:

**Definition 4** (Lower Rationality Bound)**.** *Let* $(\Omega, \mathcal{F}, P)$ *be a probability space. Then:*

(*i*) *A* lower rationality bound *is a random variable* $r : \Omega \to \mathbb{N} \cup \{0, \infty\}$.

(*ii*) *A lower rationality bound* $r$ *is* consistent *with a comprehensive collection of lists of rationality orders* $\{(r_x)_{x \in \mathcal{G}} \mid \mathcal{G} \text{ is a game}\}$ *if for every subject* $\omega$*, every game* $\mathcal{G}$ *and every* *player-type $x \in X_{\mathcal{G}}$ with depth denoted by $k_x$,*

$$r_x(\omega) \geq \min\{r(\omega), k_x\}.$$

**Remark 2.** *Similarly as above, every comprehensive collection of lists of rationality orders is consistent with at least one lower rationality bound, namely, the constant $r \equiv 0$, and is typically consistent with multiple, different lower rationality bounds. It follows that every choice-data $(\hat{a}_x)_{x \in X_{\mathcal{G}}}$ is typically consistent with multiple lower rationality bounds, hence the identification problem.*

### 2.3.2 Estimation Error and Efficiency

Section 2.3.1 defined the object of estimation for the method described in Section 2.2, which relies on the initial specification of some game $\mathcal{G}$. As it is natural to expect, different games $\mathcal{G}$ may lead to different distributions of errors in the estimation of the subjects' true value of $r$.[11] Specifically, the random variable that formalizes the *estimation error* induced by a game $\mathcal{G}$ is:

$$e_{\mathcal{G}} := |\hat{r}_{\mathcal{G}} - r|.$$

This implies two things. First, if errors are independent across games, we would need to use a large number of games to assure that the bias in the estimation disappears. Second, if errors are not independent, for example, if they depend on the structure of the game as we argue here, then it is not possible to achieve a reliable estimation by using different games and it is crucial to use the same game as an estimator to keep things constant and be able to consider the error as actual noise. Therefore, a natural prerequisite for the analyst is to choose a game $\mathcal{G}$ that yields the minimum possible error according to some reasonable criterion of what *minimum* means. In our analysis, we will rely on a very weak comparison standard, based on first-order stochastic dominance, and we will say that game $\mathcal{G}$ procures a *more efficient* estimation than game $\mathcal{G}'$ if it ensures a lower probability of error (for some size of the error):

**Definition 5** (Efficiency)**.** *Let $(\Omega, \mathcal{F}, P)$ be a probability space and let $\mathcal{G}, \mathcal{G}'$ be two games of the same depth. Then:*

*(i) $\mathcal{G}$ is* more efficient than *$\mathcal{G}'$ if, for every $\varepsilon \geq 0$,*

$$P(e_{\mathcal{G}} > \varepsilon) < P(e_{\mathcal{G}'} > \varepsilon).$$

*(ii) $\mathcal{G}$ is* efficient *if no other game of the same depth is more efficient than $\mathcal{G}$.*

[11] See Georganas, Healy and Weber (2015) and Burchardi and Penczynski (2014) for evidence in this direction.

# 3 Two Properties for Efficient Estimation

The previous section equipped us with a formal notion of estimation error and with criteria to assess the relative efficiency for estimation of different games. We are then ready to tackle the first main question of the paper: how does the *structure* of a game relate to its efficiency for estimation? Or, more specifically: is it possible to identify properties on the payoff-dependencies of the game that are *necessary* for an efficient estimation?

To answer these questions, first, in Section 3.1 we borrow some terminology from graph theory to define the *graph structure* of a game. Next, in Section 3.2, we build on this graph structure to formalize two properties on the payoff-dependencies of a game, *lower-order consistency* and *absence of framing*. Following the intuition sketched in the examples of Section 1, we argue heuristically that these properties help reducing the probabilities of overestimating the bounds. For those readers interested in the statistical foundation of these informal arguments, Section 3.3 provides a rigorous formalization thereof, and presents the *sufficient* statistical conditions for estimation, under which lower-order consistency and absence of framing are proved to be *necessary* for it to be efficient. Specifically, we prove in Proposition 1 that, under the sufficient statistical conditions, if a game satisfies both lower-order consistency and absence of framing, then any game of the same depth that fails to satisfy either property *is not efficient.*

The question regarding the existence of a game (ideally, simple) that satisfies both properties is postponed to Section 4.

## 3.1 The Graph Structure of a Game

To better visualize the player-types and the 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 6** (Graph of a Game)**.** *Let $\mathcal{G}$ be a game. The, the (directed)* graph *of $\mathcal{G}$ consists in the pair $(X_{\mathcal{G}}, L_{\mathcal{G}})$, where the set of* nodes *is the set of player-types $X_{\mathcal{G}}$, and the set of* directed links *$L_{\mathcal{G}}$ comprises the pairs of nodes $(x, x')$ where the following two conditions hold:*

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

(*ii*) *$x$'s expected payoff can depend on the actions of $x'$.*

**Definition 7** (Path)**.** *Let $\mathcal{G}$ be a game. Then, a* path *in graph $(X_{\mathcal{G}}, L_{\mathcal{G}})$ is a finite sequence of nodes $(x^{(1)}, x^{(2)}, \dots, x^{(m)})$ where the following two conditions hold:*

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

(*ii*) *All the nodes except possibly $x^{(1)}$ and $x^{(m)}$ 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 $x'$ never affects $x$'s expected payoff. Higher-order payoff dependencies are captured by *paths*, which represent different possible belief-hierarchies that the first player-type in the path ($x^{(1)}$) can conceive when thinking strategically. We now illustrate these concepts by revisiting the example sketched in Section 1:

**Example 1** (Bimatrix Games vs. Ring Games)**.** Consider again two-player bimatrix game:

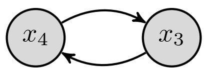
| | $a$ | $b$ | $c$ |
|---|---|---|---|
| $A$ | 80 | 20 | 140 |
| $B$ | 60 | 200 | 20 |
| $C$ | 100 | 160 | 40 |

**Player 1**

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

**Player 2**

Using Definition 1, players 1 and 2 have depths 4 and 3, respectively (action $C$ survives four rounds of iterated deletion for Player 1, and action $a$ survives 3 rounds for Player 2), and can thus be identified as player-types $x_4$ and $x_3$, respectively. Even if the game is dominance solvable (with solution $(C, a)$), there are no player-types $x_1$ and $x_2$. Thus, since no player has a dominant action, the graph structure of the game looks as follows:




As argued in the introduction, a subject capable of forming only first-order beliefs, playing as Player 1 and choosing randomly between $A$ and $C$, may have high chances of being classified as if capable of forming higher-order beliefs by playing, for example, $C$.

This problem does not arise in the four-player *ring games* employed by Kneeland (2015):[12]

| | $d$ | $e$ | $f$ |
|---|---|---|---|
| $a$ | 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**

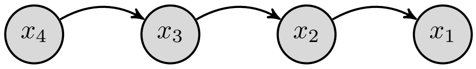
| | $a$ | $b$ | $c$ |
|---|---|---|---|
| $j$ | 120 | 160 | 140 |
| $k$ | 80 | 120 | 100 |
| $l$ | 60 | 100 | 80 |

**Player 4**

Notice that here, each player corresponds to a different player-type, Player 4 being player type $x_1$, Player 3 being $x_2$, Player 2 being $x_3$ and finally Player 1 being $x_4$. This feature allows for rejecting every possible bound 1, 2, 3 and 4 *within* the same ring-game, it is enough to

[12]Notice that these payoffs are the payoffs used in Kneeland (2015)'s G1 multiplied by 10.

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-types $x_3$ or $x_4$. 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. The following graph representation of the game makes this issue more directly apparent:




It is immediate to see that each player-type $x_k$ admits a *unique* path of length $k-1$. $\square$

The critiques to bimatrix and ring-rames outlined in the above discussion further underscore the importance of putting conditions on the structure of the game in order to adequately estimate subjects' rationality bounds. Fortunately, they also offer an intuition on how to solve this by changing this structure. We address this formally in the next section.

## 3.2 Two Properties: Lower-Order Consistency and Absence of Framing

### 3.2.1 Lower-Order Consistency

The example in the previous section shows that, while some games that can test whether a bound $k > 1$ is falsified (because they include a player-type $x_k$), can also test for all lower bounds $\ell = 1, \ldots, k-1$ (because they also include player-types $x_1, \ldots, x_{k-1}$), others are more limited and can only test for a subset of these bounds. The latter is the case for some classes of games often used for identification of hierarchies of beliefs, such as bimatrix games or $p$-beauty contest games. In such games, a subject who chooses randomly (not even 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 whether bound $k$ is falsified and, when doing so, can also test for bounds $\ell = 1, \ldots, k-1$, makes it harder for subjects with a true bound strictly below $k$ to pass all these tests, and 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 choose an $\ell$th-order rationalizable action in role $x_\ell$ for every $\ell = 1, \ldots, k$. By contrast, a subject whose rationality bound is $k' < k$ should be expected to fail to choose an $\ell$th-order rationalizable action in role $x_\ell$ for some $\ell = k'+1, \ldots, k$.

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.[13] 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

[13]Following the idea of finding a *conservative* bound (as in Kneeland, 2015) we are excluding the possibility of *trembling hand* mistakes.

identification is expected to have any external validity. Our first property can be stated as follows:

**Property 1** (Lower-Order Consistency)**.** *Game $\mathcal{G}$ is* lower-order consistent *if, for any $k \geq 2$,*

$$x_k \in X_{\mathcal{G}} \implies x_{k-1} \in X_{\mathcal{G}}.$$

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 it implies that the structure of a game of depth $n$ has to contain all the nodes from $x_1$ to $x_n$. 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 with depth $n$ that satisfies lower-order consistency. Then:*

*(i) $\mathcal{G}$ has at least $n$ distinct player types, of which one has a strictly dominant action.*

*(ii) $\mathcal{G}$ is dominance solvable in exactly $n$ rounds.*

### 3.2.2 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 $k$th-order rationalizable 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 choice and her second-order belief, about Player 2's first-order belief about Player 1's choice.

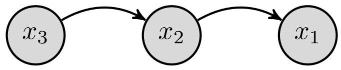



Figure 1: Game $G1$.

$G2$. *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 rationalizable 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 choice and the second-order one, about Player 2's first-order belief about Player 1's choice; (2) her first-order belief is about Player 2's choice and her second-order belief, about Player 2's first-order belief about Player 3's choice.

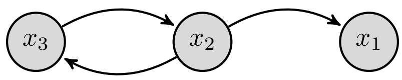



Figure 2: Game $G2$.

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.[14] 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 deletion that solves the game).[15]

[14] 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.

[15] 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 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."

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 from being immediately apparent.

**Property 2** (Absence of Framing)**.** *A lower-order consistent game $\mathcal{G}$ with depth $n \geq 2$ is* framing-free *if there exist:*

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

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

Let us provide some further intuition for the property. 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 $x_k$ is considered to be *framed* if no distinct paths of length $k-1$ that start at $x_k$ exist. 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. By contrast, 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 with depth 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.

In addition to condition ($i$), condition ($ii$) is also required for types of depth 3 and higher, 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 to player type $x_3$ being immediately apparent to $x_4$. 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.

### 3.3 Efficiency Result

We now sketch a simple, formal foundation for the heuristic arguments given above regarding the convenience of lower-order consistency and absence of framing in providing a more efficient estimation of the rationality bounds. Specifically, we find *sufficient* conditions on the probabilistic model discussed in Section 2.2 under which, both lower-order consistency and absence of framing are *necessary* for the estimation to be efficient. In Section 3.3.1 we show how to naturally decompose the estimation error, first in role-dependent components, and then in two parts: one that captures distortions due to purely observational issues, and one that captures distortions due to framing effects –as the ones discussed in the previous section. In Section 3.3.2 we introduce the statistical conditions (which rely on this decomposition) that, as shown in Proposition 1 in Section 3.3.3, guarantee the necessity of lower-order consistency and absence of framing for an efficient estimation of the rationality bounds.

#### 3.3.1 Decomposition of the Estimation Error

In Section 2.3.2 we defined the estimation error associated to game $\mathcal{G}$ as:

$$e_{\mathcal{G}} = |\hat{r}_{\mathcal{G}} - r|.$$

Now, if we suppose that the subjects' true rationality orders in game $\mathcal{G}$ are defined by the list $(r_x)_{x\in\mathcal{G}}$, we can easily rewrite the the estimation error as:

$$e_{\mathcal{G}} = \left|\min_{x\in X_{\mathcal{G}}} \hat{r}_x - r\right| = \left|\min_{x\in X_{\mathcal{G}}} (\hat{r}_x - r)\right| = \left|\min_{x\in X_{\mathcal{G}}} ((\hat{r}_x - r_x) + (r_x - r))\right| = \left|\min_{x\in X_{\mathcal{G}}} (\hat{e}_x + \bar{e}_x)\right|,$$

where, for each player-type $x \in X_{\mathcal{G}}$, we have the following two kind of errors:

- $\hat{e}_x := \hat{r}_x - r_x$ denotes the distortion in the estimation due to purely observational issues, and related to the fact that the sets of $k$th-order rationalizable structures are monotonically (weakly) decreasing on $k$, and can therefore lead to overestimations of the rationality bound (as discussed in Section 3.2.1).

- $\bar{e}_x := r_x - r$ denotes the distortion in the estimation due to framing effects, related to the fact that the inductive structure of the payoff dependencies could be very transparent in some games hence *inducing* the subject to reason iteratively (as discussed in Section 3.2.2).

### 3.3.2 Statistical Conditions

Thus far we have not imposed any statistical structure on the probability space $(\Omega, \mathcal{F}, P)$. Notice however, that the experimental literature implicitly assumes the existence of such a space, given that observations are usually considered to be random variables with particular error structures, e.g., logit or uniform. Here, we go one step further and openly model the probability space. In this way, we are able to propose four conditions on $(\Omega, \mathcal{F}, P)$ under which lower-order consistency and absence of framing become necessary for efficient estimation.

The first condition is a generic richness property that simply requires all the mutually consistent combinations of values of the errors and of the lower rationality bound to have nonnull probability.[16]

**Condition 1** (Richness condition)**.** *Let $\mathcal{G}$ be a game and let $k$ be in $\{0, 1, \ldots, n\}$. Then, for any player-type $x$ of depth $k_x > k$, every $\bar{\varepsilon} \in \{0, 1, \ldots, k_x - k - 1\}$ and every $\hat{\varepsilon} \in \{0, \ldots, k_x - k - 1 - \bar{\varepsilon}\} \cup \{\infty\}$,*

$$P(\hat{e}_x = \hat{\varepsilon}, \bar{e}_x = \bar{\varepsilon}, r = k) > 0.$$

The second condition requires that, conditional on the number of iterations performed in the role of a player-type of the highest depth, the error due to purely observational issues and the error due to framing effects are independent.

**Condition 2** (Conditional independence of the errors)**.** *Let $\mathcal{G}$ be a game and let $k$ be in $\{0, 1, \ldots, n\}$. Then, for any two player-types $x$ and $x'$, every $\hat{\varepsilon} \geq 0$ and every $\bar{\varepsilon} \geq 0$,*

$$P(\hat{e}_x = \hat{\varepsilon},\, \bar{e}_{x'} = \bar{\varepsilon} \,|\, r_{x_n} = k) = P(\hat{e}_x = \hat{\varepsilon} \,|\, r_{x_n} = k) P(\bar{e}_{x'} = \bar{\varepsilon} \,|\, r_{x_n} = k).$$

The third condition puts structure on the distributions of the errors due to purely observational issues. Loosely speaking, the first two parts of the condition formalize the assumption that these errors are, to some extent, of *noisy* nature; that is, independent across different player-types and equally distributed across different games.[17] More specifically, part (*i*) of

[16]In the condition, notice that, if $r_{x_n}(\omega) = k$ and player-type $x$ has depth $k_x > k$, then we necessarily have that $r_x(\omega) = k$ (part (*ii*) of Definition 3). Hence, the possible values that $\hat{e}_x(\omega)$ can take are, by definition, restricted to $\{0, 1, \ldots, k_x - k - 1\} \cup \{\infty\}$.

[17]Admittedly, the plausibility of these requirements hinges on the *similarity* of the problems that the different player-types of different games face. For this reason, the main games in our experiment share a common structure: for each player-type of depth $k$, one strictly dominated action, two actions that survive $k - 1$ rounds, and one action that survives $\geq k$ rounds.

Condition 3 requires that, conditional on the true number of iterations performed in the role of the player-type of the highest depth, $r_{x_n}$, the errors due to observational issues are independent from each other for player-types of depth *higher* than $r_{x_n}$. Part (*ii*) of Condition 3 requires that, for two games of the same depth and two player-types $x$ and $x'$ of the same depth, each from a different game, the distributions of $\hat{e}_x$ conditional on $r_x$ and the distribution of $\hat{e}'_{x'}$ conditional on $r'_{x'}$ are identical. It is important to interpret these two parts as properties on a distribution, not as requirements on the behavior of each individual subject. They do not require the choice of a given subject to be independent of the choice of the same individual in the role of a different player-type, but instead, that this independence holds in the distribution when considered as a property of the whole population faced with the given set of games. Finally, part (*iii*) of Condition 3 requires that the probability of $\hat{e}_x$ being higher than a given level, conditional on $r_{x_n}$, is increasing on $r_{x_n}$. To get a better intuition, notice that the higher $r_{x_n}$ is, the smaller is the set of values that $\hat{e}_x$ can take;[18] thus, what part (*iii*) of Condition 3 requires is that, as $r_{x_n}$ increases, the mass of probability that corresponds to the values that have disappeared from the new support, concentrate more on the higher values than on the low values in the new support.

**Condition 3** (Structure of the errors due to purely observational issues)**.** *Let $\mathcal{G}$ and $\mathcal{G}'$ be two games of depth $n$, and let $k$ be in $\{0, 1, \ldots, n\}$. Then, the following three hold:*

(*i*) *For every $\hat{\varepsilon} \geq 0$,*

$$P\left( \bigcap_{x \in X_{\mathcal{G}}: k_x > k} [\hat{e}_x \geq \hat{\varepsilon}] \,\middle|\, r_{x_n} = k \right) = \prod_{x \in X_{\mathcal{G}}: k_x > k} P(\hat{e}_x \geq \hat{\varepsilon} \,|\, r_{x_n} = k).$$

(*ii*) *For any two player-types $x \in X_{\mathcal{G}}$ and $x' \in X_{\mathcal{G}'}$ of depth higher than $k$, and every $\hat{\varepsilon} \geq 0$,*

$$P(\hat{e}_x = \hat{\varepsilon} \,|\, r_x = k) = P(\hat{e}'_{x'} = \hat{\varepsilon} \,|\, r'_{x'} = k).$$

(*iii*) *If $k < n$, for any player-type $x \in X_{\mathcal{G}}$ of depth higher than $k$, and every $\hat{\varepsilon} \geq 0$,*

$$P(\hat{e}_x > \hat{\varepsilon} \,|\, r_x = k) < P(\hat{e}'_{x'} > \hat{\varepsilon} \,|\, r'_{x'} = k + 1).$$

The fourth condition puts structure on the distributions of the errors for the different player-types due to having extra payoff dependencies or not, besides the ones associated to the natural hierarchy of beliefs. Before formally presenting them let us build some intuition first. Say that a player-type of depth $x_k$ is *trivial* if it only has one outwards path of length $k - 1$ or, if $k \geq 3$, only has one outwards path of length $k - 2$. Call it *nontrivial* otherwise. Then, in principle, a trivial player-type $x$ is one whose behavior is particularly simple to reason about

[18]For instance, if $x$ is of depth 6 and $r_{x_n} = 3$, then the set of possible values that $\hat{e}_x$ can take is $\{0, 1, 2, \infty\}$ whereas if $r_{x_n} = 4$, this set is $\{0, 1, \infty\}$.

for another player-type $x'$ of higher depth. In this context, Condition 4 relates the presence of trivial player-types with the arousal of framing into hierarchical thinking. To provide some intuition on the conditions, suppose first that a subject $\omega$ with $r(\omega) = 3$ is choosing in the role of $x_4$ (a player-type of depth 4). Then, whether $\omega$ performs 3 or 4 iterations in this role (i.e., $r_x(\omega) = 3$ and $r_x(\omega) = 4$, respectively) will be affected by the payoff-dependencies of $x_2$ (the player-type of depth 2), because, as $r(\omega) = 3$, $\omega$ already identifies the dominated actions of $x_2$ but may not necessarily identify the actions that survive 2 rounds. In this context, part $(i)$ of Condition 4 requires that, if $x_2$ only had one outwards link (her payoff only depends on her choice and that of $x_1$), then it would be more likely that $\omega$ performed the last iteration (and thus $r_x(\omega) = 4$) than if $x_2$ had richer payoff-dependencies. Similarly, if $r(\omega) = 2$, then the likelihood of $r_x(\omega) = 2$ and $r_x(\omega) = 3$ would be affected by how rich the payoff-dependencies of $x_3$ were. And similarly for lower values of $r(\omega)$. Part $(ii)$ of Condition 4 states that, once trivial player-types are absent, error distributions are equal.

**Condition 4** (Structure of the errors due to framing effects)**.** *Let $\mathcal{G}$ and $\mathcal{G}'$ be two games of depth $n$, and let $k$ be in $\{1, 2 \ldots, n-1\}$. Then, the following two hold:*

$(i)$ *If there exist two player-types $x \in X_{\mathcal{G}}$ and $x' \in X_{\mathcal{G}'}$ of depth $n+1-k$ such that $x$ is nontrivial and $x'$ is trivial, then, for every $\bar{\varepsilon} > 0$,*

$$P(\bar{e}_{x_n} = \bar{\varepsilon} \,|\, r = k) < P(\bar{e}'_{x'_n} = \bar{\varepsilon} \,|\, r = k).$$

$(ii)$ *If every player-type $x \in X_{\mathcal{G}}$ and $x' \in X_{\mathcal{G}'}$ with depth in $\{2, \ldots, n+1-k\}$ is nontrivial, then, for every $\bar{\varepsilon} \geq 0$,*

$$P(\bar{e}_{x_n} = \bar{\varepsilon} \,|\, r = k) = P(\bar{e}'_{x'_n} = \bar{\varepsilon} \,|\, r = k).$$

### 3.3.3 Result

Our first result establishes that, if the probability space satisfies the four conditions of the previous section, then lower-order consistency and absence of framing are *necessary* conditions for efficient estimation:

**Proposition 1.** *Let $(\Omega, \mathcal{F}, P)$ be a probability space that satisfies conditions 1, 2, 3 and 4, and let $\mathcal{G}$ and $\mathcal{G}'$ be two games of the same depth. Then, if $\mathcal{G}$ satisfies lower-order consistency and absence of framing and $\mathcal{G}'$ does not, $\mathcal{G}$ is more efficient than $\mathcal{G}'$.*

It is important to notice that these results *do not require knowledge* of the probability measure $P$ in the probability space: as long as the four conditions are considered to hold, a game satisfying lower-order consistency and absence of framing will do a better job (at least, according to the efficiency order we are considering), than any game that fails to satisfy either property. Now, it remains to be seen whether there exist games that satisfy both properties and, in case of a positive answer, whether such games can be simple enough to be implemented in a lab. The following section analyzes these issues.

# 4 E-Ring Games and Efficient Estimation of Rationality Bounds

After having motivated the properties of lower-order consistency and absence of framing as contributing towards a more efficient estimation of the rationality bounds, we now study what they imply jointly in terms of actual games to be played in the lab. In Section 4.1 we introduce the *e-ring games*, a class of games combining features of both Rubinstein's (1989) *email games* and Kneeland's (2015) *ring games*. Then, in Section 4.2 we prove that, within the class of games of depth 4, and attending to certain natural simplicity criteria, dominance solvable e-ring games of depth 4 exactly characterize the class of games satisfying lower-order consistency and absence of framing. Hence, among the games of depth 4 and under the statistical conditions of Section 3.3, the e-ring games stand as the *simplest* games delivering an *efficient estimation* of the rationality bounds.

## 4.1 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 be used to identify bounds up to 3 and bound *at least* 4.

**Example 2** (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 either gets 1 or 2 messages, whereby Player 2 either has the same number or one more message than Player 1. Each player is initially informed about the number of messages she receives, and the payoffs depend only on the number of messages a player receives as well as on the actions chosen by both players. 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 $a, b, c$ are 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** ($u_1(t_1)$)

$t_1 = 1$

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

$t_1 = 2$

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

**Player 2** ($u_2(t_2)$)

$t_2 = 1$

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

$t_2 = 2$

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

The above payoff structure has a unique (interim correlated) rationalizable action for each player and number of messages received. 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 estimate bounds up to 3 (and bound *at least* 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, again, according to Definition 1, 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 whether rationality bound $k$ is falsified. This is easy to see: $C$ is the only first-order rationalizable action for $(1, 2)$, $c$ is the only 2nd-order rationalizable action for $(2, 2)$, $C$ is the only 3rd-order rationalizable action for $(1, 1)$ and $c$ is the only 4th-order rationalizable action for $(2, 1)$. Thus the revealed rationality method yields the classification given in Table 1.

| $x_1$ | $x_2$ | $x_3$ | $x_4$ | $\hat{r}_{\mathcal{G}}$ |
|---|---|---|---|---|
| $C$ | $c$ | $C$ | $c$ | $\infty$ |
| $C$ | $c$ | $C$ | $b$ | 3 |
| $C$ | $c$ | $B$ | $b, c$ | 2 |
| $C$ | $b$ | $B, C$ | $b, c$ | 1 |

Table 1: Choice-vectors and estimated rationality bounds.

In addition, the estimated bound for a subject playing a dominated action would be 0 ($A$ or $B$ in the case of $x_1$, and $A$ or $a$ in the case of the remaining player-types).[19] $\square$

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

**Definition 8** (E-Ring Game)**.** *An* e-ring game of depth $n$ *(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. $T_i = \{1, 2, \ldots, n/2\}$ *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* $1 < t_2 \leq n/2$*, and otherwise* $\pi_1(n/2)[n/2] = 1$ *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 $2 \times 2$ payoff matrices for essentially any number of messages received. To further clarify the relation between types in an e-ring game, consider player $i$ who has received $k$ messages. By Definition 8, this player's type is $t_i = k$ and the payoff she obtains from action profile $(a_1, a_2)$ is given by $u_i(k, a_1, a_2)$. 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 $t_1 = k$ knows that, with probability $p_1$, Player 2 is of type $t_2 = k$ and that, with probability $1 - p_1$, Player 2 is of type $t_2 = k + 1$ (with the exception of type $t_1 = n/2$, who knows that Player 2 is of type $t_2 = n/2$ for sure). Similarly, Player 2 of type $t_2 = k$ knows that, with probability $p_2$, Player 1 is of type $t_1 = k - 1$ and that, with probability $1 - p_2$, Player 1 is of type $t_1 = k$ (with the exception of type $t_2 = 1$, who knows that Player 1 is of type $t_1 = 1$ for sure).

## 4.2 Characterization Result

The conceptual appeal of lower-order consistency and absence of framing for the estimation of rationality bounds has been discussed in Section 3.2, and the conditions for its mathematical necessity, in Section 3.3. We now turn to the question of the implementation of both

[19] In this example, we explain our identification strategy *as if* subjects switched roles. In the experiment detailed in Section 5, 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 dominance solvability structure.

axioms, and, more specifically, to the characterization of games satisfying the axioms and the complexity they entail.

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 of depth 4, since from the experimental literature the first four levels seem to be the empirically relevant ones (e.g., Arad and Rubinstein, 2012). Second, we order games following a *simplicity* criterion according to which, all things equal, we favor the lowest possible number of players, player-types, actions per player and directed links. This simplicity criterion 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 next result shows that, within the class of games of depth 4, the games satisfying lower-order consistency and absence of framing and satisfying the aforementioned simplicity criteria are the dominance solvable e-ring games with depth 4 and 2 actions per player.

**Proposition 2.** *Let $\mathcal{G}$ be a game. Then, $\mathcal{G}$ is simplest within the class of games of depth 4 and satisfies lower-order consistency and absence of framing if and only if $\mathcal{G}$ is a dominance solvable e-ring game of depth 4 with two actions per player.*

Combining this with Proposition 1 and assuming Conditions 1, 2, 3 and 4 further singles out the dominance solvable e-ring games with two actions as the simplest games yielding an efficient estimation of the rationality bounds. The next section shows how these games were implemented in the experiment to test the effectiveness of the proposed properties.

# 5 Experiment

## 5.1 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 for the identification of hierarchies of beliefs (ring games, dominance solvable games such as our 4×4 games and the $p$-beauty contest games). We chose these classes of games because they allow us to test the empirical validity of the two axioms proposed in Section 3, 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.[20] In both the e-ring and the ring games, there are 729 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.[21]

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.

### 5.2 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.

[20] 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. It is particularly important to control for risk preferences given that e-ring games are the only ones where they might play a role given the incompleteness of information. That is why we have constructed the games in such a way that the dominance structure is the same no matter of the individuals' risk attitudes.

[21] Of the 729 possible rational action profiles, 648 would be identified as $R1$ (88.9%), 72 as $R2$ (9.9%), 8 as $R3$ (1.1%) and 1 as $R4$ (0.1%).

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.

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. And third, they were clever but had no previous knowledge of Game Theory, which could influence their reasoning process.

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.[22]

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.[23] 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 $4 \times 4$ 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.

### 5.3 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

[22] 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.

[23] See Alaoui and Penta (2021) for a theoretical justification of this design choice that should give higher incentives to achieve higher levels in the hierarchy of beliefs.

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 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 $\ell = 1, \ldots, k-1$ 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. Thus, when testing for lower-order consistency, we draw a distinction between $R0$-$R1$ levels on one hand and $R2$-$R3$-$R4$ levels on the other, and do so for two reasons: one because those subjects that get identified as being $R0$ or $R1$ by any one of the games in our experiment are likely to have been reliably identified as such; two, the misclassification of a subject, from being identified as not having strictly speaking higher-order rationality ($R0$-$R1$) to being identified as one with strictly speaking higher-order rationality ($R2$-$R3$-$R4$) is particularly pertinent.

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). We distinguish between subjects that reveal not to have higher-order beliefs in rationality ($R0$-$R1$) from those who do ($R2$-$R3$-$R4$). To see that the 4×4 and the two beauty contest games are not as good at identifying higher-order levels $R2$-$R3$-$R4$, we look at the following 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):

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

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).[24] We calculate

[24]We do not consider all three games because the number of subjects satisfying this very strict condition is

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 $4\times4$ and the two beauty contest games (50 subjects)):

E-ring games: 35.4% (38.0%)  Ring games: 43.8% (44.9%).

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 $4\times4$ 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):

$4 \times 4$: 20.1% (25.6%)  2/3-BC: 21.3% (31.4%)  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 $4\times4$ 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 $4\times4$ and the two beauty contest games):

E-ring games: 12.2% (13.7%)  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

too small to make statistically significant comparisons.

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.

| | $R4$ | $R3$ | $R2$ |
|---|---|---|---|
| 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 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)).

**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 perform two tests.

**Test 2.1.** We consider all subjects that are identified as being of level $R0$ or $R1$ in at least two games of the 4×4 and the 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.[25]

[25]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

Both tests suggest that Property 2 reduces misclassification of subjects identified 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 properties, are less noisy in identifying higher-order rationality. These correlations are:

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

Notice that while, statistically, e-ring games may tendentially outperform the $4{\times}4$ 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.

# 6 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 the ones that do, frame individuals into thinking in higher levels, thereby compromising the very exercise.

This paper tackles this apparent contradiction in a comprehensive way. First, it formalizes the estimation problem in a probabilistic setting and links the properties of the distribution of unobservable bounds to the structure of the game. Second, using the language of graphs, the paper introduces a way of formalizing the payoff dependencies of games that allow for the axiomatic approach to be used for the first time to formalize the discussion regarding the estimation of bounds. It then shows that the axioms proposed, under some clearly stated conditions of the probability space, become necessary for a valid estimation and pin down a unique class of games: the *e-ring games*. Finally, the paper tests the properties empirically and finds evidence that suggests that both properties are indeed relevant in reducing the estimation error. As a result, e-ring games might constitute a useful starting point for the study of higher-order rationality.

The introduction of the probabilistic setting and the axiomatic approach in this literature might be important per se, as mentioned in Section 1. In fact, they enable a more transparent discussion of which statistical conditions are reasonable for the estimation exercise and what

games.

features of the game enhance the external validity of the estimation. Statistical conditions and properties of the game, once formulated explicitly, can be discussed, tested, discarded, and alternatives can be thought of, thus pushing the discussion in the literature forward in a more structured way.

The contribution of the paper is to provide a weak possibility result on the identification of higher-order rationality of individuals, which may pave the way for possibly stronger results in the following sense. The general problem of game dependent estimation errors, and what our conditions highlight, is that we need to understand more clearly the relationship between different games. In fact, our conditions show that games should be similar in terms of the distribution of noise across games. That is, our results suggest the impossibility of finding a general distribution, valid for any game and they highlight the need to establish clear classes of games. This would allow for the estimation of reliable distributions of rationality bounds within the class but not outside of it. We leave this next step for future research.

## 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 (2021). Cost-benefit analysis in reasoning. *Journal of Political Economy (forthcoming).*

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.

Arad, Ayala and Ariel Rubinstein (2012). The 11-20 money request game: A level-k reasoning study. *American Economic Review*, **102**(7), 3561–73.

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.

Coibion, Olivier, Yuriy Gorodnichenko, Saten Kumar and Jane Ryngaert (2021). Do You Know that I Know that You Know...? Higher-Order Beliefs in Survey Data. *The Quarterly Journal of Economics*, **136**(3), 1387–1446.

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.*

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**(PA), 369–400.

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

Kneeland, Terri (2015). Identifying higher-order rationality. *Econometrica*, **83**(5), 2065–2079.

Lim, Wooyoung and Siyang Xiong (2016). On identifying higher order rationality. *Mimeo.*

Nagel, Rosemarie (1995). Unraveling in guessing games: An experimental study. *American Economic Review*, **85**(5), 1313–26.

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.

# A Proofs

## A.1 Lemma 1

**Lemma 1.** *Let $\mathcal{G}$ be a game with depth $n$ that satisfies lower-order consistency. Then:*

$(i)$ *$\mathcal{G}$ has at least $n$ distinct player types, of which one has a strictly dominant action.*

$(ii)$ *$\mathcal{G}$ is dominance solvable in exactly $n$ 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 $\ell = 1, \dots, k-1$, and hence $X_{\mathcal{G}}$ contains player type $x_\ell$ for each $\ell = 1, \dots, k$, where, by definition, $x_1$ has a strictly dominant action. Part $(ii)$ follows from Definition 2 and from the definition of lower-order consistency. $\square$

## A.2 Proposition 1

In what follows, for any game $\mathcal{G}$ we denote:

$$\hat{e}_{\mathcal{G}} := \min_{x \in X_{\mathcal{G}}} \hat{e}_x,$$

and for every player-type $x \in X_{\mathcal{G}}$, we denote the depth of $x$ by $k_x$. We now first present a series of auxiliary lemmata and next, the proof of Proposition 1.

### A.2.1 First Auxiliary Result

**Lemma 2.** *Let $\mathcal{G}$ a game of depth $n$ with only one player-type of depth $n$. Then:*

$$e_{\mathcal{G}} = \hat{e}_{\mathcal{G}} + \bar{e}_{x_n}.$$

*Proof.* Pick the $k \in \{0, \dots, n\}$ where $r_{x_n} = k$. Then, we know by condition $(ii)$ in Definition 3 that $r_x = k_x$ for every player-type $x$ where $k_x \leq k$ and $r_x = k$ for every player-type $x$ where $k_x > k$. It follows that $\hat{r}_x = \infty$ for every player-type $x$ where $k_x \leq k$. Then:

$$\begin{aligned} e_{\mathcal{G}} &= \left| \min_{x \in X_{\mathcal{G}}} (\hat{e}_x + \bar{e}_x) \right| = \left| \min_{x \in X_{\mathcal{G}} : k_x > k} (\hat{e}_x + \bar{e}_x) \right| = \left| \min_{x \in X_{\mathcal{G}} : k_x > k} (\hat{e}_x + r_x - r) \right| \\ &= \left| \min_{x \in X_{\mathcal{G}} : k_x > k} (\hat{e}_x + k - r) \right| = \left| \min_{x \in X_{\mathcal{G}} : k_x > k} \hat{e}_x + (r_{x_n} - r) \right| = \min_{x \in X_{\mathcal{G}}} \hat{e}_x + (r_{x_n} - r) = \hat{e}_{\mathcal{G}} + \bar{e}_{x_n}, \end{aligned}$$

where the second to last equality relies on the fact that $r_{x_n} \geq r$. $\square$

### A.2.2 For the Necessity of Lower-Order Consistency

**Lemma 3.** *Let $(\Omega, \mathcal{F}, P)$ be a probability space satisfying conditions 1 and 3 and let $\mathcal{G}$ be a game of depth $n$ that satisfies lower-order consistency. Then, for any game $\mathcal{G}'$ of depth $n$ and* *any $\hat{\varepsilon} > 0$, the following two hold:*

(*i*) *For any $k = 0, 1, \dots, n-1$,*

$$P(\hat{e}_{\mathcal{G}} > \hat{\varepsilon} \,|\, r_{x_n} = k) \leq P(\hat{e}'_{\mathcal{G}'} > \hat{\varepsilon} \,|\, r'_{x'_n} = k).$$

(*ii*) *If $\mathcal{G}'$ does not satisfy lower-order consistency, then there exists some $k^* = 0, 1, \dots, n-1$ such that:*

$$P(\hat{e}_{\mathcal{G}} > \hat{\varepsilon} \,|\, r_{x_n} = k^*) < P(\hat{e}'_{\mathcal{G}'} > \hat{\varepsilon} \,|\, r'_{x'_n} = k^*).$$

*Proof.* Fix two games $\mathcal{G}$ and $\mathcal{G}'$ of depth $n$, and pick $k = 0, 1, \dots, n-1$ and $\hat{\varepsilon} \geq 0$. Then, we have that:

$$\begin{aligned}
P(\hat{e}_{\mathcal{G}} > \hat{\varepsilon} \,|\, r_{x_n} = k) &\overset{(1)}{=} P\left( \left[ \min_{x \in X_{\mathcal{G}}} \hat{e}_x > \hat{\varepsilon} \right] \middle| \; r_{x_n} = k \right) \\
&\overset{(2)}{=} P\left( \bigcap_{x \in X_{\mathcal{G}}} [\hat{e}_x > \hat{\varepsilon}] \middle| \; r_{x_n} = k \right) \\
&\overset{(3)}{=} \prod_{x \in X_{\mathcal{G}} : k_x > k} P(\hat{e}_x = \hat{\varepsilon} \,|\, r_{x_n} = k) \\
&\overset{(4)}{=} \prod_{\ell = k+1}^{n} \prod_{x \in X_{\mathcal{G}} : k_x = \ell} P(\hat{e}_x = \hat{\varepsilon} \,|\, r_{x_n} = k).
\end{aligned}$$

Equalities (1) and (4) are immediate; (2) follows from condition (*i*) of Definition 3 (if $r_{x_n}(\omega) = k$ then $r_x(\omega) = k_x$ and hence $\hat{r}_x(\omega) = \infty$ for every $x$ with $k_x \leq k$); (3) follows from part (*i*) of Condition 3. Similarly, we obtain that:

$$\begin{aligned}
P(\hat{e}'_{\mathcal{G}'} > \hat{\varepsilon} \,|\, r'_{x'_n} = k) &= \prod_{\ell = k+1}^{n} \prod_{x \in X_{\mathcal{G}'} : k_x = \ell} P(\hat{e}'_{x'} = \hat{\varepsilon} \,|\, r'_{x'_n} = k) \\
&= \prod_{\ell = k+1}^{n} \prod_{x \in X_{\mathcal{G}'} : k_x = \ell} P(\hat{e}_x = \hat{\varepsilon} \,|\, r_{x_n} = k),
\end{aligned}$$

where the second equality follows from part (*ii*) of Condition 3.[26] If both games satisfy lower-order consistency we have that:

$$P(\hat{e}_{\mathcal{G}} > \hat{\varepsilon} \,|\, r_{x_n} = k) = \prod_{\ell = k+1}^{n} P(\hat{e}_{x_\ell} = \hat{\varepsilon} \,|\, r_{x_n} = k),$$

[26]That is, from the fact that for every $x \in X_{\mathcal{G}}$ and every $x' \in X_{\mathcal{G}'}$ of the same depth (and higher than $k$) we have that $P(\hat{e}'_{x'} = \hat{\varepsilon} \,|\, r'_{x'_n} = k) = P(\hat{e}_x = \hat{\varepsilon} \,|\, r_{x_n} = k)$

$$P(\hat{e}'_{\mathcal{G}'} > \hat{\varepsilon} \,|\, r'_{x'_n} = k) = \prod_{\ell=k+1}^{n} P(\hat{e}_{x_\ell} = \hat{\varepsilon} \,|\, r_{x_n} = k),$$

and thus, we obtain the inequality in $(i)$ is clear. If only $\mathcal{G}$ satisfies lower-order consistency, then there exist some $k^* = 0, 1, \ldots, n-1$ and some $\ell = k^* + 1, \ldots, n$ such that there is no $x \in X_{\mathcal{G}'}$ with depth $k_x = \ell$. Since the opposite cannot happen, we have that:[27]

$$\prod_{\ell=k^*+1}^{n} P(\hat{e}_{x_\ell} = \hat{\varepsilon} \,|\, r_{x_n} = k^*) < \prod_{\ell=k^*+1}^{n} \prod_{x' \in X_{\mathcal{G}'}: k_{x'} = \ell} P(\hat{e}'_{x'} = \hat{\varepsilon} \,|\, r'_{x'_n} = k^*)$$

and thus, combining the latter with the equalities above, we obtain:

$$\begin{aligned} P(\hat{e}_{\mathcal{G}} > \hat{\varepsilon} \,|\, r_{x_n} = k^*) &= \prod_{\ell=k+1}^{n} \prod_{x \in X_{\mathcal{G}'}: k_x = \ell} P(\hat{e}_x = \hat{\varepsilon} \,|\, r_{x_n} = k^*) \\ &= \prod_{\ell=k^*+1}^{n} P(\hat{e}_{x_\ell} = \hat{\varepsilon} \,|\, r_{x_n} = k^*) \\ &< \prod_{\ell=k^*+1}^{n} \prod_{x' \in X_{\mathcal{G}'}: k_{x'} = \ell} P(\hat{e}'_{x'} = \hat{\varepsilon} \,|\, r'_{x'_n} = k^*) \\ &= P(\hat{e}'_{\mathcal{G}'} > \hat{\varepsilon} \,|\, r'_{x'_n} = k^*), \end{aligned}$$

what allows for concluding the strict inequality in $(ii)$. □

### A.2.3 For the Necessity of Absence of Framing

**Lemma 4.** *Let $(\Omega, \mathcal{F}, P)$ be a probability space satisfying Condition 4 and let $\mathcal{G}$ be a game of depth $n$ that satisfies lower-order consistency and absence of framing. Then, for any game $\mathcal{G}'$ of depth $n$ the following two hold:*

$(i)$ *For any $k = 0, 1, \ldots, n-1$ and every $\bar{\varepsilon} = 1, 2, \ldots, n-k$,*

$$P(\bar{e}_{x_n} = \bar{\varepsilon} \,|\, r = k) \leq P(\bar{e}'_{x'_n} = \bar{\varepsilon} \,|\, r = k).$$

$(ii)$ *If $\mathcal{G}'$ satisfies lower-order consistency but not absence of framing, then there exists some $k^* = 1, \ldots, n$ such that, for every $\bar{\varepsilon} = 1, 2, \ldots, n-k^*$,*

$$P(\bar{e}_{x_n} = \bar{\varepsilon} \,|\, r = k^*) < P(\bar{e}'_{x'_n} = \bar{\varepsilon} \,|\, r = k^*).$$

*Proof.* Let us first make the trivial observation that for a game that satisfies lower-order consistency, absence of framing and the absence of trivial types are *equivalent*. Given this, we

[27] The strict inequality holds due to Condition 1, which implies that $P(\hat{e}_x = \hat{\varepsilon} \,|\, r_{x_n} = k^*) < 1$ for any $k^* = 0, 1, \ldots, n-1$ and any $x \in X_{\mathcal{G}}$ of depth $\ell \in \{k^* + 1, \ldots, n\}$, and every $\hat{\varepsilon} > 0$.

know that $\mathcal{G}$ does not contain trivial types and thus, for any $k \in \{1, 2, \ldots, n-1\}$ and every $\bar{\varepsilon} = 1, 2, \ldots, n-k$ we have that:

- By part $(i)$ of Condition 4, if $\mathcal{G}'$ contains some trivial player-type of depth $\ell \leq n+1-k$, then $P(\bar{e}_{x_n} = \bar{\varepsilon} \,|\, r = k) < P(\bar{e}'_{x'_n} = \bar{\varepsilon} \,|\, r = k)$.
- By part $(ii)$ of Condition 4, if $\mathcal{G}'$ contains no trivial player-type of depth $\ell \leq n+1-k$, then $P(\bar{e}_{x_n} = \bar{\varepsilon} \,|\, r = k) = P(\bar{e}'_{x'_n} = \bar{\varepsilon} \,|\, r = k)$.

And thus, in either case we conclude that the inequality in $(i)$ holds. Similarly, if $\mathcal{G}'$ satisfies lower-order consistency but not absence of framing, we know that it contains some trivial player-type, and thus, using part $(i)$ of Condition 4 we conclude that the claim in $(ii)$ here holds. $\square$

### A.2.4 Last Auxiliary Result

**Lemma 5.** *Let $(\Omega, \mathcal{F}, P)$ be a probability space satisfying Condition 3, let $\mathcal{G}$ be a game of depth $n$, and let $k$ be in $\{0, 1, \ldots, n\}$. Then for any player-type $x \in X_{\mathcal{G}}$ of depth higher than $k$, and every $\hat{\varepsilon} \geq 0$,*

$$P(\hat{e}_{\mathcal{G}} > \hat{\varepsilon} \,|\, r_{x_n} = k) < P(\hat{e}_{\mathcal{G}} > \hat{\varepsilon} - \delta \,|\, r_{x_n} = k + \delta),$$

*for every $\delta \in \{1, \ldots, \min\{\hat{\varepsilon}, n-k\}\}$.*

*Proof.* Fix $k \in \{0, 1 \ldots, n\}$, $\hat{\varepsilon} \geq 0$ and $\delta \in \{1, \ldots, \min\{\hat{\varepsilon}, n-k\}\}$. Then, because of part $(i)$ of Condition 3 we know that:

$$\begin{aligned} P(\hat{e}_{\mathcal{G}} > \varepsilon - \delta \,|\, r_{x_n} = k) &= \prod_{x \in X_{\mathcal{G}} : k_x > k} P(\hat{e}_x > \varepsilon - \delta \,|\, r_{x_n} = k) \\ &= \left[ \prod_{x \in X_{\mathcal{G}} : k_x > k+\delta} P(\hat{e}_x > \varepsilon - \delta \,|\, r_{x_n} = k) \right] \cdot \alpha, \end{aligned} \tag{1}$$

where:

$$\alpha := \prod_{x \in X_{\mathcal{G}} : k_x \in \{k+1, \ldots, k+\delta\}} P(\hat{e}_x > \varepsilon - \delta \,|\, r_{x_n} = k) < 1.$$

Similarly, we also know that:

$$P(\hat{e}_{\mathcal{G}} > \varepsilon - \delta \,|\, r_{x_n} = k + \delta) = \prod_{x \in X_{\mathcal{G}} : k_x > k+\delta} P(\hat{e}_x > \varepsilon - \delta \,|\, r_{x_n} = k + \delta). \tag{2}$$

In addition, we know from part $(iii)$ of Condition 3 that, for any player-type $x$ of depth higher that $k + \delta$,

$$P(\hat{e}_x > \varepsilon - \delta \,|\, r_{x_n} = k) < P(\hat{e}_x > \varepsilon - \delta \,|\, r_{x_n} = k + \delta). \tag{3}$$

Combining (1), (2) and (3), we have that:

$$P(\hat{e}_{\mathcal{G}} > \varepsilon - \delta \,|\, r_{x_n} = k) < P(\hat{e}_{\mathcal{G}} > \varepsilon - \delta \,|\, r_{x_n} = k + \delta),$$

and hence, the obvious fact that:

$$P(\hat{e}_{\mathcal{G}} > \varepsilon \,|\, r_{x_n} = k) \leq P(\hat{e}_{\mathcal{G}} > \varepsilon - \delta \,|\, r_{x_n} = k),$$

completes the proof. $\square$

#### A.2.5 Proof of the Proposition

**Proposition 1.** *Let $(\Omega, \mathcal{F}, P)$ be a probability space that satisfies conditions 1, 2, 3 and 4, and let $\mathcal{G}$ and $\mathcal{G}'$ be two games of the same depth. Then, if $\mathcal{G}$ satisfies lower-order consistency and absence of framing and $\mathcal{G}'$ does not, $\mathcal{G}$ is more efficient than $\mathcal{G}'$.*

*Proof.* Fix two games $\mathcal{G}$ and $\mathcal{G}'$ of depth $n$, and pick $\varepsilon \geq 0$. For notational convenience, for each $k = 0, 1, \ldots, n-1$, set $\bar{\varepsilon}(k) := \min\{\varepsilon, n-k\}$.[28] Then, we have that:

$$\begin{aligned}
&P(e_{\mathcal{G}} > \varepsilon) = \\
&\overset{(1)}{=} \sum_{k=0}^{n-1} P(e_{\mathcal{G}} > \varepsilon,\, r = k) + P(e_{\mathcal{G}} > \varepsilon, r \geq n) \\
&\overset{(2)}{=} \sum_{k=0}^{n-1} P(\hat{e}_{\mathcal{G}} + \bar{e}_{x_n} > \varepsilon,\, r = k) + P(r \geq n) \\
&\overset{(3)}{=} \sum_{k=0}^{n-1} \sum_{\bar{\varepsilon}=0}^{\bar{\varepsilon}(k)} P(\hat{e}_{\mathcal{G}} > \varepsilon - \bar{\varepsilon},\, \bar{e}_{x_n} = \bar{\varepsilon},\, r = k) + P(r \geq n) \\
&\overset{(4)}{=} \sum_{k=0}^{n-1} \sum_{\bar{\varepsilon}=0}^{\bar{\varepsilon}(k)} P(\hat{e}_{\mathcal{G}} > \varepsilon - \bar{\varepsilon},\, \bar{e}_{x_n} = \bar{\varepsilon},\, r_{x_n} = k + \bar{\varepsilon}) + P(r \geq n) \\
&\overset{(5)}{=} \sum_{k=0}^{n-1} \sum_{\bar{\varepsilon}=0}^{\bar{\varepsilon}(k)} P(r_{x_n} = k + \bar{\varepsilon}) P(\hat{e}_{\mathcal{G}} > \varepsilon - \bar{\varepsilon},\, \bar{e}_{x_n} = \bar{\varepsilon} \,|\, r_{x_n} = k + \bar{\varepsilon}) + P(r \geq n) \\
&\overset{(6)}{=} \sum_{k=0}^{n-1} \sum_{\bar{\varepsilon}=0}^{\bar{\varepsilon}(k)} P(r_{x_n} = k + \bar{\varepsilon}) P(\hat{e}_{\mathcal{G}} > \varepsilon - \bar{\varepsilon} \,|\, r_{x_n} = k + \bar{\varepsilon}) P(\bar{e}_{x_n} = \bar{\varepsilon} \,|\, r_{x_n} = k + \bar{\varepsilon}) + P(r \geq n)
\end{aligned}$$

[28] The equation that follows clarifies the convenience of this notation. The range of $\bar{e}_{x_n}$ conditional on $r = k$ is $\{0, 1, \ldots, n-k\}$; however, it will become obvious why we want to exclude. w.l.o.g., the cases in which $\varepsilon - \bar{e}_{x_n} < 0$.

$$\overset{(7)}{=} \sum_{k=0}^{n-1} \sum_{\bar{\varepsilon}=0}^{\bar{\varepsilon}(k)} P(\hat{e}_{\mathcal{G}} > \varepsilon - \bar{\varepsilon} \,|\, r_{x_n} = k + \bar{\varepsilon}) P(\bar{e}_{x_n} = \bar{\varepsilon},\, r_{x_n} = k + \bar{\varepsilon}) + P(r \geq n)$$

$$\overset{(8)}{=} \sum_{k=0}^{n-1} \sum_{\bar{\varepsilon}=0}^{\bar{\varepsilon}(k)} P(\hat{e}_{\mathcal{G}} > \varepsilon - \bar{\varepsilon} \,|\, r_{x_n} = k + \bar{\varepsilon}) P(\bar{e}_{x_n} = \bar{\varepsilon},\, r = k) + P(r \geq n)$$

$$\overset{(9)}{=} \sum_{k=0}^{n-1} P(r = k) \sum_{\bar{\varepsilon}=0}^{\bar{\varepsilon}(k)} P(\hat{e}_{\mathcal{G}} > \varepsilon - \bar{\varepsilon} \,|\, r_{x_n} = k + \bar{\varepsilon}) P(\bar{e}_{x_n} = \bar{\varepsilon} \,|\, r = k) + P(r \geq n)$$

Equalities (1), (3), (5), (7), (8) and (9) are immediate; (2) relies on Lemma 2 and condition $(ii)$ in Definition 4 (to notice that, if $r(\omega) \geq n$ then $\hat{r}_{\mathcal{G}}(\omega) = \infty$ and thus, $e_{\mathcal{G}}(\omega) > \varepsilon$); (4) follows from the definition of $\bar{e}_{x_n}$; (6) follows immediately from Condition 2. Similarly, we obtain that:

$$P(e'_{\mathcal{G}'} > \varepsilon) = \sum_{k=0}^{n-1} P(r = k) \sum_{\bar{\varepsilon}=0}^{\bar{\varepsilon}(k)} P(\hat{e}'_{\mathcal{G}'} > \varepsilon - \bar{\varepsilon} \,|\, r'_{x'_n} = k + \bar{\varepsilon}) P(\bar{e}'_{x'_n} = \bar{\varepsilon} \,|\, r = k) + P(r \geq n).$$

Now, to simplify the rest of the proof let us denote for every $k = 0, 1, \ldots, n-1$ and every $\bar{\varepsilon} = 0, 1, \ldots, \bar{\varepsilon}(k)$,

$$q(k, \varepsilon, \bar{\varepsilon}) := P(\hat{e}_{\mathcal{G}} > \varepsilon - \bar{\varepsilon} \,|\, r_{x_n} = k + \bar{\varepsilon}) \quad \text{and} \quad p(k, \bar{\varepsilon}) := P(\bar{e}_{x_n} = \bar{\varepsilon} \,|\, r = k).$$

With this notation, $P(e_{\mathcal{G}} > \varepsilon)$ becomes:

$$P(e_{\mathcal{G}} > \varepsilon) = \sum_{k=0}^{n-1} P(r = k) \sum_{\bar{\varepsilon}=0}^{\bar{\varepsilon}(k)} q(k, \varepsilon, \bar{\varepsilon}) p(k, \bar{\varepsilon}) + P(r \geq n)$$

Now, defining $q'(k, \varepsilon, \bar{\varepsilon})$ and $p'(k, \bar{\varepsilon})$ for $\mathcal{G}'$ in analogous terms, we obtain that:

$$P(e_{\mathcal{G}} > \varepsilon) - P(e'_{\mathcal{G}'} > \varepsilon) = \sum_{k=0}^{n-1} P(r = k) \sum_{\bar{\varepsilon}=0}^{\bar{\varepsilon}(k)} \big(q(k, \varepsilon, \bar{\varepsilon}) p(k, \bar{\varepsilon}) - q'(k, \varepsilon, \bar{\varepsilon}) p'(k, \bar{\varepsilon})\big)\,.$$

Now:

(A) Since $\mathcal{G}$ satisfies lower-order consistency we know from part $(i)$ of Lemma 3 that $q(k, \varepsilon, \bar{\varepsilon}) \leq q'(k, \varepsilon, \bar{\varepsilon})$ for every $k = 0, 1, \ldots n-1$ and every $\bar{\varepsilon} = 0, 1, \ldots, \bar{\varepsilon}(k)$.

(B) Since $\mathcal{G}$ satisfies lower-order consistency and absence of framing we know from part $(i)$ of Lemma 4 that $p(k, \bar{\varepsilon}) \leq p'(k, \bar{\varepsilon})$ for every $k = 0, 1, \ldots n-1$ and every $\bar{\varepsilon} = 1, 2, \ldots, \bar{\varepsilon}(k)$.

In consequence, we have that:

$$\sum_{\bar{\varepsilon}=0}^{\bar{\varepsilon}(k)} (q(k, \varepsilon, \bar{\varepsilon}) p(k, \bar{\varepsilon}) - q'(k, \varepsilon, \bar{\varepsilon}) p'(k, \bar{\varepsilon})) \overset{(1)}{\leq}$$

$$\overset{(1)}{\leq} \sum_{\bar{\varepsilon}=0}^{\bar{\varepsilon}(k)} q'(k,\varepsilon,\bar{\varepsilon}) \big(p(k,\bar{\varepsilon}) - p'(k,\bar{\varepsilon})\big)$$

$$= \sum_{\bar{\varepsilon}=1}^{\bar{\varepsilon}(k)} q'(k,\varepsilon,\bar{\varepsilon}) \big(p(k,\bar{\varepsilon}) - p'(k,\bar{\varepsilon})\big) + q'(k,\varepsilon,0)(p(k,0) - p'(k,0))$$

$$= \sum_{\bar{\varepsilon}=1}^{\bar{\varepsilon}(k)} q'(k,\varepsilon,\bar{\varepsilon}) \big(p(k,\bar{\varepsilon}) - p'(k,\bar{\varepsilon})\big) - q'(k,\varepsilon,0) \sum_{\bar{\varepsilon}=1}^{\bar{\varepsilon}(k)} \big(p(k,\bar{\varepsilon}) - p'(k,\bar{\varepsilon})\big)$$

$$= \sum_{\bar{\varepsilon}=1}^{\bar{\varepsilon}(k)} \big(q'(k,\varepsilon,\bar{\varepsilon}) - q'(k,\varepsilon,0)\big) \big(p(k,\bar{\varepsilon}) - p'(k,\bar{\varepsilon})\big) \overset{(2)}{\leq} 0,$$

for every $k = 0, 1, \dots n-1$, where inequality (1) follows from (A), and inequality (2) follows from (B) and Lemma 5.[29] Then:

(*a*) If $\mathcal{G}'$ does not satisfy lower-order consistency then we know from part (*ii*) of Lemma 3 that that there exists some $k^* = 0, 1, \dots, n-1$ such that $q(k^*, \varepsilon, \bar{\varepsilon}) < q'(k^*, \varepsilon, \bar{\varepsilon})$.

(*b*) If $\mathcal{G}'$ satisfies lower-order consistency but not satisfy absence of framing then we know from part (*ii*) of Lemma 4 that that there exists some $k^* = 0, 1, \dots, n-1$ such that $p(k^*, \bar{\varepsilon}) < p'(k^*, \bar{\varepsilon})$ for every $\bar{\varepsilon} = 1, 2, \dots, \bar{\varepsilon}(k^*)$.

Hence, in either case (the inequality in (1) is strict if ($a$), and the one in (2) is strict if ($b$)), we conclude that:

$$\sum_{\bar{\varepsilon}=0}^{\bar{\varepsilon}(k^*)} \big(q(k^*,\varepsilon,\bar{\varepsilon}) p(k^*,\bar{\varepsilon}) - q'(k^*,\varepsilon,\bar{\varepsilon}) p'(k^*,\bar{\varepsilon})\big) < 0,$$

and thus, the proof is complete:[30]

$$P(e_{\mathcal{G}} > \varepsilon) - P(e'_{\mathcal{G}'} > \varepsilon) = \sum_{k=0}^{n-1} P(r=k) \sum_{\bar{\varepsilon}=0}^{\bar{\varepsilon}(k)} \big(q(k,\varepsilon,\bar{\varepsilon}) p(k,\bar{\varepsilon}) - q'(k,\varepsilon,\bar{\varepsilon}) p'(k,\bar{\varepsilon})\big) < 0.$$

□

### A.3 Proposition 2

**Proposition 2.** *Let $\mathcal{G}$ be a game. Then, $\mathcal{G}$ is simplest within the class of games of depth 4 and satisfies lower-order consistency and absence of framing if and only if $\mathcal{G}$ is a dominance solvable e-ring game of depth 4 with two 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

[29] To apply the lemma notice that, as $\bar{\varepsilon}(k) = \min\{\varepsilon, n-k\}$, it holds that $\bar{\varepsilon} \in \{0, 1, \dots, \min\{\varepsilon, n-k\}\}$.

[30] The strict positiveness is guaranteed by Condition 1, which implies that $P(r = k^*) > 0$.

a set of links containing $(x_4, x_3)$ and $(x_2, x_1)$. Also, by definition, there is no link starting from $x_1$, and minimality allows for excluding the presence of further player types. We distinguish now two cases:

- Suppose that $\mathcal{G}$ does not include link $(x_3, x_2)$. Then, as no link departs form $x_1$, it is not possible to have a path of length 3 that starts at $x_4$. Thus, since $\mathcal{G}$ is assumed to satisfy absence of framing, this case can be excluded.

- Suppose then that $\mathcal{G}$ contains link $(x_3, x_2)$. Then, minimality 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. $\square$