Econometrica, Vol. 82, No. 3 (May, 2014), 1133–1151

NOTES AND COMMENTS

HOW PORTABLE IS LEVEL-0 BEHAVIOR? A TEST OF LEVEL-*k* THEORY IN GAMES WITH NON-NEUTRAL FRAMES

BY SHAUN HARGREAVES HEAP, DAVID ROJO ARJONA, AND ROBERT SUGDEN[1]

We test the portability of level-0 assumptions in level-*k* theory in an experimental investigation of behavior in Coordination, Discoordination, and Hide and Seek games with common, non-neutral frames. Assuming that level-0 behavior depends only on the frame, we derive hypotheses that are independent of prior assumptions about salience. Those hypotheses are not confirmed. Our findings contrast with previous research which has fitted parameterized level-*k* models to Hide and Seek data. We show that, as a criterion of successful explanation, the existence of a plausible model that replicates the main patterns in these data has a high probability of false positives.

KEYWORDS: Level-*k* theory, Hide and Seek, coordination, discoordination.

1. INTRODUCTION

TRADITIONALLY, GAME THEORY ANALYZES the interaction of ideally rational agents whose rationality is common knowledge. Recently, however, there has been growing interest in investigating how *in fact* human agents reason in strategic situations. In such work, the most widely used approach is that of *level-k theory* (Stahl and Wilson (1994), Nagel (1995), Costa-Gomes, Crawford, and Broseta (2001)) and the closely related *cognitive hierarchy theory* (Camerer, Ho, and Chong (2004)). These theories distinguish types of players according to the level at which they reason. Assumptions about level-0 ($L0$) behavior provide an anchor for beliefs and behavior at higher levels. At each higher level, players are assumed to know the probability distributions of choices at lower levels. $L1$ players choose best replies to $L0$ choices; $L2$ players choose best replies to $L1$ choices (or, in cognitive hierarchy theory, to a probability mix of $L0$ and $L1$ choices); and so on.

In some applications, $L0$ players are assumed to pick strategies at random, but for many games this parsimonious assumption yields inaccurate predictions. A standard practice in level-*k* modeling is to tailor the $L0$ specification to the particular game being analyzed, subject to the general principle that it can be interpreted as "a strategically naïve initial assessment of others' likely responses to the game" (Crawford, Costa-Gomes, and Iriberri (2013, p. 14)). This may be a useful first step in theory development, but if level-*k* theory is

[1]We thank Vincent Crawford, Nagore Iriberri, Ariel Rubinstein, participants in many conferences and seminars, an editor, and four anonymous referees for comments on earlier versions of this paper. Hargreaves Heap's and Sugden's work was supported by the Economic and Social Research Council through the Network for Integrated Behavioural Science (Grant reference ES/K002201/1).

 DOI: 10.3982/ECTA11132

to have explanatory power, it must be based on well-defined hypotheses about $L0$ behavior that apply across a reasonably wide range of games. Further, to avoid circularity, there has to be a clear conceptual distinction between the imperfectly rational strategic thinking that the theory is intended to explain and the kind of thinking that can be attributed to $L0$ by assumption. The clearest way to make this distinction is to require, as level-k theorists sometimes do, that $L0$ behavior responds only to *non-strategic* features of games.[2] The level-k research program rests on implicit confidence that such hypotheses can be developed, and that they will generate successful predictions. But in the absence of concrete proposed hypotheses, how can one assess whether that confidence is justified? This is the question we address.

Existing tests of level-k theory are not particularly helpful for that assessment. Many such tests have used an *ex post model-fitting* approach. That approach works by specifying a parametric level-k model of behavior in a particular game and by fitting that model to (usually experimental) data. The $L0$ specification is often specific to the game, and justified by appeals to intuition; sometimes (e.g., Crawford and Iriberri (2007), Penczynski (2013)) it is partly determined in the process of model-fitting. Level-k theory is then judged in terms of the model's goodness-of-fit, as compared with that of parametric models derived from alternative theories (Crawford, Costa-Gomes, and Iriberri (2013) reviewed many such exercises). Other studies, such as the analyses of coordinated attack games by Strzalecki (2010) and Kneeland (2012), have derived empirically confirmed qualitative implications from level-k models. In these models, too, assumptions about $L0$ behavior are justified by their intuitive plausibility as representations of naïve play in the relevant games. These ways of testing level-k theory, therefore, have a common limitation. At most, they can show that in each of a variety of games *considered separately*, players' behavior is consistent with *some* plausible assumption about $L0$ behavior.[3]

In this paper, we use a different method to investigate the portability across games of assumptions about non-strategic $L0$ behavior. The essential idea is to use simple two-player games in which, after abstracting from properties of labeling, the pure strategies for each player (considered separately) are isomorphic with one another. Strategies are distinguished *only* by their labels, which differ in ways that can be expected to be salient to naïve players. The "frame"

[2]For example, Crawford, Gneezy, and Rottenstreich (2008, p. 1450) explained an $L0$ specification for coordination games by saying "Bearing in mind that $L0$ is only the starting point for players' strategic thinking, we define it via nonstrategic, behaviorally plausible general principles."

[3]A complementary research strategy, which does not depend on specific assumptions about $L0$, investigates whether players' reasoning processes have the iterated best response structure of level-k theory. Experiments that track players' sequential use of information (Costa-Gomes, Crawford, and Broseta (2001)), elicit "provisional choices" (Agranov, Caplin, and Tergiman (2012)), and analyze players' justifications of "suggested decisions" (Penczynski (2013)) have found supporting evidence. However, a level-k model of iterated reasoning in any particular game still requires assumptions about $L0$, and the portability of these assumptions remains an open question.

(i.e., the number of strategies and their labels) is the same for both players. It is natural to assume that, in games of this kind, the behavior of a non-strategic $L0$ player depends only on the frame. Any such $L0$ specification is thus directly portable across player roles in a given game, and across strategically different games that have the same frame. Our experiment tests cross-role and cross-game predictions of level-k theory that are conditional on the portability of the $L0$ specification. Crucially, we do not need to make any substantive assumptions about that specification itself. Our methodological approach is one of ex ante hypothesis testing, rather than ex post model-fitting.

Of course, we cannot claim that the portability property we test is implied by every possible general hypothesis about $L0$ behavior. But any such hypothesis must have *some* well-defined implications about portability across strategically distinct games and player roles. If $L0$ specifications were not portable even within our tightly restricted sets of games, it would seem hard to be optimistic about finding such a general theory of $L0$ behavior.

In Section 2, we define the class of games to which our analysis applies, and state the joint hypothesis that we will test to assess the portability of $L0$ assumptions. We derive implications of that hypothesis for behavior in Coordination, Discoordination, and Hide and Seek games that share the same frame. In Section 3, we explain the experimental design we used to test these implications. It is based on that used in a series of experiments conducted by Rubinstein and Tversky (1993), Rubinstein, Tversky, and Heller (1996), and Rubinstein (1999), which we describe collectively as the work of "RTH." In Section 4, we report our results. These provide little support for the joint hypothesis. In Section 5, we consider how our findings can be reconciled with apparently conflicting conclusions drawn under the ex post model-fitting methodology. We focus on Crawford and Iriberri (2007; henceforth "CI") because they fitted a level-k model to data from some of RTH's Hide and Seek games. In the final section, we discuss the wider implications of our results.

2. THEORY

We investigate behavior in a class of two-player simultaneous-move games, defined as follows. For each player, there are m alternative pure strategies, where $m \geq 3$.[4] There is a set $F = \{l_1, \dots, l_m\}$ of distinct *labels*, such that, for each player, each strategy is uniquely identified by one of these labels; this set is the *frame*. We consider three different payoff matrices for such games. In a *Hide and Seek* game, if both players' chosen strategies have the same label, one player (the *hider*) gets a payoff of 0 and the other (the *seeker*) gets 1; otherwise, the hider gets 1 and the seeker gets 0. In a *Coordination* game, both players get

[4]We do not include the case where $m = 2$ because, following a common practice in the study of Hide and Seek, we want to use frames in which one label is perceived as being different from all the other labels in some particularly salient respect.

a payoff of 1 if their chosen strategies have the same label and 0 if not. In a *Discoordination* game, they both get 1 if their chosen strategies have different labels and 0 if not.[5]

For each of the four player roles (coordinator, discoordinator, hider, and seeker) considered separately, every pure strategy is isomorphic with every other in terms of payoffs; strategies are distinguished only by their labels. In such a case, it is natural to assume that $L0$ behavior is the same for all roles. As described by level-k theory, strategic reasoning takes a similar form for players in all roles. (At each level above $L0$, coordinators and seekers choose the label that is the most frequent choice of co-players at the level below; discoordinators and hiders choose the label that is the least frequent choice of co-players at the level below.) As a starting point for analysis, we therefore assume that play is no more or less sophisticated in any one of these roles than in any other.

In accordance with the preceding arguments, we state three hypotheses about behavior in any Coordination, Discoordination, and Hide and Seek games that share a common frame $F = \{l_1, \dots, l_m\}$ and that are played by individuals drawn at random from the same population:

H1. Level-k theory holds.

H2. The probability distribution of choices at $L0$ over the labels $l_1, \dots, l_m$ is the same for all four player roles.

H3. The population distribution of players across levels $L0$, $L1$, ... is the same for all four player roles.

Our objective is to test the *joint hypothesis* "H1 and H2 and H3."[6] H2 is the portability property that is the central concern of this paper. As we explain later, the predictions we test are not particularly sensitive to H3.

Our experiment investigates *odd-one-out* frames, constructed so that one label, the *oddity* (denoted by l_1), is clearly different from the others. (In one frame, for example, the four labels are the words "love," "hate," "detest," and "dislike"; in another, they are "rude," "polite," "honest," and "friendly.") It is natural to expect that if $L0$ choices are not uniformly random, the main deviation from randomness will be that the probability of choosing l_1 is particularly high or particularly low. In principle, the direction of deviation might depend on the connotations of the oddity relative to the other labels (positive in the case of "love," negative in the case of "rude").[7] However, we do not need to

[5]How far level-k theory explains behavior in these games is an important question. For example, that theory is one of the main competing approaches to the so far unresolved problem of explaining focal points in Coordination games (see, e.g., Mehta, Starmer, and Sugden (1994), Crawford, Gneezy, and Rottenstreich (2008), Bardsley, Mehta, Starmer, and Sugden (2010)).

[6]"Level-k theory" in H1 is to be understood as excluding cognitive hierarchy theory. The predictions we test, and hence our findings about the portability of $L0$ specifications, are specific to level-k theory, narrowly defined. The problem of finding portable $L0$ specifications is just as acute for cognitive hierarchy theory, but that is not the topic of the current paper.

[7]We treat the *connotations* of a label—that is, the affective qualities of the things to which it refers—as distinct from the *payoffs* that result from the choice of that label. The hypothesis that

make assumptions about the direction of deviation. Given H2, level-k theory implies that in any Coordination game, all types $L1$, $L2$, ... choose the label that is the modal choice at $L0$. Thus, if, in any given frame, the oddity is the modal choice of coordinators, we can infer (conditional on the truth of the joint hypothesis) that the oddity is also the modal choice at $L0$. That inference is valid whatever the connotations of the oddity. We will restrict our analysis to frames with this property.

In constructing frames, we tried to ensure that no non-oddity stood out in a way that would make it uniquely *dis*favored at $L0$. We test whether we achieved this by investigating the distribution of discoordinators' choices over non-oddities. If some non-oddity label, say l_m, were chosen at $L0$ with strictly lower probability than every other label, it would be the uniquely optimal choice at $L1$, $L3$, ..., and choices at $L2$, $L4$, ... would be distributed over the other labels. Thus, under any plausible assumption about the distribution of players across levels, the overall distribution of discoordinators' choices would have a spike at l_m.[8]

For any given odd-one-out frame, we will say that $L0$ choices have an *odd-one-out distribution* if l_1 is chosen with some probability $q > 1/m$, and if each other l_j is chosen with probability $(1 - q)/(m - 1)$. We submit that *if* the experimental data show that l_1 is the modal choice of coordinators (*Validation Condition 1*), *and if* they show no systematic non-randomness in the distribution of discoordinators' choices over $l_2, \dots, l_m$ (*Validation Condition 2*), *then* in deriving implications of the joint hypothesis, it is reasonable to model $L0$ choices by an odd-one-out distribution.

We now consider the implications of the joint hypothesis, conditional on $L0$ choices having an odd-one-out distribution. Notice this condition is not an *assumption* about $L0$ behavior; it is a property that can be inferred from satisfaction of the Validation Conditions. We interpret "level-k theory" in H1 as including the assumption (also made by CI) that the relative frequency of the $L0$ type is zero. For ease of exposition, we assume that there are no players at levels higher than $L4$.[9] The relative frequencies of $L1$, $L2$, $L3$, and $L4$ types are denoted π_1, π_2, π_3, and π_4, respectively. Assuming (as CI did) that ties are broken uniformly randomly, the probability with which l_1 is predicted to be chosen by each type in each game is as shown in Table I. The odd-one-out distribution of $L0$ choices implies that, in all cases, all non-oddities are chosen with equal probability.

$L0$ types favor labels with positive connotations is orthogonal to propositions about the payoff structure of the game. For a contrasting analysis of payoff-relevant connotations, see CI's Web Appendix.

[8]For hiders and seekers, the predicted effects of a uniquely disfavored label l_m are less sharp. Overall, hiders and seekers will favor both l_1 and l_m (because l_1 is chosen by $L1$ seekers and $L2$ hiders, while l_m is chosen by $L1$ hiders and $L4$ seekers).

[9]This assumption is not strictly necessary. In all the games we analyze, $L5$ types would behave just like $L1$ types, $L6$ just like $L2$, and so on.

TABLE I

IMPLICATIONS OF LEVEL-k THEORY FOR MATCHED GAMES WITH ODD-ONE-OUT FRAMES (WITHOUT ERRORS)

	Choice Probability for l_1:					
Role	$L0$	$L1$	$L2$	$L3$	$L4$	All Players
Coordinators	q	1	1	1	1	1
Discoordinators	q	0	1	0	1	$\pi_2 + \pi_4$
Hiders	q	0	0	1	1	$\pi_3 + \pi_4$
Seekers	q	1	0	0	1	$\pi_1 + \pi_4$

Since our objective is to test the joint hypothesis using observations of choice behavior, we are concerned with implications that do not depend on specific assumptions about the distribution of players across types. We focus on two such implications:

Implication 1: Averaging over equal numbers of discoordinators, hiders, and seekers, the oddity is chosen with a probability of at least $1/3$. This follows from the fact that the average of the entries in the "all players" column of Table I for discoordinators, hiders, and seekers is $1/3 + 2\pi_4/3$ (where $\pi_4 \geq 0$).

Implication 2: For each player role (*i.e.*, *coordinator, discoordinator, hider, seeker*), *the probability with which the oddity is chosen is independent of the number of strategies and the frame*. This follows from the fact that, for each role, the "all players" probability depends only on the distribution of types in the population.

These implications are surprising and, as far as we know, specific to level-k theory. We now consider their robustness to possible relaxations of the joint hypothesis, compatible with the general spirit of level-k theory. First, an error structure can be added to the level-k model by assuming that, for each type $L1$, $L2$, ..., the probability that it forms the correct belief about the modal $L0$ choice is $1 - e_F$, where $0 \leq e_F < (m-1)/m$;[10] otherwise, each of labels $l_2, \dots, l_m$ is equally likely to be believed to be modal. Table I then describes the special case in which $e_F = 0$. The general case can be described by replacing each "1" entry in Table I by $1 - e_F$, and each "0" entry by $e_F/(m-1)$.[11] Thus, for any given frame F, players' propensity to error is revealed in the frequency of non-oddity choices in the relevant Coordination game. If error propensities are known, Implications 1 and 2 can be revised to take account of error, as we explain in Section 4.

[10]This condition is necessary to ensure that l_1 is more likely than any other label to be believed to be modal.

[11]This error structure is equivalent to that assumed by CI, except that our error parameter is allowed to depend on the frame. CI's error parameter ε is equivalent to $e_F m/(m-1)$ in our notation.

Second, although level-k theorists commonly assume that there are no players at $L0$ (e.g., Crawford, Costa-Gomes, and Iriberri (2013)), this practice is not universal. Relaxing this assumption would have similar effects to those of introducing an error structure, except that the impact on Implications 1 and 2 would be smaller.[12]

Third, since H3 is needed only for the derivation of Implication 1, relaxations of that hypothesis would not affect Implication 2. Further, for each of the roles of discoordinator, hider, and seeker *separately* (and therefore independently of H3), the choice probability of l_1 is 1 at two of the four levels $L1$, $L2$, $L3$, and $L4$. Thus, Implication 1 is likely to be robust to plausible relaxations of H3. More specifically, it has been suggested that the decision problem faced by seekers is cognitively less demanding than that faced by hiders (Rubinstein, Tversky, and Heller (1996)). One might therefore want to assume, contrary to H3, that higher-level types are more frequent among seekers (and coordinators) than among hiders (and discoordinators). But, since the seeker types that choose l_1 are $L1$ (the lowest in the model) and $L4$ (the highest), Implication 1 is unlikely to be invalidated by plausible upward shifts in the distribution of seeker types.

3. EXPERIMENTAL DESIGN

Our experiment used Coordination, Discoordination, and Hide and Seek games with common frames. It took place in two series of sessions, using different subjects. As the two sets of subjects were recruited in the same way from the same pool of potential participants, it is reasonable to treat them as independent samples drawn from a single population.

The *HS sessions* involved 200 subjects, randomly matched into pairs, which were maintained throughout. In the first part of each HS session, each pair played a series of eighteen Hide and Seek games using different frames, with no feedback until the end of the experiment. One (randomly assigned) member of the pair played all eighteen games as hider, the other as seeker. In the second part of the session, the players' roles were reversed; the same eighteen games were played, in the same order as before, and again without feedback. After both parts had been completed, one game was selected at random from each part. For each of these games, the winner was paid £10 and the loser was paid nothing.

The *CD sessions* involved 80 subjects, randomly matched into pairs, again maintained throughout. Each pair played a series of eighteen Coordination games, followed by or preceded by a series of eighteen Discoordination games using different frames, the order of the two series being counterbalanced.

[12]Our error model is equivalent to assuming an additional type L^*, with probability $\pi^* = e_F m/(m-1)$, that chooses among strategies at random, and scaling down each of $\pi_1, \dots, \pi_4$ by the factor $1 - \pi^*$. The impact on Implications 1 and 2 would be less if L^* were assumed to favor l_1, as $L0$ does.

There was no feedback until both parts had been completed. Then, for each pair, one game was selected at random from each part. For each of these games, both players were paid £5 if they had achieved the objective of the game, and nothing otherwise.

Each game was presented to players as a row of either four or eight *boxes*, each of which enclosed a word, symbol, or picture. We will refer to these sets of boxes (without reference to position) as "frames." Each player knew that her co-player was seeing the same boxes in the same positions from left to right, and was asked to choose one box, either (if a coordinator or seeker) with the aim of choosing the same box as that chosen by her co-player, or (if a discoordinator or hider) with the aim of choosing a different box. The positions of the boxes were independently randomized for each pair and for each game.

We used 36 frames, eighteen with four boxes and eighteen with eight boxes. Each row in Figure 1 shows one of the eighteen frames (1b, ..., 18b) used in eight-box games. For purposes of reference, each box is numbered from left to right in the figure.

Our intention was that each frame should create an obvious oddity, and that all the other labels should be equally undistinguished. The oddity is shown in box 1 in the frames of Figure 1. Following Rubinstein, Tversky, and Heller (1996), we used two different forms of oddity. In each of frames 1b, 4b, 5b, 6b, 8b, and 15b, seven of the boxes are identical; the box with the distinct content is

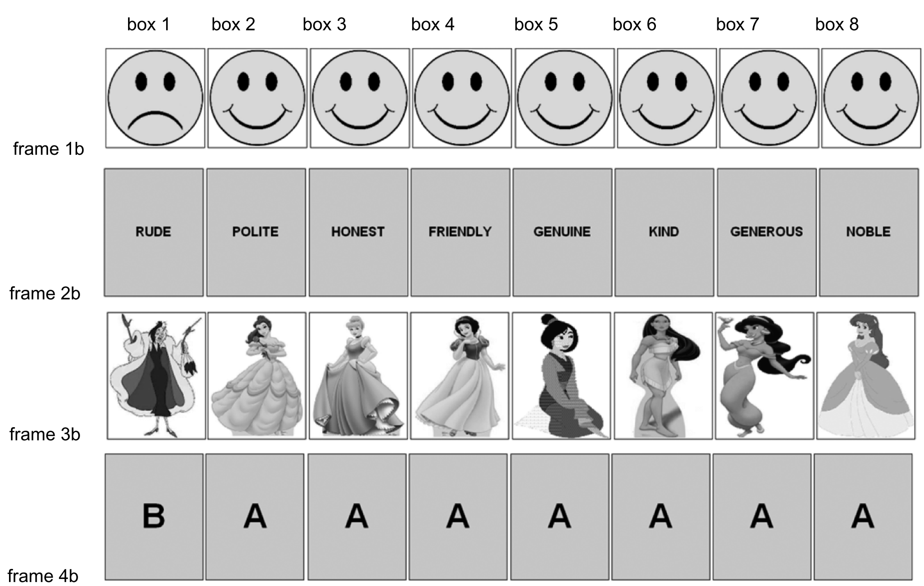

FIGURE 1.—Frames.

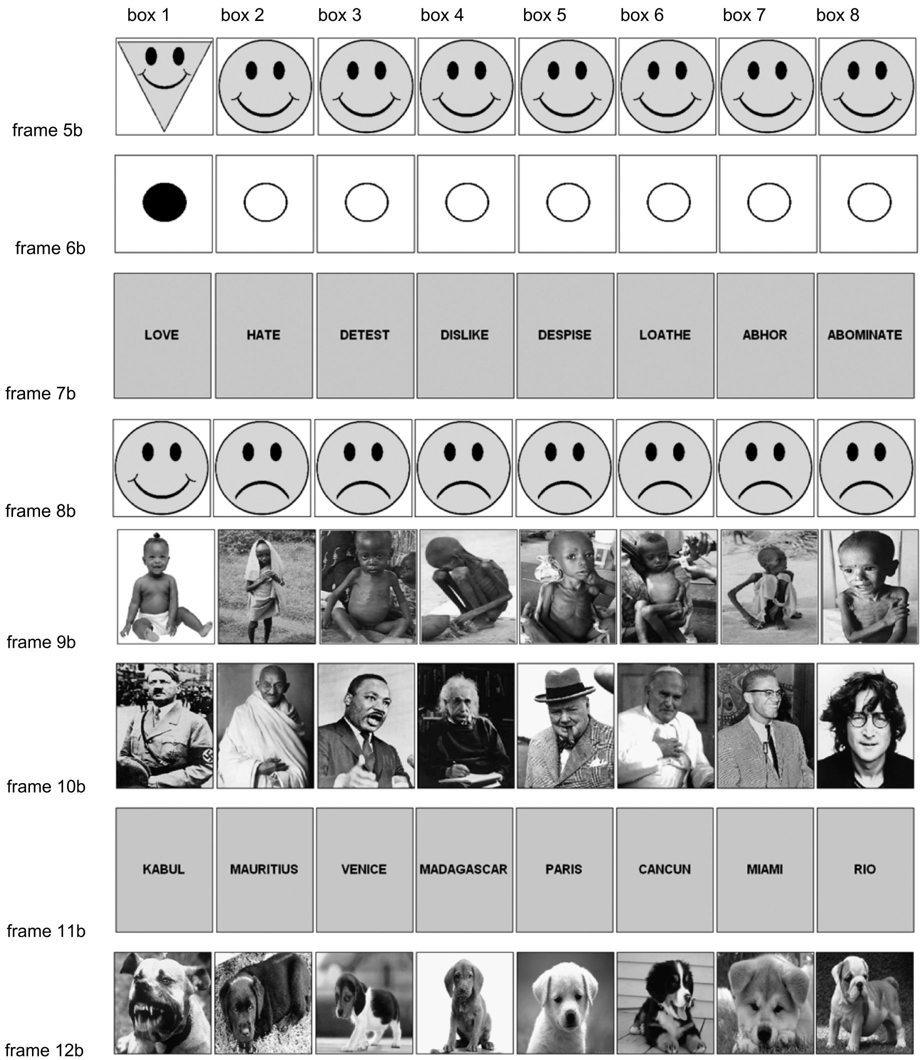

FIGURE 1.—*Continued*.

the *objective oddity*. In each of the other frames, every box is distinct, but box 1 is intended to be perceived as the *subjective oddity*. Again following Rubinstein et al., the oddity can have *positive*, *negative*, or *neutral* connotations relative to the other boxes in its frame. (The oddity is positive in frames 7b, 8b, 9b, 16b, 17b, and 18b, negative in frames 1b, 2b, 3b, 10b, 11b, and 12b, and neutral in frames 4b, 5b, 6b, 13b, 14b, and 15b.) The main function of this variety of forms of oddity was to maintain subjects' interest and attention. For the purposes of our tests, it is irrelevant what form oddity takes.

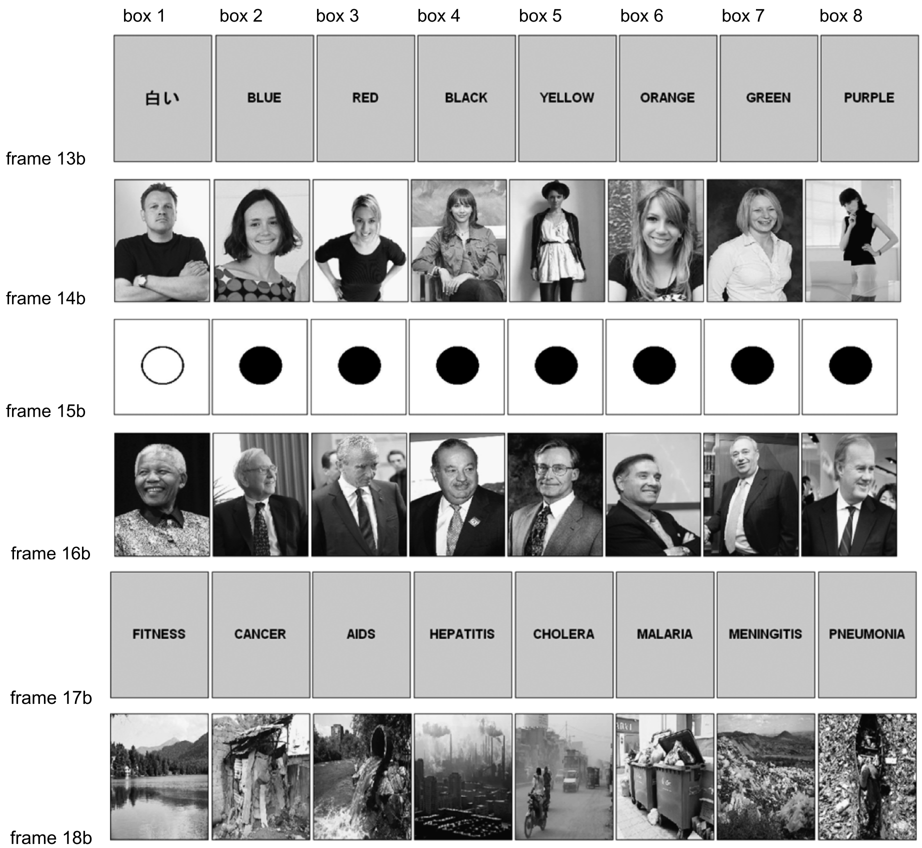

FIGURE 1.—*Continued*.

To generate as much similarity as possible between games with different values of m, the frames 1a, ..., 18a for the four-box games were formed by removing boxes 5, 6, 7, and 8 from each of the frames 1b, ..., 18b. Frames 1a, 2a, 4a, 5a, 7a, and 9a are virtually identical with those used in the six games investigated by Rubinstein et al.

The methods by which games were assigned to pairs of subjects, and by which the order of games played by pairs was randomized, are described in the Supplemental Material (Hargreaves Heap, Rojo Arjona, and Sugden (2014)). Part 1 of the HS sessions provided 50 observations of hiders and 50 observations of seekers in each of the 36 frames. We use these data for our main tests of level-k theory, but for completeness we also report the part 2 data.[13] The CD

[13] Part 2 of the HS treatment was an add-on investigation of the conjecture that, if a subject played a series of Hide and Seek games in one role and then exactly the same games in the other role, her reasoning would be more sophisticated in the second series.

treatment was designed with the intention of pooling data from parts 1 and 2, to provide 40 observations of coordinators and 40 observations of discoordinators in each of the 36 frames.

The experiment was conducted at the CBESS Laboratory at the University of East Anglia. It was implemented using z-Tree (Fischbacher (2007)). Subjects were recruited from the general student population and participated anonymously at computer workstations. Instructions (reproduced in the Supplemental Material) were presented on subjects' screens and were also read aloud by an experimenter to ensure that they were common knowledge. Sessions lasted for approximately 50 minutes. In the CD sessions, average earnings were £7.13; in the HS sessions, necessarily, they were exactly £10.

4. RESULTS

4.1. *Potential Confounds: Order and Position Effects*

The order in which blocks of games were played, and the order in which Coordination and Discoordination games were played in the CD sessions, had no systematic effect on players' behavior. As we expected, there were systematic differences between the two parts of the HS sessions.[14] We therefore pool CD data across the two parts of the experiment, but analyze the two parts of the HS data separately.

Because the positions of the boxes were randomized, we test for systematic position effects by comparing the frequencies with which subjects' chosen boxes were in each of the positions $1, \ldots, 4$ (in four-box games) or $1, \ldots, 8$ (in eight-box games). The relevant data are summarized in Table II. It is apparent that, in each class of games and for each role, each position was chosen with approximately the same frequency. The statistical tests reported in the table indicate that there may be some position effects in our data, but it seems clear that position was not a major determinant of players' choices.[15] From now on, therefore, we abstract from position effects and consider only the content of the boxes.

4.2. *Odd-One-Out Properties*

Since Implications 1 and 2 are conditional on $L0$ choices having an odd-one-out distribution, we check that the Validation Conditions were satisfied.

[14]Between part 1 and part 2, the frequency of oddity choices in both four- and eight-box games increased for hiders and decreased for seekers. Except in the case of eight-box hiders, these differences were statistically significant. Since part 1 seekers tended to favor oddities while part 1 hiders tended to avoid them, these results suggest some tendency for seekers who had previously played as hiders to play best responses to their own previous hiding behavior, and vice versa.

[15]Further analysis of position effects is reported in the Supplemental Material.

TABLE II

FREQUENCY OF CHOICES BY POSITION[a]

	Percentage of Choices That Are of Position:							
Role	1	2	3	4	5	6	7	8
Four-box games								
Coordinators ($n = 40$)	**27.5****	24.4	**26.1**	21.9				
Discoordinators ($n = 40$)	**26.4****	20.3**	**26.8**	**26.5**				
Hiders (part 1: $n = 100$)	25.0	22.2	24.9	**27.9**				
Hiders (part 2: $n = 100$)	24.3**	24.6*	24.9	**26.2**				
Seekers (part 1: $n = 100$)	22.9	**27.7***	**28.3**	21.1**				
Seekers (part 2: $n = 100$)	22.2**	**25.9**	**31.2*****	20.7***				
Eight-box games								
Coordinators ($n = 40$)	**12.8***	**14.7**	10.0	12.1***	**16.9*****	**11.5**	12.2	9.7**
Discoordinators ($n = 40$)	10.1	**13.5**	**16.0****	10.0	**12.6**	**12.6**	**14.9**	10.3
Hiders (part 1: $n = 100$)	12.1	**13.0**	**13.0**	10.2	9.8*	**13.3**	**15.6**	**13.0**
Hiders (part 2: $n = 100$)	**13.7**	**14.2**	11.7***	11.1**	10.8*	**14.4**	**12.9**	11.2
Seekers (part 1: $n = 100$)	10.0*	**13.4**	**13.8**	**13.2**	**13.6**	**12.8**	12.3	10.9
Seekers (part 2: $n = 100$)	10.2***	10.7	**15.8*****	**14.0**	**15.4**	**14.4**	11.2	8.2***

[a] n denotes the number of subjects who faced games of the relevant type. Each coordinator and discoordinator faced 18 games of that type; each hider and seeker faced 9 games of that type. Relative frequencies greater than the random-choice benchmark (25 percent for four-box games, 12.5 percent for eight-box games) are shown in bold. For each type of game and each position j, we find the number of choices made in position j by each subject i, and then run a chi-squared test of whether the distribution of these n numbers is different from the binomial distribution implied by random choice. Significant differences are shown by asterisks (*, **, and *** denoting significance at the 10, 5, and 1 percent levels).

Table III reports the frequency with which, in each frame, the oddity was chosen by coordinators (and, for completeness, by players in the other roles). In every frame but one (frame 11a), the oddity was the modal choice of coordinators, satisfying Validation Condition 1. In 33 of the 36 frames, the oddity was chosen by a majority (often a very large majority) of players, even though choices were distributed over four or eight labels. Although there was some tendency for oddities to be chosen less frequently if their connotations were negative than if they were neutral or positive, it is clear that coordinators' choices were very strongly skewed toward oddities, irrespective of connotations. Given this regularity, the most natural explanation of the anomalous behavior in frame 11a is that many subjects failed to perceive l_1 (Kabul) as the oddity.[16] We therefore drop this frame from our analysis.

Our test of Validation Condition 2 found no systematic non-randomness in discoordinators' choices over non-oddities.[17] We conclude that tests of Implications 1 and 2 are valid tests of the joint hypothesis.

[16]The three labels most frequently chosen by coordinators in this frame were Venice (17 choices), Kabul (10 choices), and Madagascar (also 10 choices).

[17]Since we have abstracted from position effects, this test is relevant only for the 24 games with subjective oddities. For each of these games, we compare the observed distribution of discoordi-

TABLE III

FREQUENCY OF ODDITY CHOICES BY FRAME[a]

		Percentage of Oddity Choices											
		Four-Box Games						Eight-Box Games					
Frame	Connotation	C	D	H1	H2	S1	S2	C	D	H1	H2	S1	S2
1a, b	negative	73	28	14	20	20	10	85	23	10	12	24	8
2a, b	negative	70	28	14	12	24	6	58	13	10	10	24	12
3a, b	negative	68	15	6	10	22	6	60	3	12	6	22	6
4a, b	neutral	80	30	18	20	16	18	83	25	16	8	28	20
5a, b	neutral	80	23	14	24	28	22	95	25	8	10	38	24
6a, b	neutral	93	25	6	22	22	20	93	18	12	8	28	14
7a, b	positive	85	40	14	22	24	38	88	10	8	8	36	24
8a, b	positive	93	28	12	12	20	22	88	20	12	8	28	32
9a, b	positive	88	13	8	18	32	22	93	13	8	6	34	18
10a, b	negative	53	30	20	18	24	16	40	20	6	14	16	12
11a, b	negative	25	33	22	30	24	22	28	13	12	16	16	8
12a, b	negative	68	25	12	18	26	20	70	10	0	6	24	14
13a, b	neutral	80	20	22	18	24	12	80	23	4	2	10	10
14a, b	neutral	68	13	16	12	16	24	63	13	2	2	6	6
15a, b	neutral	85	15	14	18	30	22	93	10	6	14	24	18
16a, b	positive	83	20	10	16	48	36	90	13	10	20	36	26
17a, b	positive	90	18	22	16	30	18	80	10	6	10	34	16
18a, b	positive	90	20	18	22	34	40	85	20	12	16	28	22
All frames		76	23	15	18	26	21	76	15	9	10	25	16

[a] C, D, H1, H2, S1, S2 denote coordinators ($n = 40$), discoordinators ($n = 40$), part 1 hiders ($n = 100$), part 2 hiders ($n = 100$), part 1 seekers ($n = 100$), and part 2 seekers ($n = 100$), respectively.

4.3. *Frequency of Oddity Choices*

The observed relative frequencies of oddity choices in the four roles are reported (as percentages) in the second column of Table IV. The first and third columns report the lower and upper bounds of the 95 percent confidence intervals for these relative frequencies, calculated by treating subjects as the units of observation.

Implication 1 is that, in the absence of error, the expected value of the average of the relative frequencies of oddity choices for discoordinators, hiders, and seekers (the *synthetic average*) is at least 0.333. As explained in Section 3, we use the part 1 HS data in testing Implications 1 and 2 (our main conclusions would be unchanged if we used part 2 data instead). The observed synthetic av-

nators' choices over non-oddities with the distribution implied by uniformly random choice over non-oddities (given actual choices of oddities). The null hypothesis of random choice is rejected at the 5 percent level in only one of the 21 cases in which there are enough non-oddity choices to allow a reliable chi-squared test. This 1/21 rejection rate is almost exactly what one would expect from randomly generated data.

TABLE IV

FREQUENCY OF ODDITY CHOICES BY PLAYERS IN ALL ROLES[a]

Role	Percentage of All Choices That Are of Oddity		
	Lower Bound	Observed	Upper Bound
Four-box games			
Coordinators ($n = 40$)	71.4	79.0	86.6
Discoordinators ($n = 40$)	17.1	22.8	28.5
Hiders (part 1: $n = 100$)	12.1	15.0	17.9
Hiders (part 2: $n = 100$)	14.8	18.6	22.5
Seekers (part 1: $n = 100$)	23.5	27.5	31.5
Seekers (part 2: $n = 100$)	18.1	22.0	25.9
Eight-box games			
Coordinators ($n = 40$)	69.6	76.0	82.4
Discoordinators ($n = 40$)	9.3	15.4	21.5
Hiders (part 1: $n = 100$)	5.4	8.6	11.7
Hiders (part 2: $n = 100$)	6.8	9.8	12.8
Seekers (part 1: $n = 100$)	21.2	25.3	29.5
Seekers (part 2: $n = 100$)	11.9	16.1	20.4

[a] n denotes the number of subjects who faced games of the relevant type. Each coordinator and discoordinator faced 18 games of that type (of which 17 or 18 appear in the analysis, depending on whether frame 11a was faced). Each hider and seeker faced 9 games of that type (of which 8 or 9 appear in the analysis).

erage is only 0.218 for four-box games and 0.164 for eight-box games. Even if, for each of the three roles, we were to use the upper bound of the relevant confidence interval rather than the observed value, the synthetic average would be only 0.260 for four-box games and 0.209 for eight-box games.

If errors are taken into account, Implication 1 generalizes to the prediction that the expected value of the synthetic average is at least $(1/3)(1 - e_F) + (2/3)(e_F/[m - 1])$. The average frequencies of oddity choices by coordinators imply values of e_F of 0.21 and 0.24 in four- and eight-box games, respectively.[18] Using these values, the synthetic average is predicted to be at least 0.310 for four-box games and at least 0.276 for eight-box games—still well outside the confidence intervals of our observations. We conclude that those observations are not consistent with Implication 1.

Implication 2 is that, in the absence of error and for any given player role, the expected proportion of oddity choices is the same for four- and eight-box games. The observed proportions of oddity choices at the two values of m are not significantly different for coordinators and seekers, but for discoordinators and hiders, the four-box proportions are greater than the corresponding eight-

[18] We conjecture that the similarity between these values is the result of two opposing effects offsetting one another. The more boxes there are to choose from, the more candidates there are for the status of "most salient box"; but the unique property of an oddity is more obvious, the more items lacking that property are presented alongside it.

box proportions by margins of 0.074 and 0.064, respectively. These differences are significant at the 5 percent level and are too large to be explained by error.[19]

As a robustness check, we ran the same tests of Implications 1 and 2 using only the data from games with neutral frames. The results were very similar: Implication 1 was still rejected, as was Implication 2 for hiders (see Supplemental Material for details).

5. RECONCILING RESULTS OF EX ANTE HYPOTHESIS-TESTING AND EX POST MODEL-FITTING

Our results give little support to the joint hypothesis, and hence to the idea that $L0$ specifications are portable across strategy-isomorphic games with common frames. However, CI showed that a parameterized level-k model has a good fit on data from six of RTH's Hide and Seek games. They argued that this model explains a "robust" and role-asymmetric *fatal attraction pattern* in those data: the strategy with the "least salient" label "was the strongly modal choice for both hiders and seekers, and was even more prevalent for seekers than hiders" (pp. 1732–1733). Since CI's model satisfies H2 and H3 with respect to the roles of hiders and seekers, their results provide some evidence of the portability of $L0$ assumptions. In this section, we try to resolve the apparent conflict between CI's conclusions and ours.

The most important difference between the two analyses is that CI's is restricted to Hide and Seek games, even though RTH's experiments included Coordination and Discoordination games with the same "AABA" and "ABAA" frames as Hide and Seek games in CI's data set. In these cases, as in our experiment, RTH found that the oddity was the strongly modal choice of coordinators. The implication (if level-k theory is correct) is that the oddity was the modal $L0$ choice *of coordinators*. In our analysis, but not in CI's, this implication constrains the $L0$ specification for hiders and seekers.

CI instead assumed that a label is more likely to be chosen at $L0$, the more "salient" it is, with salience defined as follows. In four of the six games they analyzed,[20] each player chose one "item" from a row of four letters—three As and one B, with the B in second or third place. CI assumed that "central A" is the least salient item, on the grounds that B is salient as the oddity and that end

[19]Using the inferred error parameters, the level-k model implies that discoordinators choose the oddity with probability $0.790(\pi_2 + \pi_4) + 0.210(\pi_1 + \pi_3)/3$ in four-box games and $0.760(\pi_2 + \pi_4) + 0.240(\pi_1 + \pi_3)/7$ in eight-box games. The former exceeds the latter by $0.030 + 0.006(\pi_1 + \pi_3)$. The corresponding excess for hiders is $0.030 + 0.006(\pi_1 + \pi_2)$. Both values are in the interval [0.030, 0.036] for all distributions of player types.

[20]CI's analysis excluded three games from Rubinstein, Tversky, and Heller (1996) in which the frames are essentially the same as our 2a, 7a, and 8a, and the oddity is in second or third position. These games are similar to ABAA and AABA games, but are excluded because the oddity has positive or negative connotations (see CI's Web Appendix). Only one of these games shows the fatal attraction pattern.

locations "may be inherently salient,"[21] and assumed that the two "end As" are equally salient. They allowed the data to decide whether B is the most salient item (*Specification 1*) or whether the end As are jointly most salient (*Specification 2*). In the remaining two games, the items are the numbers 1, 2, 3, 4. Without further explanation, CI assumed that 3 is least salient, that 1 and 4 are equally salient, and that 2 is most salient if and only if Specification 1 applies to ABAA and AABA games. *Given these assumptions about salience*, choices in all six games show the fatal attraction pattern. But conversely, the claim that this pattern is a robust feature of the data set depends on this particular combination of assumptions.[22]

CI fitted two alternative models to the pooled data from the six games. In both models, the distribution of player types and the $L0$ specification are the same for hiders and seekers, and types other than $L1$, $L2$, $L3$, and $L4$ have zero relative frequency. The models differ according to whether Specification 1 or 2 determines $L0$ behavior. The best-fitting of this class of models has the somewhat counterintuitive Specification 2. CI presented this as a "convincing" and "plausible" explanation of RTH's results (p. 1748) and, in particular, of the fatal attraction pattern. The implicit claim is that the ability of a model with plausible level-k assumptions to replicate an apparently surprising pattern in those data is evidence of the explanatory power of level-k theory.[23]

Since the $L0$ specification of this model does *not* explain behavior in ABAA and AABA Coordination games, CI's claim about the explanatory power of level-k theory can be accepted only if one does not require (as our hypothesis H2 does) that $L0$ assumptions are portable from Hide and Seek to Coordination. Conversely, if one views H2 as a natural extension of level-k theory, it may seem surprising that a model that is inconsistent with that hypothesis can fit experimental data. But just how surprising *is* this?

[21]The latter claim is supported by an unexplained citation of an experiment by Christenfeld (1995), which, in fact, found that when individuals pick from a row of identical items, they tend to *avoid* the end locations (CI, p. 1732).

[22]One might also ask whether our experiment found the fatal attraction pattern in those games that were most like CI's AABA and ABAA games, that is, four-box games in which an objective oddity with neutral connotations appeared in the second or third position. In these games, seekers tended to favor the central non-oddity (which CI interpreted as "least salient"), but hiders did not. Summing over part 1 and part 2 HS sessions, there are 185 (non-independent) observations of such games. The distributions of choices over "end non-oddity," "oddity," and "central non-oddity" were $(107, 30, 48)$ for hiders and $(73, 51, 61)$ for seekers, compared with the random-choice expectation $(92.5, 46.25, 46.25)$.

[23]Responding to possible doubts about the explanatory power of their estimated model, CI tested its portability by applying it to experimental data from two other games. However, the frames of those games are too different from RTH's Hide and Seek games to allow portability *of the L0 specification*. In effect, CI used a new $L0$ specification for each of the games. CI also reported that if their model was estimated using data from any one of the six RTH games in their data set, it had a good fit with the rest of the data. But this merely confirms the similarity of the data from the six games, when classified according to CI's assumptions about salience.

Let us assume, for the sake of the argument, that the only available data are those from the six Hide and Seek games analyzed by CI, and that Specifications 1 and 2 are the only plausible representations of $L0$ behavior. Thus, we have a data set with the fatal attraction pattern. A level-k model can be configured to predict that pattern. But for how many *other* possible patterns can the same claim be made?

Recall that in both specifications, the end As (or 1 and 4) are chosen with equal frequency at $L0$.[24] This forces a level-k model to predict that the end As are chosen with equal frequency by *all* types. Thus, CI's model has to explain the distributions of hiders' and seekers' choices between only three categories: B, central A, and *the combination of* the two end As. There are only four independent observations to be explained by a model with two alternative specifications, each with three free parameters (describing the population distribution of types).

The fatal attraction pattern is described by the triple (central A, central A, seeker) whose elements are respectively the modal choice of hiders, the modal choice of seekers, and which of the two roles has the higher mode. Ignoring ties, there are eighteen such patterns. (Since the model treats "end A" as a single category, we define its choice frequency as the average of the choice frequencies of the two end A items.) Following CI's arguments (p. 1743), we define a level-k model to be *plausible* if it has Specification 1 or 2 and has the properties $\pi_2 > \pi_1$ and $\pi_3 > \pi_4$ (i.e., the distribution of levels is "hump-shaped"). It can be shown that, of the eighteen possible patterns, seven (i.e., 39 percent of the total) are predicted by plausible level-k models.[25] The implication of this result, we suggest, is that even though CI's modeling framework imposes portability of the $L0$ specification between hiders and seekers, it allows too much freedom for the ex post rationalization of observed behavior. Thus, CI's claim that level-k theory *explains* the fatal attraction pattern (rather than that the theory *can be configured to fit* it) is open to question.[26]

6. CONCLUSION

We have studied a class of frames in which one label is the obvious odd-one-out. When those frames are used in Coordination games, the oddity is

[24]CI justified this assumption on the model-fitting grounds that "the end A frequencies are almost equal in the data" (p. 1738). Thus, the model does not *explain* that observation.

[25]The proof of this claim is given in the Supplemental Material.

[26]This critique applies with even greater force to Penczynski's (2013) analysis of behavior in Hide and Seek games with ABAA frames. Penczynski fitted a level-k model in which both the distribution of player types and the $L0$ specification were allowed to differ between hiders and seekers. This gave many more free parameters than in CI's analysis, and so even more scope for ex post rationalization. Since the fatal attraction pattern is an asymmetry between roles, it is hardly surprising that the best-fitting model has role asymmetries. (In this model, seekers reason at higher levels than hiders. In one of six cases studied, $L0$ behavior is significantly different for the two roles.)

overwhelmingly the modal choice. Thus, if level-k theory is correct, the oddity must be the modal choice of $L0$ types in those games. If, for each given frame, the specification of $L0$ is assumed to be the same for coordinators, discoordinators, hiders, and seekers, level-k theory implies that when frames with oddities are used in Discoordination and Hide and Seek games, the choice frequency for the oddity will be disproportionately high. But this hypothesis is disconfirmed by our data.

This result poses a severe challenge for the project of completing level-k theory by developing *general* hypotheses about $L0$ behavior. To be consistent with our data, $L0$ behavior must vary according to whether a given frame is confronted by a coordinator, discoordinator, hider, or seeker. But, since each of the three games is strategy-isomorphic when roles are considered separately, it is difficult to see how a general theory of $L0$ behavior can discriminate between those roles without using strategic properties of the games as explanatory variables; and to do that would be to compromise the aspiration of level-k theorists to anchor strategic reasoning on assumptions about non-strategic behavior.

Of course, it may still be possible to find a general definition of “naïve” $L0$ play that allows for some strategic awareness without simply reporting theorists’ intuitions about naïveté in particular games, and that generates accurate predictions. Or perhaps a more fundamental theory of analogical learning might allow such conceptions of naïveté to vary between players according to their private experience (compare Jehiel (2005)). But completing level-k theory in these ways will require a lot more work, and there is no guarantee of success.

REFERENCES

AGRANOV, M., A. CAPLIN, AND C. TERGIMAN (2012): “Naïve Play and the Process of Choice in Guessing Games,” Unpublished Paper, California Institute of Technology. Available at http://hss.caltech.edu/~magranov/documents/scp2013.pdf. [1134]

BARDSLEY, N., J. MEHTA, C. STARMER, AND R. SUGDEN (2010): “Explaining Focal Points: Cognitive Hierarchy Theory *versus* Team Reasoning,” *Economic Journal*, 120, 40–79. [1136]

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. [1133]

CHRISTENFELD, N. (1995): “Choices From Identical Options,” *Psychological Science*, 6 (1), 50–55. [1148]

COSTA-GOMES, M. A., V. P. CRAWFORD, AND B. BROSETA (2001): “Cognition and Behavior in Normal-Form Games: An Experimental Study,” *Econometrica*, 69 (5), 1193–1235. [1133,1134]

CRAWFORD, V. P., AND N. IRIBERRI (2007): “Fatal Attraction: Salience, Naïveté, and Sophistication in Experimental Hide-and-Seek Games,” *American Economic Review*, 97 (5), 1731–1750. [1134,1135]

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. [1133,1134,1139]

CRAWFORD, V. P., U. GNEEZY, AND Y. ROTTENSTREICH (2008): “The Power of Focal Points Is Limited: Even Minute Payoff Asymmetry May Yield Large Coordination Failures,” *American Economic Review*, 98 (4), 1443–1458. [1134,1136]

FISCHBACHER, U. (2007): "Z-Tree: Zurich Toolbox for Ready-made Economic Experiments," *Experimental Economics*, 10 (2), 171–178. [1143]

HARGREAVES HEAP, S., D. ROJO ARJONA, AND R. SUGDEN (2014): "Supplement to 'How Portable Is Level-0 Behavior? A Test of Level-*k* Theory in Games With Non-Neutral Frames'," *Econometrica Supplemental Material*, 82, http://www.econometricsociety.org/ecta/supmat/11132_miscellaneous.pdf; http://www.econometricsociety.org/ecta/supmat/11132_Data_and_Programs.zip. [1142]

JEHIEL, P. (2005): "Analogy-Based Expectation Equilibrium," *Journal of Economic Theory*, 123, 81–104. [1150]

KNEELAND, T. (2012): "Coordination Under Limited Depth of Reasoning," Unpublished Paper, University of British Columbia. Available at http://terri.microeconomics.ca/research/cldr.pdf. [1134]

MEHTA, J., C. STARMER, AND R. SUGDEN (1994): "The Nature of Salience: An Experimental Investigation of Pure Coordination Games," *American Economic Review*, 84 (3), 658–673. [1136]

NAGEL, R. (1995): "Unraveling in Guessing Games: An Experimental Study," *American Economic Review*, 85 (5), 1313–1326. [1133]

PENCZYNSKI, S. P. (2013): "Strategic Thinking: The Influence of the Game," Unpublished Paper, University of Mannheim. Available at http://www.penczynski.de/attach/Focality.pdf. [1134,1149]

RUBINSTEIN, A. (1999): "Experience From a Course in Game Theory: Pre and post-Class Problem Sets as a Didactic Device," *Games and Economic Behavior*, 28 (1), 155–170. [1135]

RUBINSTEIN, A., AND A. TVERSKY (1993): "Naïve Strategies in Zero-Sum Games," Working Paper 17-93, Sackler Institute of Economic Studies, Tel Aviv University. [1135]

RUBINSTEIN, A., A. TVERSKY, AND D. HELLER (1996): "Naïve Strategies in Competitive Games," in *Understanding Strategic Interaction: Essays in Honor of Reinhard Selten*, ed. by W. Albers, W. Güth, P. Hammerstein, B. Moldovanu, and E. van Damme. Berlin: Springer-Verlag, 394–402. [1135,1139,1140,1147]

STAHL, D. O., AND P. WILSON (1994): "Experimental Evidence on Players' Models of Other Players," *Journal of Economic Behavior and Organization*, 25 (3), 309–327. [1133]

STRZALECKI, T. (2010): "Depth of Reasoning and Higher Order Beliefs," Unpublished Paper, Harvard University. Available at http://scholar.harvard.edu/tomasz/files/strzalecki-depth.pdf. [1134]

Dept. of Political Economy, King's College London, Strand, London WC2R 2LS, U.K.; s.hargreavesheap@kcl.ac.uk,

Economic Science Institute, Chapman University, One University Drive, Orange, CA 92866, U.S.A.; rojoarjo@chapman.edu,

and

School of Economics, University of East Anglia, Norwich NR4 7TJ, U.K.; r.sugden@uea.ac.uk.

Manuscript received October, 2012; final revision received December, 2013.