ECONOMETRICA

JOURNAL OF THE ECONOMETRIC SOCIETY

An International Society for the Advancement of Economic Theory in its Relation to Statistics and Mathematics

http://www.econometricsociety.org/

Econometrica, Vol. 83, No. 5 (September, 2015), 2065–2079

IDENTIFYING HIGHER-ORDER RATIONALITY

TERRI KNEELAND
University College London, London WC1E 6BT, U.K.

Econometrica, Vol. 83, No. 5 (September, 2015), 2065–2079

IDENTIFYING HIGHER-ORDER RATIONALITY

BY TERRI KNEELAND[1]

Strategic choice data from a carefully chosen set of ring-network games are used to obtain individual-level estimates of higher-order rationality. The experimental design exploits a natural exclusion restriction that is considerably weaker than the assumptions underlying alternative designs in the literature. In our data set, 93 percent of subjects are rational, 71 percent are rational and believe others are rational, 44 percent are rational and hold second-order beliefs that others are rational, and 22 percent are rational and hold at least third-order beliefs that others are rational.

KEYWORDS: Rationality, higher-order rationality, epistemic game theory.

1. INTRODUCTION

ONE OF THE MAIN ASSUMPTIONS UNDERLYING STRATEGIC BEHAVIOR is the assumption of rationality and higher-order rationality. A player is rational if she plays a best response to her beliefs. She satisfies higher-order rationality if she believes others are rational, if she believes others believe others are rational, if she believes others believe others believe others are rational, and so on. The extent to which this assumption holds true is important for both understanding and modeling strategic behavior.

There are two main approaches to investigating rationality assumptions: one is to elicit both strategic choices and beliefs and the other is to recover rationality directly from choice data. Under the belief elicitation approach, subjects choose actions and state first-order beliefs about the actions of their opponents.[2] A subject is then considered rational if her choice of action is a best response to her stated first-order beliefs. However, this approach cannot easily be extended to higher-order rationality because higher-order rationality does not imply a best response relationship between higher-order beliefs.[3] The other approach, the level-k model, imposes structural assumptions on players' beliefs and then recovers rationality directly from choice data.[4] This approach allows for the identification of higher order of rationality but at the expense of

[1]I would like to thank my supervisor Yoram Halevy for support and guidance throughout the process of developing and writing this paper. I am grateful to Ryan Oprea for invaluable feedback and encouragement and Mike Peters for numerous conversations, without which this paper would not have been possible. For very helpful comments I would like to thank Doug Bernheim, Colin Camerer, Vince Crawford, Donna Feir, David Freeman, Amanda Friedenberg, Ed Green, Li Hao, Andrew Hill, Kevin Leyton-Brown, Wei Li, Ariel Rubinstein, Ran Spiegler, Charles Sprenger, and anonymous referees.

[2]For examples, see Costa-Gomes and Weizsäcker (2008) and Healy (2011).

[3]For example, even if a subject believes her opponents are rational, her first-order beliefs about the actions of her opponent need not be a best response to her second-order beliefs (beliefs about the beliefs her opponent holds about the actions of her opponents).

[4]While not commonly interpreted as such, the level-k type distribution can be considered an estimate of an order of rationality distribution. The level-k model makes specific assumptions

 DOI: 10.3982/ECTA11983

possible misidentification due to the strong assumptions imposed. This paper provides a method to identify higher-order rationality from strategic choice data under considerably weaker assumptions.[5]

We make use of an existing characterization of behavior established by Bernheim (1984), Pearce (1984), and Tan and Werlang (1988): a subject who satisfies kth-order rationality must play a kth-order rationalizable action.[6] However, this characterization does not fully separate different orders of rationality because the behavioral implications of higher orders of rationality are contained in the behavioral implications of lower orders of rationality (i.e., the kth-order rationalizable sets always nest one another). To deal with this identification problem, we focus on a special class of games, *ring-network games*. Ring games are used to isolate the behavioral implications of different orders of rationality under the natural exclusion restriction: lower-order rational players do not respond to changes in higher-order payoffs.[7]

A ring game is a series of 2-player normal form games where the opponent structure is relaxed relative to standard game forms. For example, in a 3-player ring game, player 1's payoff depends on the action of player 2. Player 2's payoff depends on the action of player 3. And, player 3's payoff depends on the action of player 1. The opponent structure of the ring game allows one to define games which induce differences in higher-order payoffs without affecting lower-order payoffs. This is not possible in standard game forms (e.g., bimatrix games) because there is a tight link between higher- and lower-order payoffs. Adding new players into the game creates additional degrees of freedom. Thus, the ring game can be used to identify orders of rationality by focusing on a carefully chosen set of games that differ only in higher-order payoffs. To the best of our knowledge, this is the first paper to use ring games (or, more generally, network games) to study strategic reasoning empirically.[8]

The experiment consists of a set of eight 4-player ring games. We classify subjects into orders of rationality based on their choices in the eight games.

about the beliefs held by players that satisfy different orders of rationality: a rational player always believes her opponent is playing each action with equal probability, a rational player that believes her opponent is rational always believes that her opponent is playing a best response to the belief that her opponent is playing each action with equal probability, and so on.

[5]The estimated order of rationality distribution in this paper is interpretable as a level-k level distribution that is independent of the level-0 specification. The only other paper to provide a level-0 independent distribution is Burchardi and Penczynski (2014). They analyzed incentivized communication between group members in order to determine a subject's depth of reasoning.

[6]A kth-order rationalizable action is one that survives k rounds of iterated deletion of strategies that are never best responses.

[7]A player's kth-order payoffs are her kth-order beliefs about payoffs. However, since we consider only complete information games, a player's beliefs about payoffs are just given by the payoffs themselves; thus we drop the language of beliefs when discussing higher-order payoffs.

[8]Cubitt and Sugden (1994) used a ring game to illustrate a paradox related to iterated deletion of weakly dominated strategies.

We find that 93 percent of subjects are rational, 71 percent are rational and believe others are rational (second-order rational), 44 percent are second-order rational and hold second-order beliefs that others are rational (third-order rational), and 22 percent are third-order rational and hold third-order beliefs that others are rational (fourth-order rational).

These results suggest somewhat higher orders of rationality than many previous experiments have found. In particular, the results from the level-k literature tend to put more weight on rational and second-order rational types and little weight on higher-order rational types.[9] The motivating claim of this paper is that ring games allow us to impose weaker identification restrictions, thus giving more direct and therefore more reliable results. An alternative explanation is that the change in game form may increase the ease of expressing higher-order rationality relative to standard game forms. We report a robustness treatment that casts doubt on this type of explanation. The treatment perturbs the ring game to increase the difficulty of iterative reasoning. This has no effect on our estimates.

This paper is closely related to the literature on iterated dominance (Beard and Beil (1994), Schotter, Weigelt, and Wilson (1994), Van Huyck, Wildenthal, and Battalio (2002), Ho, Camerer, and Weigelt (1998), and Costa-Gomes, Crawford, and Broseta (2001)), as we estimate a subject's order of rationality using her choices in dominance solvable games.[10] The existing literature tends to focus on violations of iterated dominance, with the exception of Ho, Camerer, and Weigelt (1998) which identifies a subject's capability to perform a certain order of iterated dominance as in the level-k model.

The paper proceeds as follows. The next section provides an example that illustrates the challenge in identifying higher-order rationality from strategic choice data and motivates the ring game as a solution to the identification problem. Section 3 applies the tools of epistemic game theory to formally define rationality, higher-order rationality, and our exclusion restriction. Section 4 discusses the experimental design. The results are discussed in Section 5. Section 6 is a conclusion.

2. EXAMPLE

Consider the bimatrix game B1 in Figure 1 (the first matrix represents player 1's payoffs and the second matrix represents player 2's payoffs). If player 2 is rational, she must play action c since d is strictly dominated (d will never be a best response to any beliefs held by player 2). If player 1 is rational, she can play either a or b since a is the best response if she believes player 2

[9]For pioneering works in the level-k literature, see Stahl and Wilson (1994, 1995), Nagel (1995), Costa-Gomes, Crawford, and Broseta (2001), and Camerer, Ho, and Chong (2004). For a recent survey of this literature, see Costa-Gomes, Crawford, and Iriberri (2013).

[10]Iterated dominance and rationalizability are equivalent concepts in the games we consider.

Player 1 (rows: Player 1's actions; columns: Player 2's actions)

	c	d
a	15	0
b	5	10

Player 2 (rows: Player 2's actions; columns: Player 1's actions)

	a	b
c	10	5
d	5	0

FIGURE 1.—B1.

is playing *c* and *b* is the best response if she believes player 2 is playing *d*. If she satisfies second-order rationality (is rational and believes player 2 is rational), then she must play action *a* since she can only believe player 2 is playing *c* (since she believes player 2 is rational).

Suppose we observe a subject play the action *a* as player 1. She may have played *a* because she satisfies second-order rationality. However, we cannot rule out the possibility that the subject only satisfies rationality. This is because, in any game, the first-order rationalizable set necessarily contains the second-order rationalizable set.[11] This leads to an identification problem.

We resolve this problem by considering behavior in related sets of games under the natural exclusion restriction:[12] *subjects satisfying only lower-order rationality do not respond to changes in higher-order payoffs.* Consider the bimatrix game B2 in Figure 2. If player 2 is rational, she must play action *d*. If player 1 satisfies second-order rationality, she must play action *b*. In addition, B2 is related to B1 in a structured way: player 1 has the same payoffs in both B1 and B2. Thus, under the exclusion restriction, a subject that satisfies rationality but not second-order rationality will not respond to changes in first-order

Player 1 (rows: Player 1's actions; columns: Player 2's actions)

	c	d
a	15	0
b	5	10

Player 2 (rows: Player 2's actions; columns: Player 1's actions)

	a	b
c	5	0
d	10	5

FIGURE 2.—B2.

[11] If an action survives two rounds of iterated deletion of never best responses, then it obviously also survives one round.

[12] This exclusion restriction is natural if you believe that subjects that do not express higher-order rationality do not then base their decisions on higher-order information. This is elaborated on below.

payoffs.[13] Under this assumption, such a subject would play either (a, a) or (b, b) as player 1 in games B1 and B2, respectively; however, a subject who satisfies second-order rationality would play action profile (a, b).[14] Observing the action profile of a subject in both games B1 and B2 allows us to separate the behavioral implications of second-order rationality from rationality.

The exclusion restriction is empirically valid. In our experimental data, subjects follow the restrictions of this assumption 84 percent of the time. In addition, there are theoretical reasons to support this assumption. If subjects have finite depths of reasoning, subjects will not process higher-order information, and behavior will not depend on higher-order payoff information. For example, if a subject is rational (and not higher-order rational) because she has a depth of reasoning of 1, then she will not base her decision on any payoffs besides her own.[15]

Following this logic, the behavioral implications of kth-order rationality can be separated from lower orders of rationality by looking at games that differ only in $(k-1)$th-order payoffs but have different kth-order rationalizable implications. However, no such bimatrix games exist. Any two bimatrix games with the same payoffs up to the first order must be the same game (and hence have the same rationalizable implications).

This paper solves this problem by making use of a novel class of games: ring games. In a bimatrix game, player 1 and player 2 are each other's mutual opponent. But, in a 3-player ring game, player 1's opponent is player 2 and player 2 has an entirely different opponent, player 3. This unique opponent structure makes it possible to induce changes in higher-order payoffs independently of lower-order payoffs.

Consider the game R1 in Figure 3. Player 3 must play e if she is rational, player 2 must play c if she satisfies second-order rationality, and player 1 must play a if she satisfies third-order rationality (rational, believes her opponent is rational, and believes her opponent believes her opponent is rational). Game R2, in Figure 4, has different rationalizable implications: player 3 must play f if she is rational, player 2 must play d if she satisfies second-order rationality, and player 1 must play b if she satisfies third-order rationality.

Under the exclusion restriction, a subject that satisfies rationality or second-order rationality but not third-order rationality will not respond to changes in

[13] A player's zeroth-order payoffs are her own payoffs. Her first-order payoffs are her opponent's payoffs. Her second-order payoffs are her opponent's opponent's payoffs, and so on.

[14] We use the notation (s_1, s_2) to refer to a single player's action profile in two different games.

[15] There is empirical support for this type of behavior from the limited depth of reasoning literature. Costa-Gomes, Crawford, and Broseta (2001), Costa-Gomes and Crawford (2006), Wang, Spezio, and Camerer (2009), Brocas, Carrillo, Wang, and Camerer (2014), Johnson, Camerer, Sen, and Rymon (2002), and Johnson, Camerer, Rymon, and Sen (1993) all analyzed strategic behavior by investigating the information search pattern of subjects. They found a correlation between the play of kth-order rationalizable strategies and patterns of information search that are associated with k depths of reasoning.

Player 1

Player 1's actions \ Player 2's actions	c	d
a	15	0
b	5	10

Player 2

Player 2's actions \ Player 3's actions	e	f
c	15	0
d	5	10

Player 3

Player 3's actions \ Player 1's actions	a	b
e	10	5
f	5	0

FIGURE 3.—R1.

second-order payoffs (represented by player 3's payoffs). Under this assumption, such a subject would play either (a, a) or (b, b) as player 1 in games R1 and R2, respectively; however, a subject who satisfies third-order rationality would play action profile (a, b). Thus, observing the behavior of player 1 in R1 and R2 allows us to separate the implications of third-order rationality from lower orders of rationality. Each additional player in the ring game allows an additional degree of independence between orders of payoffs and allows us to separately identify an additional order of rationality.

3. THE MODEL

Now we will define rationality, higher-order rationality, and our exclusion restriction more formally. We first define an n-player ring game.

DEFINITION: An *n-player ring game* Γ is a tuple $\Gamma = \langle I = \{1, \dots, n\}; S_1, \dots, S_n; \pi_1, \dots, \pi_n; o\rangle$ where I and S_i are finite sets, $\pi_i : S_i \times S_{o(i)} \to \mathbb{R}$, and $o : I \to I$ with $o(i) = 1 + i \bmod n$.

The set I represents the set of players, S_i represents the set of actions for player i, π_i represents the payoffs for player i which depend upon player i's action and the action of her opponent $o(i)$, and o represents the opponent mapping function where $o(i)$ is the opponent of player i. The restriction $o(i) = 1 + i \bmod n$ restricts the opponent relationship to that of a ring:

Player 1

Player 1's actions \ Player 2's actions	c	d
a	15	0
b	5	10

Player 2

Player 2's actions \ Player 3's actions	e	f
c	15	0
d	5	10

Player 3

Player 3's actions \ Player 1's actions	a	b
e	5	0
f	10	5

FIGURE 4.—R2.

player 1's opponent is player 2, player 2's opponent is player 3, and so on, with player n's opponent being player 1.

Given a game, an epistemic type space describes players' beliefs about strategies.[16]

DEFINITION: Let $\Gamma = \langle I; S_1, \dots, S_n; \pi_1, \dots, \pi_n; o\rangle$ be an n-player ring game. A finite Γ-based *epistemic type space* is a tuple $\langle T_1, \dots, T_n; b^1, \dots, b^n; \hat{s}_1, \dots, \hat{s}_n\rangle$ where T_i is a finite set, $b^i : T_i \to \Delta(T_{o(i)})$, and $\hat{s}_i : T_i \to \Delta(S_i)$.

The set T_i is the set of types of player i. The function $\hat{s}_i$ defines a strategy for each type of player i by mapping each type to a probability distribution over S_i. And, the function b^i represents each type's beliefs about the types of her opponent by mapping each type to a probability distribution over $T_{o(i)}$. Thus, the function $b^i(t_i)$ together with the function $\hat{s}_{o(i)}$ defines type t_i's beliefs about the strategies of her opponent.

The expected utility of a type t_i can be defined[17]

$$u_i(s_i, t_i) \equiv \sum_{t_{o(i)} \in T_{o(i)}} b^i(t_i)(t_{o(i)})\pi\big(s_i, \hat{s}(t_{o(i)})\big).$$

Given the specification for expected utility, a type is rational if the strategy $\hat{s}_i(t_i)$ is a best response for player t_i given her beliefs.

DEFINITION: A type t_i is *rational* if $\hat{s}_i(t_i)$ maximizes player i's expected payoff under the measure $b^i(t_i)$, that is, if, for any $s \in \text{supp}(\hat{s}_i(t_i))$,

$$u_i(s, t_i) \geq u_i\big(s', t_i\big) \quad \forall s' \in S_i.$$

To define higher-order rationality, we first define what we mean by belief. We say a type believes an event if she places probability 1 on that event happening.

DEFINITION: A type t_i *believes* an event $E \subseteq T_{o(i)}$ if $b^i(t_i)(E) = 1$. Let the set $B^i(E) = \{t_i \in T_i | b^i(t_i)(E) = 1\}$ be the set of types for player i that believe event E.

Higher-order rationality is then defined recursively in the following way:

$$R_i^1 = \{t_i \in T_i | t_i \text{ is rational}\},$$
$$R_i^{m+1} = R_i^m \cap B^i\big(R_{o(i)}^m\big).$$

DEFINITION: If $t_i \in R_i^m$, then we say that t_i satisfies mth-*order rationality*.

[16]See Aumann and Brandenburger (1995) for development of the epistemic framework used here.

[17]Let $\pi(s_i, \hat{s}(t_{o(i)}))$ be redefined in the standard way whenever $\hat{s}$ is a mixed-strategy.

3.1. *Identifying Orders of Rationality*

To state our exclusion restriction, we first formalize what is meant by higher-order payoffs.

Every ring game can be characterized by its payoff hierarchy for each player. For a given ring game $\Gamma = \langle I = \{1, \dots, n\}; S_1, \dots, S_n; \pi_1, \dots, \pi_n; o\rangle$, the payoff hierarchy $\bar{h}^i(\Gamma)$ of player i is defined by

$$\bar{h}^i(\Gamma) = \{\pi_{o^k(i)}\}_{k=0}^{\infty}.$$

The payoffs of player i are given by $\bar{h}^i_0(\Gamma)$, her first-order payoffs are given by $\bar{h}^i_1(\Gamma)$ (i.e., the payoffs of player $o(i)$), her second-order payoffs are given by $\bar{h}^i_2(\Gamma)$ (i.e., the payoffs of player $o^2(i)$), and so on.

Our identification strategy exploits the flexible payoff structure of the ring game. The assumption used to identify a player's order of rationality requires a player that satisfies kth-order rationality but not $(k+1)$th-order rationality to not respond to changes in her mth-order payoffs whenever m is greater than or equal to k. Assumption ER is our exclusion restriction:

ER: *A player i who satisfies kth-order rationality but not $(k+1)$th-order rationality $(k \geq 1)$ will play the same action in any two games, Γ and Γ', where $\bar{h}^i_j(\Gamma) = \bar{h}^i_j(\Gamma')$ for all $j < k$.*

4. EXPERIMENTAL DESIGN

We can use the preceding results to empirically identify orders of rationality from strategic choice data. Each subject's order of rationality is estimated from choice data in eight 4-player ring games defined by G1 and G2 in Figures 5 and 6 (2 games $\times$ 4 player positions). The games are similar to the games R1 and R2 from Section 2 except that G1 and G2 are 4-player, 3 action games. Adding an additional player permits identification of a player's order of ratio-

Player 1

Player 1's actions \ Player 2's actions	a	b	c
a	8	20	12
b	0	8	16
c	18	12	6

Player 2

Player 2's actions \ Player 3's actions	a	b	c
a	14	18	4
b	20	8	14
c	0	16	18

Player 3

Player 3's actions \ Player 4's actions	a	b	c
a	20	14	8
b	16	2	18
c	0	16	16

Player 4

Player 4's actions \ Player 1's actions	a	b	c
a	12	16	14
b	8	12	10
c	6	10	8

FIGURE 5.—G1.

Player 1

Player 2's actions

Player 1's actions	a	b	c
a	8	20	12
b	0	8	16
c	18	12	6

Player 2

Player 3's actions

Player 2's actions	a	b	c
a	14	18	4
c	20	8	14
c	0	16	18

Player 3

Player 4's actions

Player 3's actions	a	b	c
a	20	14	8
b	16	2	18
c	0	16	16

Player 4

Player 1's actions

Player 4's actions a	a	b	c
a	8	12	10
b	6	10	8
c	12	16	14

FIGURE 6.—G2.

nality up to the fourth order. Adding an additional action lowers the likelihood of misidentification.[18]

Under the exclusion restriction ER, these eight games separate subjects into five rationality categories: irrational (R0), rational but not second-order rational (R1), second-order rational but not third-order rational (R2), third-order rational but not fourth-order rational (R3), and fourth-order rational (R4). The predicted action profiles for each of the types, R1–R4, are given in Table I. The eight games are defined by either game G1 or G2 and the player position. The action profile (x, y) represents the actions played by the given player in games G1 and G2, respectively. An R4 subject always plays the rationalizable profiles when she plays games G1 and G2 as any player (i.e., she must play (a, c) when she plays as player 1, (b, a) when she plays as player 2, (a, b) as player 3, and (a, c) as player 4). An R3 subject must play the rationalizable profiles when she plays as player 2, 3, or 4. She can play any action, a, b, or c, when she plays as player 1, but must play the same action in G1 and G2 (i.e., (a, a), (b, b), or (c, c)). An R2 subject must play the rationalizable profiles when she plays as player 3 or 4. She can play any action, a, b, or c, as player 1

TABLE I

PREDICTED ACTIONS UNDER RATIONALITY AND ASSUMPTIONS ER IN THE EIGHT GAMES

	Games			
	P1	P2	P3	P4
Type	Gl G2	Gl G2	Gl G2	Gl G2
R1	$(a, a)(b, b)(c, c)$	$(a, a)(b, b)(c, c)$	$(a, a)(b, b)(c, c)$	(a, c)
R2	$(a, a)(b, b)(c, c)$	$(a, a)(b, b)(c, c)$	(a, b)	(a, c)
R3	$(a, a)(b, b)(c, c)$	(b, a)	(a, b)	(a, c)
R4	(a, c)	(b, a)	(a, b)	(a, c)

[18]Moving to three actions is enough to ensure that the likelihood of a subject who is playing randomly getting assigned as a rational type is quite small. This is discussed in more detail below.

or 2, but must play the same action in G1 and G2. An R1 subject must play the rationalizable profile when she plays as player 4. And, she can play any action, a, b, or c, as player 1, 2, or 3, but must play the same action in G1 and G2.

Each subject can play three possible actions in each of the eight games for a total of 6561 possible action profiles. There are 40 action profiles that are exact matches to the predicted action profiles of types R1–R4. If a subject's action profile matches one of the predicted action profiles of type R1–R4 exactly, then we assign that subject as that type. Additionally, if a subject's action profile deviates from an action profile of type R1–R4 in only one of the eight games (one error), then we assign that subject as that type. If a subject's action profile is within one error of two types, we assign that subject as the lower type (assignment to the lower type implicitly assumes that errors due to assumption ER are more likely than errors due to rationality).[19] An additional 255 action profiles are one error away from the predicted action profiles of types R1–R4. Thus, it is unlikely for a subject to be assigned to a rational type by random chance. Only 5 percent of the action profiles (295 out of 6561) would get a subject assigned to an R1–R4 type. Playing any of the other 95 percent of action profiles (6266 of 6561) would get a subject assigned to the irrational (R0) type.

Our classification is conservative: it is harder to be assigned to a higher-order type than a lower-order type. Only 11 of the 6561 action profiles would get a subject assigned to an R4 type, versus 26 for R3 types, 78 for R2 types, 180 for R1 types, and 6266 for R0 types. Regardless, our results are not overly sensitive to the classification assumptions as most of the subjects in our sample will be assigned based on exactly matching one of the predicted action profiles of R1–R4 types.[20]

Our experiment consists of two treatments: a main treatment and a robustness treatment. Subjects play the same eight games in both treatments. However, the games are presented differently. In the main treatment, the payoff matrices were presented in a particular order, with a subject's own payoff matrix being the leftmost matrix, followed by her opponent's payoff matrix as the

[19] We additionally assume that a subject has to play the rationalizable profile as player 4 to be matched to R1.

[20] Since assignment based on exact match is quite high, we should not be too concerned about misidentification due to our classification assumptions. Thus, the only other source of misidentification is the failure of ER. One situation in which ER could fail is when a player is holding beliefs that make her indifferent between actions. In this case, we would not necessarily expect a player to play a deterministic profile, (a, a), (b, b), or (c, c), but instead mix between actions. If this is the case, she could end up playing the rationalizable profile with positive probability, which would lead to misidentification (assigning a subject to a higher-order type when she is not). However, we can assess this likelihood by looking at the rate at which the reverse-rationalizable profile is played (i.e., if (x, y) is the rationalizable profile, then (y, x) is the reverse-rationalizable profile). If a player is mixing, then the reverse-rationalizable profile will be played with the same probability as the rationalizable profile. Less than 4 percent of subjects in our data set play the reverse-rationalizable profile in any player position. This suggests that misidentification due to stochastic choice should not be a concern.

second, and so on (i.e., if a subject was playing the game defined by G1 and player 1, then the game was presented as in Figure 5 with payoff matrices ordered P1, P2, P3, P4). In the robustness treatment, the payoff matrices were presented in a random order (e.g., for the game defined by G1 and player 1, a subject's own payoff matrix might be the rightmost matrix, with the other three matrices presented in some random order (i.e., P2, P4, P3, P1)).[21] The two treatments are motivated in Section 5.

4.1. *Laboratory Implementation*

Sessions were conducted in Arts ISIT computer labs with undergraduate students at the University of British Columbia. Subjects were recruited through the Online Recruitment System for Economic Experiments (ORSEE) (Greiner (2004)). No subject participated in more than one session. Subjects made all decisions through an online interface. In order to ensure independence across subjects, subjects did not interact with one another during the experiment and were not informed of one another's decisions.

Each subject played the games G1 and G2 in each of the player positions, for a total of eight games.[22] Subjects played the games in a random order without feedback. Subjects were required to spend at least 90 seconds on each of the games. Once subjects made choices in all games, they were given the opportunity to revise their choices (without feedback).

One game was randomly selected for payment at the end of the experiment. Subjects were randomly and anonymously matched into 4-player groups and paid based on their choice and the choices of their group members in the selected game. Subjects received the dollar value of their payoff in the selected game. The average session lasted 45 minutes and the average subject earned approximately $17 dollars (maximum payment was $25 and minimum payment was $7), including a showup fee of $5. Payments were in Canadian dollars.

Instructions were read aloud by the experimenter at the beginning of the session. Instructions to all subjects were the same. Subjects then completed a short quiz to make sure they understood the instructions. The instructions and quiz can be found in the Supplemental Material (Kneeland (2015)).

The experiment consists of 116 subjects. The main treatment contains 80 subjects gathered over six sessions. The robustness treatment contains 36 subjects gathered over three sessions.

[21]In this treatment, the eight games were randomly presented in one of four different orders: (P1, P4, P3, P2), (P4, P1, P2, P3), (P2, P3, P1, P4), or (P3, P2, P4, P1).

[22]In the main treatment, six additional games were played by each subject. Those games were designed to assess additional features of strategic reasoning. The data from those games are not analyzed in this paper.

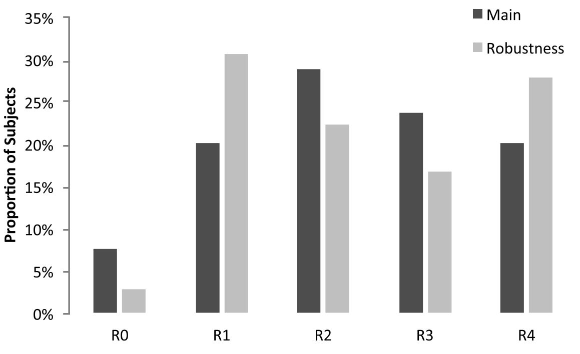

FIGURE 7.—Subjects classified by order of rationality, by treatment.

5. EXPERIMENTAL RESULTS

Figure 7 reports the proportion of subjects classified as each type in both the main and robustness treatments. Combining treatments, 6 percent of subjects are classified as R0 types, 23 percent are classified as R1 types, 27 percent are classified as R2 types, 22 percent are classified as R3 types, and 22 percent are classified as R4 types (there are no significant differences between treatments: Fisher exact test, p-value $= 0.580$).[23]

Most of our subjects—79 of the 116 subjects—made decisions that matched one of the 40 action profiles that yield an exact type match. This high rate of exact classification cannot be random[24] and strongly supports the assumptions of ER and rationality that undergird our empirical strategy. Importantly, 93 percent of R3 and R4 subjects were classified by exact match. Thus, different assumptions made about type classification under error would not decrease the proportion of higher-order types.

The results are surprising. We find considerably more weight on higher-order types, R3–R4, than the level-k literature typically finds.[25] The motivating reason behind the use of ring games is that they allow us to use weaker identification assumptions. This gives us more direct and therefore more reliable results. One concern, however, is that the change in game form also has the potential to change other determinative factors of strategic decision making. Indeed, any

[23] Of the 116 subjects, 27 failed the quiz. Subjects were more likely to fail if they had a lower order of rationality. Of the R0 subjects, 71 percent failed, while only 8 percent of the R4 subjects failed. This suggests that irrational (R0) subjects may not have had a clear understanding of the games they were asked to play.

[24] If subjects were playing randomly, the odds of them playing any of the actions predicted by R1–R4 types are less than half a percent. If this were the case, you would expect to see at most 1 of the 116 subjects getting assigned as an R1–R4 type through exact match.

[25] For examples, see Costa-Gomes and Crawford (2006), Costa-Gomes, Crawford, and Broseta (2001), and Nagel (1995). Though there are some exceptions; for example, Arad and Rubinstein (2012) found the most weight on R3 types.

change to a strategic environment has the potential to change the framing of choices and even the ease of expressing rationality. For that reason, we must be cautious in drawing conclusions about the generality of these (or any) results across games. 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. One way of examining whether the high orders of rationality measured in our ring games are caused by such reductions in the cost of iterative reasoning is to perturb ring games in such a way as to decrease the transparency of the dominance structure of the game. This is just what we do in the robustness treatment.

Surprisingly, as Figure 7 shows, perturbing the game to obfuscate the dominance structure has no significant effect on behavior and, in particular, does not change our finding that a substantial proportion of subjects are of higher-order types. Subjects are as likely to be classified as a higher-order type (type R3 or R4) in the main treatment as in the robustness treatment (44 percent in both treatments) and as likely to be classified as a lower-order type (type R1 or R2) in the main treatment as in the robustness treatment (49 and 53 percent, respectively). In fact, subjects are modestly *more likely* to be classified as R4 in the robustness treatment than in the main treatment. These results suggest that our findings are not driven by an increased ease of iterative reasoning in ring games. However, we encourage additional robustness tests along these lines in future research. The interpretation of the robustness treatment as increasing the cost of iterative reasoning is only one interpretation and it may be that costs do not actually increase between treatments. Furthermore, little is known about rationality and behavior in asymmetric n-player games in general, as the study of 2-player games has dominated the focus of experimental research. Future research will be necessary to develop our understanding of the behavioral relationships between different game forms.

6. CONCLUDING REMARKS

This paper develops a novel experimental design based on ring games to identify higher-order rationality from strategic choice data. Ring games allow for the application of a natural exclusion restriction due to the flexibility of their payoff structure. This design has advantages over other approaches used in the literature because it allows us to make reliable choice based inferences about higher-order reasoning under weak identification assumptions.

Our experimental design has a number of other interesting applications. As discussed above, different features like the level of payoffs, the complexity of the game, or even the framing of the game may affect strategic decision making and hence the expression of rationality. Understanding how these different features affect reasoning is important both for understanding the differences in behavior across games and for increasing our understanding of behavior in the real world. By perturbing these features within the ring game, we can investigate how they affect higher-order reasoning.

Additionally, since our design allows us to simply and cleanly identify orders of rationality, it can easily be applied to study the relationship between higher-order rationality and other features of interest, for example, whether order of rationality is related to cognitive abilities, preference features like loss aversion or present bias, or to behavior in decision theoretic problems.

Our design can even be used to investigate assumptions about game theoretic models like the level-k model. Much of the predictive power of the level-k model stems from the level-0 assumption. But, tests of the level-k model typically assume that level-0 behavior is known to the experimenter and is the same for all players. However, our experimental design would let the experimenter test both her assumptions about level-0 behavior and the heterogeneity of the level-0 model across subjects. This is because our ER assumption and the ring game framework allows us to characterize a subject's depth of reasoning without making any assumptions about the beliefs a subject holds, nor does it require beliefs to be the same for all subjects. Thus, observing the behavior of a subject who does not respond to changes in higher-order payoffs will reveal information about her underlying beliefs, which we can interpret as the underlying level-0 behavior.

REFERENCES

ARAD, A., AND A. RUBINSTEIN (2012): "The 11–20 Money Request Game: Evaluating the Upper Bound of Level-k Reasoning," *American Economic Review*, 102 (7), 3561–3573. [2076]

AUMANN, R., AND A. BRANDENBURGER (1995): "Epistemic Conditions for Nash Equilibrium," *Econometrica*, 63, 1161–1180. [2071]

BEARD, T. R., AND R. O. BEIL (1994): "Do People Rely on the Self-Interested Maximization of Others? An Experimental Test Management Science," *Management Science*, 40 (2), 252–262. [2067]

BERNHEIM, D. (1984): "Rationalonalizable Strategic Behavior," *Econometrica*, 52, 1007–1028. [2066]

BROCAS, I., J. D. CARRILLO, S. W. WANG, AND C. F. CAMERER (2014): "Imperfect Choice or Imperfect Attention? Understanding Strategic Thinking in Private Information Games," *Review of Economic Studies*, 81, 944–970. [2069]

BURCHARDI, K. B., AND S. P. PENCZYNSKI (2014): "Out of Your Mind: Eliciting Individual Reasoning in One Shot Games," *Games and Economic Behavior*, 84, 39–57. [2066]

CAMERER, C. F., T.-H. HO, AND J.-K. CHONG (2004): "A Cognitive Hierarchy Model of Games," *Quarterly Journal of Economics*, 119 (3), 861–898. [2067]

COSTA-GOMES, M., V. P. CRAWFORD, AND B. BROSETA (2001): "Cognition and Behavior in Normal-Form Games: An Experimental Study," *Econometrica*, 69 (5), 1193–1235. [2067,2069, 2076]

COSTA-GOMES, M. A., AND V. P. CRAWFORD (2006): "Cognition and Behavior in Two-Person Guessing Games: An Experimental Study," *American Economic Review*, 96 (5), 1737–1768. [2069,2076]

COSTA-GOMES, M. A., AND G. WEIZSÄCKER (2008): "Stated Beliefs and Play in Normal-Form Games," *The Review of Economic Studies*, 75, 729–762. [2065]

CRAWFORD, V. P., M. A. COSTA-GOMES, AND N. IRIBERRI (2013): "Structural Models of Nonequilibrium Strategic Thinking: Theory, Evidence, and Applications," *Journal of Economic Literature*, 51, 5–62. [2067]

CUBITT, R. P., AND R. SUGDEN (1994): "Rationally Justifiable Play and the Theory of Non-Cooperative Games," *The Economic Journal*, 104, 798–803. [2066]

GREINER, B. (2004): "The Online Recruitment System ORSEE 2.0—A Guide for the Organization of Experiments in Economics," Working Paper Series in Economics, 10, University of Cologne. [2075]

HEALY, P. J. (2011): "Epistemic Foundations for the Failure of Nash Equilibrium," Working Paper. [2065]

HO, T.-H., C. CAMERER, AND K. WEIGELT (1998): "Iterated Dominance and Iterated Best Response in Experimental 'p-Beauty Contests'," *American Economic Review*, 88 (4), 947–969. [2067]

JOHNSON, E. J., C. CAMERER, T. RYMON, AND S. SEN (1993): "Cognition and Framing in Sequential Bargaining for Gains and Losses," in *Frontiers of Game Theory*. Cambridge: MIT Press, 27–47. [2069]

JOHNSON, E. J., C. CAMERER, S. SEN, AND T. RYMON (2002): "Detecting Failures of Backward Induction: Monitoring Information Search in Sequential Bargaining," *Journal of Economic Theory*, 104, 16–47. [2069]

KNEELAND, T. (2015): "Supplement to 'Identifying Higher-Order Rationality'," *Econometrica Supplemental Material*, 83, http://dx.doi.org/10.3982/ECTA11983. [2075]

NAGEL, R. (1995): "Unraveling in Guessing Games: An Experimental Study," *American Economic Review*, 85 (5), 1313–1326. [2067,2076]

PEARCE, D. (1984): "Rationalizable Strategic Behavior and the Problem of Perfection," *Econometrica*, 52, 1029–1050. [2066]

SCHOTTER, A., K. WEIGELT, AND C. WILSON (1994): "A Laboratory Investigation of Multiperson Rationality and Presentation Effects," *Games and Economic Behavior*, 6, 445–468. [2067]

STAHL, D. O., AND P. W. WILSON (1994): "Experimental Evidence on Player's Models of Other Players," *Journal of Economic Behavior and Organization*, 25 (3), 309–327. [2067]

STAHL, D. O., AND P. W. WILSON (1995): "On Player's Models of Other Players: Theory and Experimental Evidence," *Games and Economic Behavior*, 10 (1), 218–254. [2067]

TAN, T. C.-C., AND S. R. C. WERLANG (1988): "The Bayesian Foundations of Solution Concepts of Games," *Journal of Economic Theory*, 45, 370–391. [2066]

VAN HUYCK, J. B., J. M. WILDENTHAL, AND R. C. BATTALIO (2002): "Tacit Cooperation, Strategic Uncertainty, and Coordination Failure: Evidence From Repeated Dominance Solvable Games," *Games and Economic Behavior*, 38 (1), 156–175. [2067]

WANG, J. T.-Y., M. SPEZIO, AND C. F. CAMERER (2009): "Pinocchio's Pupil: Using Eyetracking and Pupil Dilation to Understand Truth Telling and Deception in Sender–Receiver Games," *American Economic Review*, 100 (3), 984–1007. [2069]

Dept. of Economics, University College London, Gower Street, London WC1E 6BT, U.K.; t.kneeland@ucl.ac.uk.

Manuscript received October, 2013; final revision received April, 2015.