

Fatal Attraction:

Saliency, Naiveté, and Sophistication in Experimental “Hide-and-Seek” Games

Vincent P. Crawford and Nagore Iriberrí¹

20 August 2004; revised 25 February 2006; this revision 6 February 2007

“Any government wanting to kill an opponent... would not try it at a meeting with government officials.”

—comment on the poisoning of Ukrainian presidential candidate (now president) Viktor Yushchenko, quoted in C. J. Chivers (2004)

“...in Lake Wobegon, the correct answer is usually ‘c’.”

—Garrison Keillor (1997) on multiple-choice tests, quoted in Yigal Attali and Maya Bar-Hillel (2003)

Abstract: “Hide-and-seek” games are zero-sum two-person games in which one player wins by matching the other's decision and the other wins by mismatching. Although such games are often played on cultural or geographic “landscapes” that frame decisions non-neutrally, equilibrium ignores such framing. This paper reconsiders the results of experiments by Rubinstein, Tversky, and others whose designs model non-neutral landscapes, in which subjects deviated systematically from equilibrium in response to them. Comparing alternative explanations theoretically and econometrically suggests that the deviations are well explained by a structural non-equilibrium model of initial responses based on “level- k ” thinking, suitably adapted to non-neutral landscapes.

Keywords: behavioral game theory, experiments, hide-and-seek games, framing effects, saliency, bounded rationality, level- k thinking

JEL classification numbers: C70, C92

¹University of California, San Diego (Crawford); and Departament d'Economia i Empresa, Universitat Pompeu Fabra, Barcelona, Spain (Iriberrí). Email: vcrawfor@dss.ucsd.edu and nagore.iriberri@upf.edu. Most of Iriberrí's work on this paper was completed at the University of California, San Diego. We are grateful to the National Science Foundation (Crawford) and the Centro de Formación del Banco de España (Iriberrí) for research support; to Miguel Costa-Gomes, Victor Ferreira, Navin Kartik, Barry Nalebuff, Steven Scroggin, Ricardo Serrano-Padial, Joel Sobel, David Swinney, Mark Voorneveld, Mark Walker, Joel Watson, and three anonymous referees for helpful comments or discussions; to Dale Stahl for helpful discussions and for providing a copy of Michael Bacharach and Stahl (1997a); to Stahl and Daniel Zizzo for searching for Michael Bacharach and Stahl (1997b); to Barry O'Neill and Amnon Rapoport (with the help of Ido Erev and Mark Walker, respectively) for providing data from their experiments; and to Ariel Rubinstein for providing a copy of Rubinstein and Amos Tversky (1993), searching for additional data, and helpful discussions. Glenn Close and Michael Douglas (<http://www.imdb.com/title/tt0093010/>) were no help at all.

Game theorists have been intrigued by hide-and-seek games—zero-sum two-person games in which one player wins by matching the other’s decision and the other wins by mismatching—for more than 50 years (John von Neumann (1953)). These games cleanly model a strategic problem that is central to many economic, political, and social settings as well as the obvious military and security applications. Examples include entry games where entry requires a differentiated product and blocking it requires matching the entrant’s design; election campaigns in which a challenger can win only by campaigning in different areas than the incumbent; and fashion games in which hoi polloi wish to mimic the elite but the elite prefer to distinguish themselves.

Although zero-sum two-person games are one of game theory’s success stories, equilibrium analysis of hide-and-seek games is not very helpful as a guide to prediction or decision-making. There seem to be two main reasons for this, both illustrated by our epigraphs: Hide-and-seek games are often played without clear precedents, so equilibrium depends on strategic thinking rather than learning; but such thinking may not follow the fixed-point logic of equilibrium. A game theorist would reply to our first epigraph, “But if investigators thought that way, a meeting with government officials is precisely where a government *would* try to kill an opponent.” Further, hide-and-seek games are usually played on naturally occurring cultural or geographic “landscapes” that are non-neutral across locations in framing and/or payoffs. Equilibrium ignores such landscapes except as they affect payoffs, but non-equilibrium thinking may respond to them.

Both reasons are also well illustrated by the experimental results of Rubinstein and Tversky (1993; “RT”) and Rubinstein et al. (1996; “RTH”); see also Rubinstein (1999; “R”). RT, RTH, and R (collectively “RTH”) elicited subjects’ initial responses to hide-and-seek games. RTH explained the games as “stories,” probably increasing subjects’ comprehension. In a leading example, R told seekers: “You and another student are playing the following game: Your opponent has hidden a prize in one of four boxes arranged in a row. The boxes are marked as follows: A, B, A, A. Your goal is, of course, to find the prize. His goal is that you will not find it. You are allowed to open only one box. Which box are you going to open?” Hiders were told an analogous story. Thus the entire structure, including the order and labeling of locations, was publicly announced.²

This story makes the framing of locations non-neutral in two ways. The “B” location is distinguished by its label, and so is salient in one of Thomas Schelling’s (1960) senses. And the

²RT’s and RTH’s subjects’ payments appeared sufficient to motivate them, and the binary-lottery structure of the payoff function implies under standard assumptions that players maximize expected money payoffs, without regard to risk preferences. Note however that in R’s experiments, subjects were not paid or screened for exposure to game theory.

two “end A” locations, though not distinguished by their labels, may be inherently salient, as RT and RTH argue, citing Nicholas Christenfeld (1995). As RT note, these two saliencies interact to give the remaining location, “central A,” its own brand of uniqueness as “the least salient location.” This aspect of their designs is important as a tractable abstract model of a naturally occurring landscape.

Figure 1 translates RTH's story into a payoff matrix. Because the seeker chooses without observing the hider's choice, their choices are strategically simultaneous. If the seeker chooses the same location as the hider, he wins a payoff normalized here to one; if not, the hider wins that payoff. This game has a unique equilibrium in mixed strategies, in which both players randomize uniformly across locations, independent of their framing. Because the seeker can choose only one of the four locations, he is at a disadvantage in equilibrium, finding the prize only 25% of the time. Despite this clear equilibrium prediction, RTH's publicly announced order and labeling of locations create a potential for framing effects, and their subjects deviated systematically from equilibrium in ways that were highly sensitive to framing.³ Table 1 gives the aggregate choice frequencies for the “RTH-4” treatment described above and RTH's most closely related other treatments (Table A1 in the web appendix gives frequencies from less closely related treatments). In RTH-4, central A was the strongly modal choice for both hidere and seekers, and was even more prevalent for seekers than hidere. As a result, assuming independence, seekers could expect to find the prize 32% of the time, substantially more than the 25% equilibrium predicts. These qualitative patterns extend, properly interpreted, to the other five treatments in Table 1, which we shall argue are closely analogous to RTH-4. They also extend with minor exceptions to a very large sample from more recent internet experiments (Ariel Rubinstein, private communication).

Like the beliefs that underlie our epigraphs, these patterns are an intriguing puzzle: If hidere and seekers are equally intelligent, on average, why don't hidere tempted to hide in central A realize that seekers will also be tempted to look there? Why do hidere allow seekers to find them 32% of the time when they could hold it down to 25% via the equilibrium mixed strategy? The puzzle is

³The p -values from chi-square tests for significant differences from equilibrium in hidere' and seekers' choice frequencies in RTH's six treatments are shown in Table 1. Although in this game any strategy, pure or mixed, is a best response to equilibrium beliefs, *systematic* deviations of aggregate choice frequencies from equilibrium probabilities must (with high probability) have a cause that is partly common across players, and are therefore indicative of systematic deviations from equilibrium. Other studies of framing effects in different kinds of games include Jörn Scharlemann et al. (2001), who studied trust games in which otherwise anonymous players were “labeled” by photographs; and Judith Mehta et al. (1994) and Nicholas Bardsley et al. (2006), who studied coordination games in which decisions had naturally occurring labels, as in Schelling's (1960) classic experiments.

deepened by the patterns' robust asymmetry across player roles. Why do seekers choose central A more often than hiders? Although the payoff structure is asymmetric, this asymmetry is not explained by noisy generalizations of equilibrium such as Richard McKelvey and Thomas Palfrey's (1995) quantal response equilibrium ("QRE"), which coincides with equilibrium in RTH's games.

RTH took their subjects' deviations from equilibrium as *prima facie* evidence that the subjects did not think strategically; and with one exception, explained below, they did not consider alternative explanations of their results.⁴ But in our view, such robust patterns of behavior are unlikely to lack a coherent explanation; and given the simplicity of the strategic question hide-and-seek games pose, the explanation is unlikely to be nonstrategic. On the contrary, deviations from equilibrium in games where its rationale is especially strong seem to offer a particularly promising opportunity to explore alternative strategic theories of initial responses to games.

In this paper we propose an explanation of RTH's and related results using a non-equilibrium model of initial responses to games based on "level- k " thinking, building on Bacharach and Stahl's (1997a) analysis of a simplified version of RTH's games.⁵ Level- k models were introduced by Stahl and Paul Wilson (1994, 1995) and Rosemarie Nagel (1995) and further developed by Teck-Hua Ho et al. (1998), Costa-Gomes et al. (2001), Vincent Crawford (2003), Colin Camerer et al. (2004), Costa-Gomes and Crawford (2006), and Crawford and Nagore Iriberry (2005). They have strong experimental support, which should allay the concern that once one relaxes equilibrium, anything is possible. Like RTH, we focus on hide-and-seek and related games played on landscapes that are non-neutral in the framing of locations but neutral with regard to payoffs.⁶ Although learning may converge to equilibrium over time, we also follow RTH in studying subjects' initial responses, which reveal their strategic thinking most clearly. Our goals are to resolve a long-

⁴In RT's words, "The finding that both choosers and guessers selected the least salient alternative suggests little or no strategic thinking." In RTH's words, "...the players employed a naïve strategy (avoiding the endpoints), that is not guided by valid strategic reasoning. In particular, the hiders in this experiment either did not expect that the seekers too, will tend to avoid the endpoints, or else did not appreciate the strategic consequences of this expectation."

⁵Bacharach and Stahl's (1997a) analysis of hide-and-seek games did not appear in the published version of their paper, Bacharach and Stahl (2000). Bacharach and Stahl (1997b), whose title suggests a more detailed version of their hide-and-seek analysis, is no longer available. The relationship between their analysis and ours is explained below.

⁶Robert Rosenthal et al. (2003) and papers cited there report experiments on hide-and-seek games with non-neutral payoffs but neutral framing. Camerer et al. (2004, p. 883) show that level- k models can explain subjects' responses to such games. Crawford, Gneezy, and Rottenstreich (2007) propose a level- k explanation along the lines suggested here of coordination experiments in which subjects respond strongly to both labeling and salience created by payoff perturbations, which opposes label salience in one player role but reinforces it in the other. Von Neumann (1953) characterized equilibria in hide-and-seek games whose payoffs vary with location, including a two-dimensional game in which a hider hides in a matrix and a seeker guesses the hider's row or column.

standing behavioral puzzle; to explore the specification of level- k models for games played on non-neutral landscapes; and to establish a link between RTH's results and experimental work on strategic thinking, and so bring new evidence to bear on modeling initial responses to games.

Our level- k model allows behavior to be heterogeneous, but it assumes that each player follows a rule drawn from a common distribution over a particular hierarchy of decision rules or *types*. Type L_k for $k > 0$ anchors its beliefs in a nonstrategic L_0 type and adjusts them via thought-experiments with iterated best responses: L_1 best responds to L_0 , L_2 to L_1 , and so on. L_1 and higher types have accurate models of the game and they are rational in that they choose best responses to beliefs. Their only departure from equilibrium is replacing its assumed perfect model of others' decisions with simplified models that avoid the complexity of equilibrium analysis.

In applications the population type frequencies are usually estimated from the current dataset or translated from previous work. The estimated distribution tends to be stable across games and hump-shaped, with most weight on L_1 , L_2 , and L_3 . Thus the anchoring L_0 type exists mainly in the minds of higher types. Even so, its specification is the main issue that arises in defining a level- k model for games with non-neutrally framed locations, and the key to its explanatory power.

We show that a level- k model with an L_0 that is insensitive to payoffs and symmetric across player roles, but sensitive to framing in that L_0 is constrained to favor salient locations, as seems plausible for a payoff-insensitive anchoring type, allows a simple explanation of RTH's and related results. Estimating the model with population type frequencies and L_0 's choice probabilities constrained to be equal across roles yields a type distribution like those previously estimated for other games, though with somewhat more weight on higher types than usual, perhaps due to the transparency of the strategic question hide-and-seek games pose. The estimated mixture of types' best responses to such an L_0 explains subjects' robust, role-asymmetric deviations from equilibrium, with no difference in behavioral assumptions for hidiers and seekers.

Although our proposed level- k model yields a plausible resolution of the puzzle posed by RTH's results, its large number of parameters and the freedom to specify L_0 raise concerns about overfitting within RTH's sample and portability, the extent to which estimating the model for RTH's games helps to predict behavior in other games. We address these concerns by comparing it with four alternatives: two level- k models that relax the constraints that L_0 favors salience and/or is role-symmetric; and two equilibrium models as close as possible to the mainstream, with "hard-wired" payoff perturbations that can describe an instinctive attraction to salience for seekers and an

instinctive aversion for hidiers, one with their magnitudes (but not signs) restricted to be the same for hidiers and seekers and one with their magnitudes unrestricted. The web appendix extends these equilibrium comparisons to QRE with perturbations, which never performs better than equilibrium with perturbations, and sometimes performs worse. The alternative models can also explain RTH's results, usually with a small likelihood advantage over our proposed model. But they too have large numbers of parameters, and they raise similar concerns about overfitting and portability.

We test for overfitting within RTH's sample by using each model to compute estimates separately for each of their six treatments and using each estimated model to “predict” the results of the other five treatments. Overall, our proposed level- k model fits slightly worse than all four alternatives. But it has a lower mean squared prediction error (“MSE”) than each alternative but the level- k model that relaxes the constraint that $L0$ is role-symmetric, whose error is 10% lower.

We evaluate portability by using each model to “predict” subjects' initial responses in Barry O'Neill's (1987) and Amnon Rapoport and Richard Boebel's (1992) experiments, whose games are strategically close to RTH's games but with more complex win-loss patterns, different framing, and in one case five locations. The equilibrium with perturbations models are easily adapted to O'Neill's game using the same general notions of salience as for RTH's games; but in each case the perturbations estimated under the natural constraints regarding how players react to salience are 0, reducing the model to an equilibrium model, which predicts O'Neill's subjects' initial responses poorly. The level- k models adapt just as easily and predict well. For O'Neill's game our proposed level- k model with a role-symmetric $L0$ that favors salience has lower MSE than any of the alternative models. The level- k models with role-symmetric $L0$ adapt easily to Rapoport and Boebel's more complex game; but there is no plausible, parsimonious way to adapt a model with a role-asymmetric $L0$ or the equilibrium with perturbations models to that game. Our proposed model has total MSE slightly lower than equilibrium or a level- k model with a role-symmetric $L0$ that avoids salience. The analysis traces our proposed model's advantage in portability to its reliance on general principles of strategic behavior that are not overly sensitive to the details of the structure.

The rest of the paper is organized as follows. Section I explains the analogies among RTH's six treatments and the case for pooling their data the way we do. Sections II and III describe our equilibrium with perturbations and level- k models. Section IV reports econometric estimates, Section V discusses overfitting, and Section VI discusses portability. Section VII is the conclusion.

I. Analogies across RTH's Treatments

This section describes the five other RTH treatments in Table 1 and explains the senses in which they are analogous to the RTH-4 treatment. We then argue that, properly interpreted, the patterns of deviations from equilibrium in RTH-4— that central A was modal for both hidere and seekers, and even more prevalent for seekers—extend to the other five treatments. We also test whether the choice frequencies in RTH's six treatments can be pooled in the econometric analysis.

In RTH-4 the hider hid a desirable “prize,” which we call a “treasure” as in RTH (1996). The five other treatments include three more treasure treatments, RT-AABA-Treasure, RT-1234-Treasure, and R-ABAA. They also include two “mine” treatments, RT-AABA-Mine and RT-1234-Mine, identical to the corresponding treasure treatments except that the hidden object is undesirable, so that hidere and seekers' 0-1 payoffs are interchanged. This yields an equivalent normal form with players' roles reversed, leaving equilibrium predictions unchanged.

However, because hidere inherently move first, even though seekers do not observe their choices mine treatments have different extensive forms than treasure treatments with reversed roles. RTH, suspecting that this difference might make it easier for seekers to mentally simulate hidere choices, used mine treatments to test whether it explains why seekers in treasure treatments did better than equilibrium predicts. But the mine treatments yielded results very close to the treasure treatments with roles reversed, which suggests that the seekers' advantage was somehow driven by subjects' responses to the normal-form structure, as in all of the models considered here.

In the three ABAA or AABA Treasure treatments and the AABA Mine treatment, central A was modal for both hidere and seekers. This pattern extends to the 1234 Treasure and Mine treatments if we follow RT's suggestion that “the least salient response...may correspond to 3, or perhaps 2” and take 2 as analogous to B and 3 to central A. Given this correspondence, central A was more prevalent for seekers in all four treasure treatments and more prevalent for hidere in both mine treatments. Thus if hidere in treasure and seekers in mine treatments are treated as equivalent, this pattern is also the same in all six treatments. Further, the frequencies with which seekers found a treasure *or* a mine exceeded 25%, so that seekers (hidere) had higher (lower) expected payoffs than in equilibrium in treasure treatments, and vice versa in mine treatments.

From now on we build in these analogies by treating 2 as equivalent to B, and treating mine treatments as equivalent to treasure treatments with reversed player roles. To avoid unnecessary repetition, we use “central A” to refer to either a central A or a 3 location, and so on; and we refer

to mine treatments as if they were treasure treatments with reversed roles. No theory of which we are aware could predict these equivalences across different frames; their justification is that they allow us to use a more general theory of strategic behavior to interpret the patterns in the data.

After transforming the data accordingly, chi-square tests for differences in subjects' aggregate choice frequencies reveal no significant difference for seekers (p -value 0.4836) or hidiers (p -value 0.1635) across the six treatments. We therefore pool the data except as noted. The pooled sample includes 624 hidiers and 560 seekers, with the aggregate choice frequencies in Table 3.⁷

II. An Equilibrium with Payoff Perturbations Model of RTH's Games

To fix ideas, we begin with an explanation of RTH's results as close as possible to the mainstream, an equilibrium model with hard-wired payoff perturbations that can describe an instinctive attraction to salience for seekers and an instinctive aversion for hidiers. Because the end As' frequencies are almost equal in the data, we set their perturbations equal for simplicity. Thus, we assume that seekers gain an additional payoff of e for an end A location or f for the B location; and that hidiers lose the same payoffs for those choices, yielding the perturbed matrix in Figure 2. Given the signs of the perturbations, intuition suggests that $e, f > 0$; but we do not assume this here.

First restrict the magnitudes of e and f to be the same for hidiers and seekers as in Figure 2. If $-1 < (f - 2e), (2e - 3f), (2e + f) < 3$, the perturbed game has a unique, symmetric, totally mixed equilibrium, with hidiers and seekers both playing A, B, A, A with probabilities $1/4 - e/2 + f/4$, $1/4 + e/2 - 3f/4$, $1/4 + e/2 + f/4$, and $1/4 - e/2 + f/4$. Both hidiers and seekers play central A with probability $1/4 + e/2 + f/4$, which is greater than $1/4$ when $2e + f > 0$. A seeker finds the treasure with probability:

$$(1) \quad \left(\frac{1}{4} - \frac{e}{2} + \frac{f}{4}\right)^2 + \left(\frac{1}{4} + \frac{e}{2} - \frac{3f}{4}\right)^2 + \left(\frac{1}{4} + \frac{e}{2} + \frac{f}{4}\right)^2 + \left(\frac{1}{4} - \frac{e}{2} + \frac{f}{4}\right)^2 = \frac{1}{4} \left(1 + (2e - \sqrt{3}f)^2 + 4(\sqrt{3} - 1)ef\right),$$

which is greater than $1/4$ for $e, f > 0$. Thus, equilibrium with perturbations of equal magnitudes but opposite signs for hidiers and seekers can explain why central A is modal for both roles if $2e + f > 0$, but not its greater prevalence for seekers. It can only explain the role-asymmetry in RTH's results by invoking differences in the magnitudes of e and f , with $2e + f$ nearly twice as large for hidiers as for seekers (Table 3). There is no logical reason why the game's role-asymmetric payoffs should not evoke instinctive reactions that differ this way, but neither to our knowledge is there any

⁷We made minor adjustments to RTH's published data to reconcile reported frequencies and sample sizes (web appendix). Pairwise tests suggest that RTH-4 has somewhat higher frequencies of B (as well as central A) for both hidiers and seekers. Although this difference is intriguing, we focus on explaining the more robust patterns discussed in the text.

plausible theory to explain such differences. Without such a theory, perturbations of unrestricted magnitudes give the model enough flexibility to explain virtually any pattern of behavior, raising concerns about overfitting; and the lack of a theory also raises concerns about portability. (These issues are discussed in more detail in the web appendix.)

In RTH's unperturbed game, QRE coincides with equilibrium for any value of its precision parameter. But because QRE explains deviations from equilibrium in some other experiments, and it normally responds to payoff asymmetries, it is natural to hope that QRE with perturbations can explain RTH's results without unexplained differences in the magnitudes of the perturbations.⁸ We consider this possibility in the web appendix, focusing on the popular logit specification and assuming $e, f > 0$ in Figure 2, which seems compelling. With perturbations of equal magnitudes QRE, like equilibrium, can explain why central A is model for hiders and seekers, but it robustly predicts that central A is more prevalent for *hiders*, the opposite of what RTH found. Perturbations of unrestricted magnitudes yield an effectively infinite estimate of QRE's precision parameter, reducing it to equilibrium with unrestricted perturbations. Thus QRE is unhelpful in this setting.

III. A Level- k Model of RTH's Games

This section introduces our proposed level- k model for RTH's games and alternative level- k models. Each player follows one of five types, $L0$, $L1$, $L2$, $L3$, or $L4$, with given probabilities in each role r , s , t , u , and v . Type Lk for $k > 0$ anchors its beliefs in type $L0$ and adjusts them via thought-experiments with iterated best responses: $L1$ best responds to $L0$, $L2$ to $L1$, and so on.⁹ For $k > 0$ these responses determine types' normal choices, with ties broken uniformly randomly.

For econometric reasons (Section IV) we assume $r = 0$, so the anchoring type $L0$ exists only in the minds of higher types, as the starting point for their strategic thinking. Even so, $L0$ is the key to the model's explanatory power. We also assume that $L0$ is payoff-insensitive as in most previous level- k analyses; this choice is discussed below. In games with symmetric framing and sets of decisions like those we study, it is plausible to assume that a payoff-insensitive $L0$ is symmetric

⁸Rosenthal et al. (2003) find that QRE gives a reasonable explanation of the qualitative features of subjects' role-asymmetric deviations from equilibrium in 2x2 hide-and-seek games with neutral framing but varying payoffs.

⁹Costa-Gomes and Crawford (2006) summarize the experimental evidence for the level- k model, and give support for our assumptions that $L2$ best responds to an $L1$ without decision errors, and to $L1$ alone rather than a mixture of $L1$ and $L0$, etc., unlike in Stahl and Wilson (1994, 1995) or (in the latter case) Camerer et al. (2004). In RTH's games $L5$'s choice probabilities (but not its expected payoffs) are the same as $L1$'s, so $L5$ is equivalent to $L1$, $L6$ to $L2$, and so on.

across player roles.¹⁰ We further assume that $L0$ probabilistically favors salient locations: B and the end As in RTH's games. Because the end A frequencies are almost equal in the data, we set their choice probabilities for $L0$ (and so for higher types) equal for simplicity. Thus, in RTH's games $L0$ hidiers and seekers both choose A, B, A, A with probabilities $p/2, q, 1-p-q, p/2$, where $p > 1/2$ and $q > 1/4$.

When $r = 0$, the model's implications apart from errors are determined by $L1$ hidiers' and seekers' best responses to $L0$, which determine the normal choices of $L2, L3$, and $L4$ as well as $L1$; and by the type frequencies s, t, u , and v . Table 2 lists types' normal choice probabilities when $p > 1/2$ and $q > 1/4$, distinguishing the cases $p < 2q$ —in which $L0$ responds more strongly to the salience of a B than an end A location—and $p > 2q$. $L1$ hidiers choose central A to avoid $L0$ seekers, and $L1$ seekers avoid central A in their searches for $L0$ hidiers, choosing B if $p < 2q$ or the end As (with equal probabilities) if $p > 2q$. For similar reasons, $L2$ hidiers choose central A with probability $1/3$ or $1/2$ and $L2$ seekers choose it with probability 1 ; $L3$ hidiers avoid central A and $L3$ seekers choose it with probability $1/3$ or $1/2$; and $L4$ hidiers and seekers avoid central A.¹¹

These choice patterns allow the model to explain RTH's results with behaviorally plausible type frequencies. Table 2's bottom lines give the population choice probabilities as functions of the type frequencies and the error rate ϵ . Section IV's econometric analysis (Table 3) confirms that a role-symmetric $L0$ with $p > 2q$ and a population with 19% $L1$, 32% $L2$, 24% $L3$, and 25% $L4$ players in each role closely matches RTH's hidiers' and seekers' choice frequencies. This almost hump-shaped type distribution is like those estimated for other settings (Stahl and Wilson (1994, 1995); Camerer et al. (2004); Costa-Gomes and Crawford (2006)). It is, however, shifted upward somewhat, perhaps due to the transparency of the strategic question hide-and-seek games pose.¹²

Imagine, more concretely, that in the situation of our first epigraph there are four dinners at which the government could try to poison Yushchenko. Dinner "B" is with government officials; but none of the others is salient the way RTH's end locations are. As the only salient location, B is the modal choice for $L0$ poisoners and investigators. $L1$ poisoners avoid B, but $L1$ investigators

¹⁰ One might argue that the framing isn't completely role-symmetric because the roles have different labels, but RTH's failure to find significant differences between the behavior of hidiers in treasure treatments and seekers in mine treatments, and vice versa (Section I), suggests that role-asymmetric framing is not what is driving subjects' behavior.

¹¹ These choices remain the same whenever $p + 2q > 1$ and $3p + 2q > 2$; we maintain $p > 1/2$ and $q > 1/4$ for simplicity.

¹² This shift hints at the possibility of a more general theory that relates the type frequencies to the transparency and cognitive difficulty of the game. In contrast to RTH's interpretation of their results as evidence of strategic naïveté, using our level- k model to explain their results implies *lower* bounds on the population's sophistication in that, for central A to be more prevalent for seekers than hidiers, there must be sufficiently more $L2$ and $L3$ than $L1$ subjects.

choose it. $L2$ poisoners and investigators both avoid B. $L3$ poisoners choose B, but $L3$ investigators avoid it. $L4$ poisoners and investigators both choose B. Thus our first epigraph may reflect the reasoning of an $L1$ poisoner, or equivalently of an $L2$ investigator reasoning about an $L1$ poisoner.

The game referred to in our second epigraph, between a test-taker trying to guess the answer and a test-maker trying to minimize a guesser's chances, is exactly like RTH's games if location "c" is treated as equivalent to central A. Thus our second epigraph may reflect the reasoning of an $L1$ test-maker, or an $L2$ guesser reasoning about an $L1$ test-maker. Attali and Bar-Hillel (2003) present intriguing field evidence on this game, from the Israeli analog of the U.S. Scholastic Aptitude Test.

Like equilibrium with perturbations, our level- k model is quite flexible, and the freedom to specify $L0$ makes it even more flexible. Our payoff-insensitive $L0$ assumes that a player processes the game's decisions and framing before its payoffs. We show in Section IV that two alternative level- k models fit RTH's data slightly better than our proposed model: one that merges those processing steps as in Bacharach and Stahl's (1997a) payoff-sensitive, role-asymmetric $L0$ in which seekers favor salience and hidiers avoid it; and one with a payoff-insensitive, role-symmetric $L0$ that *avoids* salience.¹³ Section II's equilibrium with unrestricted perturbations model fits RTH's data best of all the models discussed here, but still only slightly better than our proposed model.

Despite our proposed model's slightly worse fit in RTH's dataset, we find its explanation of their results more convincing because its assumptions seem behaviorally more plausible and it does not rely on unexplained role differences in behavior or payoffs. To our knowledge, only a level- k model with a role-symmetric $L0$ can explain the robust patterns in RTH's data without unexplained differences in behavioral assumptions across roles. A payoff-insensitive $L0$ that avoids salience seems perverse, given that it responds only to decisions and framing. And a payoff-sensitive, role-asymmetric $L0$ suffers from the same over-flexibility and lack of theory to guide its specification as equilibrium with perturbations. Sections V's and VI's analyses give support for these subjective judgments by showing that our proposed model has advantages in overfitting and portability. (These issues are discussed in more detail in the web appendix.)

¹³Bacharach and Stahl (1997a) analyze a simplified version of RTH's game with three locations, one clearly salient and one less clearly salient. They give an evolutionary justification for their role-asymmetric $L0$: "For early humans, 'looking' problems were more generic and 'hiding' problems more strategic" (their footnote 13). Such a model can explain the main patterns in RTH's results with only types $L0$, $L1$, and $L2$, if there are more $L2$ than $L0$ subjects. This type distribution seems lower and therefore closer to previous estimates than ours; but the difference is largely semantic because the behavior of their $L0$ is similar to that of our $L1$, and their $L1$ to that of our $L2$, and so on.

IV. Econometric Analysis

This section describes our econometric specification and then separately estimates the equilibrium with perturbations model and each possible level- k model econometrically, using the pooled data from RTH's six treatments. We compare our proposed level- k model, which has a role-symmetric $L0$ that favors salience, with level- k alternatives that relax the constraint that $L0$ is role-symmetric, that $L0$ favors salience, or both. Our goal is to illustrate the models' possibilities, not to take a definitive position on the behavioral parameters. Estimating the models rather than calibrating them constrains our discretion, and yields likelihoods that provide an objective criterion by which to compare models.

Our econometric model is a mixture model as in Stahl and Wilson (1994, 1995) or Costa-Gomes et al. (2001), with one type for equilibrium with perturbations and up to four for the level- k models. Let X_{ij} be the total numbers of hiders or seekers (with $i = h, s$) who choose location j (with A, B, A, A denoted 1, 2, 3, 4). Let π_k be the probability that a given subject is type k , and ρ_{ikj} be the probability that a hider or seeker of type k chooses j . The full-sample likelihood can be written:

$$(2) \quad L \equiv \prod_{i=h,s} \prod_{j=1,2,3,4} \left[\sum_k \pi_k \rho_{ikj} \right]^{X_{ij}} .$$

The equilibrium with perturbations models have ρ_{ikj} as in Section II's analysis. Because the equilibrium that best fits RTH's results is totally mixed, errors would be redundant in this case. In the level- k models a player's type determines his normal choices as in Section III's analysis, with ties broken uniformly randomly, so with ρ_{ikj} as illustrated for our proposed model in Table 2. To avoid specification bias we use the simplest possible error structure: With probability ε a player makes an error, in which case he chooses his location uniformly randomly; errors are independent across players. The population choice probabilities are determined as functions of the behavioral parameters as in Table 2, with $\pi_0 = r$, $\pi_1 = s$, $\pi_2 = t$, $\pi_3 = u$, and $\pi_4 = v$ in Section III's notation.

In the estimation, we rule out knife-edge (p, q) combinations that make $L1$ hiders or seekers indifferent between locations, because they implicitly allow $L1$ to randomize, risking specification bias.¹⁴ We further constrain $r = 0$. This constraint is not binding for our proposed model, but

¹⁴We distinguish between ties that occur because two location choices are expected to yield the same outcome, which must be broken randomly in any consequentialist theory; and ties that exist only for knife-edge (p, q) combinations. With a role-symmetric $L0$ the likelihood is occasionally maximized at such combinations because they make $L1$ randomize in a way that happens to fit the data better than our (deliberately naïve) error structure. We view such maxima as spurious because it is implausible that there are very many subjects with such knife-edge combinations.

without it many of the alternative models have identification problems and those with a role-asymmetric $L0$ can achieve a near-perfect fit with $r = 1$, “explaining” the data by tabulating them.¹⁵

Given $r = 0$, estimating a level- k model amounts to choosing the (p, q) combination for $L0$, or equivalently the implied normal choices for $L1$ seekers and hidiers, such that the $L1$ choices and the choices they imply for higher types maximize the likelihood, given the restrictions imposed. Figure 3 graphs the regions in (p, q) -space for which $L0$ hidiers or seekers yield each possible combination of choices for $L1$ hidiers and seekers, numbered 1 through 6. In the most general level- k model, with no restriction on the role-symmetry of $L0$ or its response to salience, $L0$ hidiers and seekers can be independently assigned to regions 1 through 6, or equivalently (Figure 3) $L1$ seekers’ and hidiers’ choices can be chosen independently from {end As, central A, B}, which yields 9 possible cases.¹⁶ With a role-symmetric $L0$ but no restriction on its response to salience, $L0$ hidiers’ and seekers’ regions must be the same but can otherwise be chosen freely, which yields 6 possible cases. In our proposed model, with a role-symmetric $L0$ that favors salience, this joint choice is restricted to regions 1 and 2, in which $L1$ hidiers choose central A and $L1$ seekers choose either B (region 1, $p < 2q$) or end As (region 2, $p > 2q$). The ρ_{ikj} in these regions can be read, for given type frequencies and error rate, from Table 2 with $r = 0$; and the ρ_{ikj} in other regions can be deduced from Figure 3.¹⁷

Table 3 summarizes the parameter estimates, likelihoods, and predicted choice frequencies for all models, with the observed frequencies for comparison. All the models are flexible enough to fit RTH’s data very well.¹⁸ Equilibrium with unrestricted perturbations fits best. Next best is a level- k model with a role-symmetric $L0$ that avoids salience, with $L0$ hidiers and seekers both in region 5. This model is equivalent when estimated to the most general level- k model, with $L0$

¹⁵Only “near-perfect” because our models restrict $L0$ ’s choice probabilities to be the same for end A locations within each player role, so their predicted frequencies are also the same, which is not quite true in the data. Equilibrium with perturbations imposes an analogous restriction, so this should not bias the comparisons. Without the $r = 0$ constraint all maximum-likelihood estimates for models with a role-asymmetric $L0$ are equivalent to $r = 1$. Our proposed model estimates $r = 0$ because, while a role-asymmetric $L0$ can fit hidiers and seekers independently, a role-symmetric $L0$ cannot fit their different choices as well as higher types can. Our finding that there are no $L0$ subjects is consistent with the common finding that people underestimate others’ sophistication relative to their own (Georg Weizsäcker (2003)).

¹⁶The most general level- k model falls short of full generality only because we define the types as discussed in footnote 9, restrict $L0$ ’s choice probabilities to be the same for end A locations within each player role, impose $r = 0$, and rule out (p, q) combinations that make $L1$ hidiers or seekers indifferent between locations.

¹⁷If the regions defined by p and q for hidiers and seekers are each viewed as a single, discrete parameter, the most general level- k model has 6 independent parameters: a region each for hidiers and seekers, 3 independent type frequencies, and an error rate. Restrictions on $L0$ ’s response to salience do not affect this, while a restriction to a role-symmetric $L0$ as in our proposed model reduces it to 5. Equilibrium with unrestricted perturbations, with no error rate, has 4 parameters; and equilibrium with restricted perturbations has 2.

¹⁸The likelihood gap is larger than it may appear, because the log-likelihood associated with perfect prediction of the choice frequencies is not 0 but -1561.7 —negative because in these models even perfect prediction is only probabilistic.

hidiers in region 4 or equivalently 5 and $L0$ seekers in region 6 or 5. Next is a model with a role-asymmetric $L0$ with seekers favoring salience, in region 1 or 2, and hidiers avoiding it, in region 4 or 5, as in Bacharach and Stahl (1997a).¹⁹ Next is our proposed model with a role-symmetric $L0$ that favors salience, with $L0$ hidiers and seekers both in region 2, which fits only slightly worse than the three previous models. Equilibrium with restricted perturbations fits worst of all.²⁰

The estimated payoff perturbations in the equilibrium models have the expected positive signs, with unrestricted magnitudes nearly twice as large for hidiers as seekers. The estimated type distribution in our proposed level- k model is behaviorally plausible, with the characteristic hump-shape of previous estimates but a shift toward higher, more sophisticated types (Section III). In this case the restriction $r = 0$ is non-binding, so the model also reaffirms previous findings that $L0$ exists mainly in the minds of higher types. $L0$, in region 2, responds more to the salience of an end A than a B location. $\varepsilon = 0$, so types $L1$ to $L4$, taken together, explain the data better than uniform errors.²¹

The estimates for the alternative Bacharach and Stahl-style model, with a role-asymmetric $L0$ with seekers favoring salience and hidiers avoiding it, are somewhat perverse, with plausible type frequencies but an error rate of $\varepsilon = 0.72$, and identification problems even when $r = 0$ (footnote 19).

The estimates for the best-fitting alternative level- k model, with a role-symmetric $L0$ that avoids salience, are similar to those for our proposed model, again with $\varepsilon = 0$ but with an implausible type distribution of 36% $L1$, 9% $L2$, 36% $L3$, and 18% $L4$ players, far from the typical hump shape. Ultimately, this alternative model will stand or fall on the plausibility of a payoff-insensitive $L0$ that avoids salience and on its overfitting and portability performance.

V. Overfitting

This section tests for overfitting by using each model to compute separate estimates for each of RTH's six treatments and using the re-estimated models to “predict” the choice frequencies in

¹⁹In this model subscripts H and S distinguish $L0$ hidiers' and seekers' choice probabilities. This model is unidentified even when $r = 0$, because there are linear dependencies among types' choice probabilities that are the same for hidiers and seekers. We report an estimate constraining $u = v = 0$ (ruling out $L3$ and $L4$), following Bacharach and Stahl (1997a) in including only the lowest types, on the grounds that they are behaviorally more plausible. This choice has no substantive implications in RTH's dataset, but it matters in Section VI's analysis of portability.

²⁰We strongly reject the restrictions of equal magnitudes across roles (p -value 0.0022). We focus below on equilibrium with unrestricted perturbations, which is equivalent in RTH's games to QRE with unrestricted perturbations.

²¹Without payoff perturbations our uniform errors are perfectly confounded with the equilibrium mixed strategies. Thus $\varepsilon = 0$ also suggests the absence of subjects who play equilibrium strategies, and it allows us to reject explanations in which some of the population choose locations with given probabilities (like $L0$) and the rest, like the *Sophisticated* type in Crawford (2003), play equilibrium in a game among themselves, taking those probabilities into account.

the other five treatments. We evaluate the models by their mean squared prediction errors (“MSEs”).

Table 4 gives the MSEs and Table A2 in the web appendix reports the estimates on which they are based. Even though our proposed level- k model fits slightly worse than each alternative, it has a lower MSE than each alternative but the Bacharach and Stahl-style level- k model with a role-asymmetric $L0$ with seekers favoring salience and hidiers avoiding it, whose error is 10% lower.

VI. Portability

This section takes up the issue of portability, the extent to which estimating the model for RTH’s games helps to predict behavior in other games. We use each model, with the parameters estimated for RTH’s games, to “predict” subjects’ initial responses in the closest analogs of RTH’s experiments we know of, O’Neill’s (1987) card-matching experiment and Rapoport and Boebel’s (1992) closely related experiment. (Their data are included in our data appendix, with permission.) Their games provide a good test of portability, raising the same strategic issues as RTH’s games but with different framing, in one case five locations, and more complex win-loss patterns (Figures 4 and 5).

O’Neill and Rapoport and Boebel presented their games to subjects as stories, with locations ordered as in Figures 4 and 5. Rapoport and Boebel’s subjects were also given a matrix like Figure 5, but with Ws or Ls for wins or losses.²² O’Neill’s experiment had 25 subjects per role and Rapoport and Boebel’s had 10. Subjects played the games repeatedly in fixed pairs, with feedback after each play. We focus on subjects’ first-round choices, interpreting them as initial responses.²³

Tables 5 and 6 give the games’ unique equilibrium mixed strategies and subjects’ aggregate first-round choices. Equilibrium reflects the payoff-symmetry of actions A, 2, and 3 in O’Neill’s game and of actions F, I, and O in Rapoport and Boebel’s game. Equilibrium is also symmetric

²²O’Neill’s story (p. 2708) was: “Each player has four cards—Ace, two, three, and a joker. . . . [Player 1] wins if there is a match of jokers (two jokers played) or a mismatch of number cards (two, three, for example). [Player 2] wins if there is a match of number cards (three, three, for example) or a mismatch of a joker (one joker, one number card).” This wording leaves room for doubt whether Ace was also a “number” card, but his practice rounds made it clear to the subjects that it was. O’Neill had only one treatment, while Rapoport and Boebel had two, which differed only in the scaling and expected magnitudes of payoffs in ways that do not affect the predictions of the theories considered here. There is a statistically significant difference between Rapoport and Boebel’s treatments for player 2s (p -value 0.0087) but not for player 1s (p -value 0.8557). We keep their treatments separate in both player roles.

²³Rapoport and Boebel’s subjects played first in one player role, then the other. We use only a given subject’s first response to the game, in either role. Even so, interpreting first-round choices as initial responses is less straightforward here than for RTH’s subjects, who played a game only once or in a series without feedback, because with feedback it is theoretically possible for first-round choices to influence future play (even with random pairing, via “contagion”). Nonetheless, future influences do not seem to have been an important source of distortion in this case.

across player roles, as in RTH's games. We keep symmetric actions separate because equilibrium with perturbations and the level- k models break the symmetries in response to differences in salience. Relative to equilibrium, O'Neill's subjects have a large, positive Joker ("J") effect for both players, with J even more prevalent for player 2s.²⁴ Rapoport and Boebel's subjects have a large, positive I effect for player 1s in both treatments; a large, positive C effect for player 2s in treatment 1; and a large, positive L effect for player 2s in treatment 2.²⁵

In adapting the equilibrium model's perturbations to O'Neill's and Rapoport and Boebel's games, we use the same general notions of salience and players' instinctive reactions to it as for RTH's games. Although O'Neill's game is not a hide-and-seek game, player 1 (row) can be viewed as a hider when he chooses one of the number cards A, 2, and 3 and a seeker when he chooses J; and player 2 as the reverse. The end positions of A and J, and J's unique role in the payoff structure, reinforce the salience of their labels. We assume a player choosing a salient card for which he is a seeker receives an additional payoff of α for A and ι for J; and one choosing a salient card for which he is a hider loses analogous payoffs (Figure 6). We constrain $\alpha, \iota \geq 0$, so players favor salient cards for which they are seekers and avoid salient cards for which they are hidiers. We also allow α and ι to differ across roles, but we suppress this for ease of notation.

If $3\alpha - \iota < 1$ and $\alpha + 3\iota < 2$, the perturbed O'Neill game has a unique, symmetric, totally mixed equilibrium in which player 1s and player 2s both play A, 2, 3, and J with probabilities $(1-3\alpha+\iota)/5$, $(1+2\alpha+\iota)/5$, $(1+2\alpha+\iota)/5$, and $(2-\alpha-3\iota)/5$, with obvious changes if α and ι differ across roles. Either way, the probability of J is maximized subject to $\alpha, \iota \geq 0$ when $\alpha = \iota = 0$, which yields an equilibrium J probability of 0.4, well below the observed frequency in each role. Thus, equilibrium with perturbations, even with new estimates of the perturbations for O'Neill's game, cannot explain his subjects' initial responses better than equilibrium without perturbations, which explains them poorly.

In adapting level- k models to O'Neill's game we define $L0$ using the same notions of salience and players' responses to it as for RTH's games, and the same estimated type frequencies. Thus in our proposed model with a role-symmetric $L0$ that favors salience, $L0$ chooses A, 2, 3, J with

²⁴Discussions of O'Neill's data have been dominated by a small, positive "Ace effect" when the data are aggregated over all rounds (player 1s and 2s played A 22.0% and 22.6% of the time, versus the equilibrium 20%); see for example McKelvey and Palfrey (1995, p. 20). The Joker effect for initial responses, which is an order of magnitude larger, came as a surprise—a fortunate one for us because an Ace effect for player 1s is hard to square with a plausible level- k model (Table A3 in the web appendix). Although O'Neill speculated that "players were attracted by the powerful connotations of an Ace," our analysis suggests that the aggregate Ace effect is due to learning, not salience.

²⁵We put less weight on Rapoport and Boebel's effects than on O'Neill's Joker effect because of Rapoport and Boebel's smaller sample sizes and the large differences in choice frequencies across their two treatments.

probabilities a , $(1-a-j)/2$, $(1-a-j)/2, j$ (treating 2 and 3 equally), where $a > 1/4$ and $j > 1/4$. The model with a role-symmetric $L0$ that avoids salience is the same, but with $a < 1/4$ and $j < 1/4$. The Bacharach and Stahl-style model with a role-asymmetric $L0$ has $a_1 < 1/4$ and $j_1 > 1/4$ for player 1s (hidiers when they choose A, seekers when they choose J) and $a_2 > 1/4$ and $j_2 < 1/4$ for player 2s.²⁶

Table 5 summarizes the observed choice frequencies and the alternative models' predicted frequencies and MSEs for O'Neill's game, with equilibrium (without perturbations) as the best-fitting equilibrium model, and using Table 3's RTH estimates of s , t , u , v , and ε for the level- k models.²⁷ Our proposed level- k model with a role-symmetric $L0$ that favors salience has lower MSE than any of the alternatives. This model makes J modal for player 1s and 2s and in one case reproduces its greater prevalence for 2s, so the same behavioral assumptions we used to explain the prevalence of central A and its greater prevalence for seekers in RTH's experiments explain O'Neill's Joker effect.

Adapting the equilibrium with perturbations model to Rapoport and Boebel's game raises new difficulties. Although their game is very close in structure to RTH's and O'Neill's games, it no longer makes a player unambiguously a hider or seeker for all choices, and so allows no plausible, parsimonious parameterization of perturbations.²⁸ We could use a more flexible parameterization and just "let the data speak." But this would take us further from the general notions of instinctive reactions to salience we used to adapt the perturbations from RTH's to O'Neill's game, and even a parameterization as tight as we used for those games tends to overfit (Section V). The prospects for a useful equilibrium with perturbations analysis of Rapoport and Boebel's results seem dim.

In adapting level- k models to Rapoport and Boebel's game, we assume that their $L0$ s respond to salience qualitatively as in RTH's and O'Neill's games. We assume that the abstract labels C, L,

²⁶Our proposed model's choices remain the same whenever $3a + j > 1$, even if $a < 1/4$ or $j < 1/4$, with similar relaxations for the alternative model with a role-symmetric $L0$ that avoids salience and the alternative Bacharach and Stahl-style model with a role-asymmetric $L0$ that favors salience for seekers but avoids it for hidiers. Figure A3 in the web appendix, like Figure 3 for RTH's games, graphs the regions in (a, j) -space for which $L0$ yields each possible choice for $L1$ player 2s or 1s. For each model, we choose the region that is consistent with its constraints and that yields the best fit, given the type frequencies estimated for that model from RTH's data. In our proposed model, $a > 1/4$ and $j > 1/4$ (or the weaker condition $3a + j > 1$) restrict $L1$ player 1s and 2s to one of three choice combinations: J for player 1s and A for player 2s (when $3j - a > 1$); 2,3 for player 1s and A for player 2s (when $3j - a < 1$ and $a + 2j > 1$); or 2,3 for player 1s and J for player 2s (when $a + 2j < 1$, which given $3a + j > 1$ implies $3j - a < 1$). In the model with a role-symmetric $L0$ that avoids salience, $L1$ player 1s and 2s are similarly restricted to A for player 1s and J for player 2s. In the Bacharach and Stahl-style model, $L1$ player 1s' and 2s' choices can be chosen independently from {A; 2,3} for player 1s and {A; 2,3; J} for player 2s.

²⁷Tables A3-A4 in the web appendix, like Table 2 for RTH's games, gives types' expected payoffs and choice probabilities for our proposed model in O'Neill's game in the leading regions.

F, I, O are not salient per se, but that the end and center locations are inherently salient (Attali and Bar-Hillel (2004)). Thus, in our proposed model with a role-symmetric $L0$ that favors salience, $L0$ player 1s and 2s both choose C, L, F, I, O with probabilities $m/2$, $(1-m-n)/2$, n , $(1-m-n)/2$, $m/2$ (again treating end locations equally), with $m > 2/5$ and $n > 1/5$. The alternative model with a role-symmetric $L0$ that avoids salience is the same, but with $m < 2/5$ and $n < 1/5$.²⁹ However, specifying a Bacharach and Stahl-style model with a payoff-sensitive, role-asymmetric $L0$ raises difficulties like those for equilibrium with perturbations, in that there are no natural restrictions on m and n .

Table 6 summarizes the observed choice frequencies and the alternative models' predicted frequencies and MSEs for Rapoport and Boebel's game, with equilibrium (without perturbations) as the best-fitting equilibrium model, and using Table 3's estimates of s , t , u , v , and ε for the level- k models, choosing among regions as for O'Neill's game (footnote 26). None of the models fit well, which may reflect the large variation in player 2s' frequencies of C and L across their treatments. Our proposed level- k model has MSEs lower than equilibrium and the level- k model that avoids salience except for player 2s in treatment 2, which are best captured by equilibrium. Our proposed model reproduces a small fraction of the I effect for player 1s in both treatments and the C effect for player 2s in treatment 1, but it completely misses the L effect for player 2s in treatment 2.

VII. Conclusion

This paper has compared alternative explanations of the systematic deviations from equilibrium in RTH's experiments with hide-and-seek games with non-neutral framing of locations, and in O'Neill's and Rapoport and Boebel's experiments with closely related games. Our analysis explores the issues that arise in specifying level- k and alternative models for games played on non-neutral landscapes. It then shows that a structural non-equilibrium model of initial responses based on "level- k " thinking, with a payoff-insensitive anchoring $L0$ type, can explain RTH's results, including their puzzling role-asymmetries, without assuming differences in behavior across roles.

Our proposed model fits RTH's data slightly worse than some of the alternatives, but we find its explanation of their results more convincing because its assumptions seem behaviorally more plausible and it (alone) does not rely on unexplained role differences in behavior or payoffs. Our

²⁸Player 1 (2) is a seeker (hider) for location C. When C is eliminated, player 1 (2) is a hider (seeker) for location L. But even when L and C are eliminated, the player roles for location F cannot be classified this way.

analyses of overfitting and portability give support for these subjective judgments. They also trace the model's advantages to two features. First, it is based on general decision rules or “types” that apply to any game, whose population frequencies tend to be stable across different games. Second, its LO is based on simple principles—how salience is determined by the set of decisions and their framing, and how people respond to it—for which there is strong support, whose simplicity facilitates transfer to new games, just as the sensitivity to the details of the structure of alternative specifications of LO or the payoff perturbations in an equilibrium model inhibit transfer.

²⁹In each case $L1$ Player 1s' and 2s' choices remain the same under weaker conditions. Table A5 in the web appendix gives types' expected payoffs and choice probabilities for our proposed model in Rapoport and Boebel's game in the leading regions.

References

- Attali, Yigal, and Maya Bar-Hillel. 2003. "Guess Where: The Position of Correct Answers in Multiple-Choice Test Items as a Psychometric Variable." *Journal of Educational Measurement*, 40(2): 109-128.
- Bacharach, Michael, and Dale Stahl. 1997a. "Variable-Frame Level- N Theory." Unpublished; posted at <http://www.eco.utexas.edu/~stahl/vflnt4a.pdf>.
- Bacharach, Michael, and Dale Stahl. 1997b. "Variable-Frame Level- n Theory with Application to Hide and Seek." Institute of Economics and Statistics, University of Oxford, Working Paper.
- Bacharach, Michael, and Dale Stahl. 2000. "Variable-Frame Level- n Theory." *Games and Economic Behavior*, 32(2): 220–246.
- Bardsley, Nicholas, Judith Mehta, Chris Starmer, and Robert Sugden (2006): "The Nature of Salience Revisited: Cognitive Hierarchy Theory versus Team Reasoning." Unpublished.
- Camerer, Colin, Teck-Hua Ho, and Juin Kuan Chong. 2004. "A Cognitive Hierarchy Model of Games." *Quarterly Journal of Economics*, 119(3): 861-898.
- Chivers, C. J. 2004. "A Dinner in Ukraine Made for Agatha Christie." *The New York Times*, December 20, 2004, A1.
- Christenfeld, Nicholas. 1995. "Choices from Identical Options." *Psychological Science*, 6(1), 50-55.
- Costa-Gomes, Miguel, Vincent Crawford, and Bruno Broseta. 2001. "Cognition and Behavior in Normal-Form Games: An Experimental Study." *Econometrica*, 69(5): 1193-1235.
- Costa-Gomes, Miguel, and Vincent Crawford. 2006. "Cognition and Behavior in Two-Person Guessing Games: An Experimental Study." *American Economic Review*, 96(5): 1737-1768.
- Crawford, Vincent. 2003. "Lying for Strategic Advantage: Rational and Boundedly Rational Misrepresentation of Intentions." *American Economic Review*, 93(1): 133-149.
- Crawford, Vincent, Uri Gneezy, and Yuval Rottenstreich. 2007. "The Power of Focal Points is Limited: Even Minute Payoff Asymmetry May Yield Large Coordination Failures." Unpublished.
- Crawford, Vincent P., and Nagore Iriberrí. 2005. "Level- k Auctions: Can a Non-Equilibrium Model of Strategic Thinking Explain the Winner's Curse and Overbidding in Private-Value Auctions?" Unpublished.

- Ho, Teck-Hua, Colin Camerer, and Keith Weigelt. 1998. "Iterated Dominance and Iterated Best Response in Experimental '*p*-Beauty Contests'." *American Economic Review*, 88(4): 947-969.
- Keillor, Garrison. 1997. *Wobegon Boy*. New York: Penguin.
- McKelvey, Richard, and Thomas Palfrey. 1995. "Quantal Response Equilibria for Normal-Form Games." *Games and Economic Behavior*, 10(1): 6-38.
- Mehta, Judith, Chris Starmer, and Robert Sugden. 1994. "The Nature of Salience: An Experimental Investigation of Pure Coordination Games." *American Economic Review*, 84(3): 658-673.
- Nagel, Rosemarie. 1995. "Unraveling in Guessing Games: An Experimental Study." *American Economic Review*, 85(5): 1313-1326.
- O'Neill, Barry. 1987. "Nonmetric Test of the Minimax Theory of Two-Person Zerosum Games." *Proceedings of the National Academy of Sciences of the United States of America*, 84(7): 2106-2109.
- Rapoport, Amnon, and Richard Boebel. 1992. "Mixed Strategies in Strictly Competitive Games: A Further Test of the Minimax Hypothesis." *Games and Economic Behavior*, 4(2): 261-283.
- Rosenthal, Robert, Jason Shachat, and Mark Walker. 2003. "Hide and Seek in Arizona." *International Journal of Game Theory*, 32(2): 273-293.
- Rubinstein, Ariel, and Amos Tversky. 1993. "Naïve Strategies in Zero-Sum Games." Sackler Institute of Economic Studies, Tel Aviv University, Working Paper 17-93.
- Rubinstein, Ariel, Amos Tversky, and Dana Heller. 1996. "Naive Strategies in Competitive Games," in *Understanding Strategic Interaction—Essays in Honor of Reinhard Selten*, ed. Wulf Albers, Werner Güth, Peter Hammerstein, Bemy Moldovanu, and Eric van Damme, 394-402. Berlin: Springer-Verlag.
- Rubinstein, Ariel. 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; see especially "Second Edition" posted at <http://arielrubinstein.tau.ac.il/99/gt100.html>.
- Scharlemann, Jörn, Catherine Eckel, Alex Kacelnik, and Rick Wilson. 2001. "The Value of a Smile: Game Theory with a Human Face." *Journal of Economic Psychology*, 22(5): 617-640.
- Schelling, Thomas. 1960. *The Strategy of Conflict*. Cambridge: Harvard University Press.
- Stahl, Dale, and Paul Wilson. 1994. "Experimental Evidence on Players' Models of Other Players." *Journal of Economic Behavior and Organization*, 25(3): 309-327.

- Stahl, Dale, and Paul Wilson. 1995. "On Players' Models of Other Players: Theory and Experimental Evidence." *Games and Economic Behavior*, 10(1): 218-254.
- Von Neumann, John. 1953. "A Certain Zero-Sum Two-Person Game Equivalent to the Optimal Assignment Problem," in *Contributions to the Theory of Games, Vol. II*, ed. H.W. Kuhn and A.W. Tucker, 5-12. *Annals of Mathematics Studies*, 28, Princeton, N.J.: Princeton Univ. Press.
- Weizsäcker, Georg. 2003. "Ignoring the Rationality of Others: Evidence from Experimental Normal-Form Games." *Games and Economic Behavior*, 44(1): 145-171.

Table 1. Aggregate Choice Frequencies in RTH's Treatments

| | | | | |
|----------------------------|----------|----------|----------|----------|
| RTH-4 | A | B | A | A |
| Hider (53; $p = 0.0026$) | 9% | 36% | 40% | 15% |
| Seeker (62; $p = 0.0003$) | 13% | 31% | 45% | 11% |
| RT-AABA-Treasure | A | A | B | A |
| Hider (189; $p = 0.0096$) | 22% | 35% | 19% | 25% |
| Seeker (85; $p = 9E-07$) | 13% | 51% | 21% | 15% |
| RT-AABA-Mine | A | A | B | A |
| Hider (132; $p = 0.0012$) | 24% | 39% | 18% | 18% |
| Seeker (73; $p = 0.0523$) | 29% | 36% | 14% | 22% |
| RT-1234-Treasure | 1 | 2 | 3 | 4 |
| Hider (187; $p = 0.0036$) | 25% | 22% | 36% | 18% |
| Seeker (84; $p = 3E-05$) | 20% | 18% | 48% | 14% |
| RT-1234-Mine | 1 | 2 | 3 | 4 |
| Hider (133; $p = 6E-06$) | 18% | 20% | 44% | 17% |
| Seeker (72; $p = 0.149$) | 19% | 25% | 36% | 19% |
| R-ABAA | A | B | A | A |
| Hider (50; $p = 0.0186$) | 16% | 18% | 44% | 22% |
| Seeker (64; $p = 9E-07$) | 16% | 19% | 54% | 11% |

Sample sizes and p -values for significant differences from equilibrium in parentheses; salient labels in italics; order of presentation of locations to subjects as shown.

Table 2. Types' Expected Payoffs and Choice Probabilities in RTH's Games when $p > 1/2$ and $q > 1/4$

| Hider | Exp. Payoff $p < 2q$ | Choice Pr. $p < 2q$ | Exp. Payoff $p > 2q$ | Choice Pr. $p > 2q$ | Seeker | Exp. Payoff $p < 2q$ | Choice Pr. $p < 2q$ | Exp. Payoff $p > 2q$ | Choice Pr. $p > 2q$ |
|--------------------------|--|------------------------|--|------------------------|--------------------------|--|------------------------|--|------------------------|
| <i>L0 (Pr. r)</i> | | | | | <i>L0 (Pr. r)</i> | | | | |
| A | - | $p/2$ | - | $p/2$ | A | - | $p/2$ | - | $p/2$ |
| B | - | q | - | q | B | - | q | - | q |
| A | - | $1-p-q$ | - | $1-p-q$ | A | - | $1-p-q$ | - | $1-p-q$ |
| A | - | $p/2$ | - | $p/2$ | A | - | $p/2$ | - | $p/2$ |
| <i>L1 (Pr. s)</i> | | | | | <i>L1 (Pr. s)</i> | | | | |
| A | $1 - p/2 < 3/4$ | 0 | $1 - p/2 < 3/4$ | 0 | A | $p/2 > 1/4$ | 0 | $p/2 > 1/4$ | 1/2 |
| B | $1 - q < 3/4$ | 0 | $1 - q < 3/4$ | 0 | B | $q > 1/4$ | 1 | $q > 1/4$ | 0 |
| A | $p + q > 3/4$ | 1 | $p + q > 3/4$ | 1 | A | $1-p-q < 1/4$ | 0 | $1-p-q < 1/4$ | 0 |
| A | $1 - p/2 < 3/4$ | 0 | $1 - p/2 < 3/4$ | 0 | A | $p/2 > 1/4$ | 0 | $p/2 > 1/4$ | 1/2 |
| <i>L2 (Pr. t)</i> | | | | | <i>L2 (Pr. t)</i> | | | | |
| A | 1 | 1/3 | 1/2 | 0 | A | 0 | 0 | 0 | 0 |
| B | 0 | 0 | 1 | 1/2 | B | 0 | 0 | 0 | 0 |
| A | 1 | 1/3 | 1 | 1/2 | A | 1 | 1 | 1 | 1 |
| A | 1 | 1/3 | 1/2 | 0 | A | 0 | 0 | 0 | 0 |
| <i>L3 (Pr. u)</i> | | | | | <i>L3 (Pr. u)</i> | | | | |
| A | 1 | 1/3 | 1 | 1/3 | A | 1/3 | 1/3 | 0 | 0 |
| B | 1 | 1/3 | 1 | 1/3 | B | 0 | 0 | 1/2 | 1/2 |
| A | 0 | 0 | 0 | 0 | A | 1/3 | 1/3 | 1/2 | 1/2 |
| A | 1 | 1/3 | 1 | 1/3 | A | 1/3 | 1/3 | 0 | 0 |
| <i>L4 (Pr. v)</i> | | | | | <i>L4 (Pr. v)</i> | | | | |
| A | 2/3 | 0 | 1 | 1/2 | A | 1/3 | 1/3 | 1/3 | 1/3 |
| B | 1 | 1 | 1/2 | 0 | B | 1/3 | 1/3 | 1/3 | 1/3 |
| A | 2/3 | 0 | 1/2 | 0 | A | 0 | 0 | 0 | 0 |
| A | 2/3 | 0 | 1 | 1/2 | A | 1/3 | 1/3 | 1/3 | 1/3 |
| Total | $P < 2q$ | | $p > 2q$ | | Total | $p < 2q$ | | $p > 2q$ | |
| A | $Rp/2+(1-\varepsilon)[t/3+u/3]+(1-r)\varepsilon/4$ | | $rp/2+(1-\varepsilon)[u/3+v/2]+(1-r)\varepsilon/4$ | | A | $rp/2+(1-\varepsilon)[u/3+v/3]+(1-r)\varepsilon/4$ | | $rp/2+(1-\varepsilon)[s/2+v/3]+(1-r)\varepsilon/4$ | |
| B | $Rq+(1-\varepsilon)[u/3+v]+(1-r)\varepsilon/4$ | | $rq+(1-\varepsilon)[t/2+u/3]+(1-r)\varepsilon/4$ | | B | $rq+(1-\varepsilon)[s+v/3]+(1-r)\varepsilon/4$ | | $rq+(1-\varepsilon)[u/2+v/3]+(1-r)\varepsilon/4$ | |
| A | $r(1-p-q)+(1-\varepsilon)[s+t/3]+(1-r)\varepsilon/4$ | | $r(1-p-q)+(1-\varepsilon)[s+t/2]+(1-r)\varepsilon/4$ | | A | $r(1-p-q)+(1-\varepsilon)[t+u/3]+(1-r)\varepsilon/4$ | | $r(1-p-q)+(1-\varepsilon)[t+u/2]+(1-r)\varepsilon/4$ | |
| A | $Rp/2+(1-\varepsilon)[t/3+u/3]+(1-r)\varepsilon/4$ | | $rp/2+(1-\varepsilon)[u/3+v/2]+(1-r)\varepsilon/4$ | | A | $rp/2+(1-\varepsilon)[u/3+v/3]+(1-r)\varepsilon/4$ | | $rp/2+(1-\varepsilon)[s/2+v/3]+(1-r)\varepsilon/4$ | |

Table 3. Parameter Estimates and Likelihoods for the Leading Models in RTH's Games

| Model | Ln L | Parameter estimates | Observed or predicted choice frequencies | | | | | MSE |
|---|---------|---|--|--------|--------|--------|--------|---------|
| | | | Pl. | A | B | A | A | |
| Observed frequencies (624 hidiers, 560 seekers) | | | H | 0.2163 | 0.2115 | 0.3654 | 0.2067 | - |
| | | | S | 0.1821 | 0.2054 | 0.4589 | 0.1536 | |
| Equilibrium without perturbations | -1641.4 | | H | 0.2500 | 0.2500 | 0.2500 | 0.2500 | 0.00970 |
| | | | S | 0.2500 | 0.2500 | 0.2500 | 0.2500 | |
| Equilibrium with restricted perturbations | -1568.5 | $e_H \equiv e_S = 0.2187$ $f_H \equiv f_S = 0.2010$ | H | 0.1897 | 0.2085 | 0.4122 | 0.1897 | 0.00084 |
| | | | S | 0.1897 | 0.2085 | 0.4122 | 0.1897 | |
| Equilibrium with unrestricted perturbations | -1562.4 | $e_H = 0.2910, f_H = 0.2535$ $e_S = 0.1539, f_S = 0.1539$ | H | 0.2115 | 0.2115 | 0.3654 | 0.2115 | 0.00006 |
| | | | S | 0.1679 | 0.2054 | 0.4590 | 0.1679 | |
| Level- k with a role-symmetric LO that favors salience | -1564.4 | $p > 1/2$ and $q > 1/4, p > 2q,$ $r = 0, s = 0.1896, t = 0.3185,$ $u = 0.2446, v = 0.2473, \varepsilon = 0$ | H | 0.2052 | 0.2408 | 0.3488 | 0.2052 | 0.00027 |
| | | | S | 0.1772 | 0.2047 | 0.4408 | 0.1772 | |
| Level- k with a role-asymmetric LO that favors salience for seekers and avoids it for hidiers | -1563.8 | $p_H < 1/2$ and $q_H < 1/4,$ $p_S > 1/2$ and $q_S > 1/4,$ $r = 0, s = 0.66, t = 0.34, \varepsilon = 0.72;$ $u \equiv v \equiv 0$ imposed | H | 0.2117 | 0.2117 | 0.3648 | 0.2117 | 0.00017 |
| | | | S | 0.1800 | 0.1800 | 0.4600 | 0.1800 | |
| Level- k with a role-symmetric LO that avoids salience | -1562.5 | $p < 1/2$ and $q < 1/4, p < 2q,$ $r = 0, s = 0.3636, t = 0.0944,$ $u = 0.3594, v = 0.1826, \varepsilon = 0$ | H | 0.2133 | 0.2112 | 0.3623 | 0.2133 | 0.00006 |
| | | | S | 0.1670 | 0.2111 | 0.4549 | 0.1670 | |

Table 4. MSEs Treatment by Treatment in RTH's Games

| Predicted treatment Estimated treatment | RTH-4 | RT-AABA- Treasure | RT-AABA- Mine | RT-1234- Treasure | RT-1234- Mine | R-ABAA |
|--|--------|----------------------|------------------|----------------------|------------------|--------|
| Level-k with symmetric $L0$ that favors salience: overall MSE 0.00341 | | | | | | |
| RTH-4 | 0.0020 | 0.0032 | 0.0098 | 0.0031 | 0.0019 | 0.0032 |
| RT-AABA-Treasure | 0.0070 | 0.0014 | 0.0029 | 0.0011 | 0.0004 | 0.0048 |
| RT-AABA-Mine | 0.0132 | 0.0042 | 0.0011 | 0.0029 | 0.0023 | 0.0085 |
| RT-1234-Treasure | 0.0072 | 0.0016 | 0.0029 | 0.0007 | 0.0002 | 0.0037 |
| RT-1234-Mine | 0.0054 | 0.0017 | 0.0035 | 0.0009 | 0.0000 | 0.0034 |
| R-ABAA | 0.0040 | 0.0023 | 0.0073 | 0.0016 | 0.0010 | 0.0023 |
| Equilibrium with unrestricted perturbations: overall MSE 0.00418 | | | | | | |
| RTH-4 | 0.0002 | 0.0088 | 0.0156 | 0.0079 | 0.0050 | 0.0087 |
| RT-AABA-Treasure | 0.0089 | 0.0001 | 0.0039 | 0.0013 | 0.0017 | 0.0022 |
| RT-AABA-Mine | 0.0153 | 0.0034 | 0.0005 | 0.0031 | 0.0032 | 0.0070 |
| RT-1234-Treasure | 0.0076 | 0.0009 | 0.0031 | 0.0005 | 0.0004 | 0.0025 |
| RT-1234-Mine | 0.0053 | 0.0018 | 0.0037 | 0.0009 | 0.0000 | 0.0036 |
| R-ABAA | 0.0085 | 0.0019 | 0.0071 | 0.0027 | 0.0032 | 0.0004 |
| Level-k with symmetric $L0$ that avoids salience: overall MSE 0.00359 | | | | | | |
| RTH-4 | 0.0035 | 0.0062 | 0.0081 | 0.0050 | 0.0029 | 0.0054 |
| RT-AABA-Treasure | 0.0091 | 0.0001 | 0.0040 | 0.0012 | 0.0018 | 0.0020 |
| RT-AABA-Mine | 0.0148 | 0.0033 | 0.0005 | 0.0028 | 0.0029 | 0.0067 |
| RT-1234-Treasure | 0.0069 | 0.0008 | 0.0028 | 0.0007 | 0.0003 | 0.0031 |
| RT-1234-Mine | 0.0054 | 0.0017 | 0.0035 | 0.0010 | 0.0001 | 0.0041 |
| R-ABAA | 0.0073 | 0.0008 | 0.0054 | 0.0021 | 0.0023 | 0.0010 |
| Level-k with asymmetric $L0$: overall MSE 0.00306 | | | | | | |
| RTH-4 | 0.0077 | 0.0016 | 0.0026 | 0.0009 | 0.0005 | 0.0025 |
| RT-AABA-Treasure | 0.0086 | 0.0007 | 0.0038 | 0.0007 | 0.0010 | 0.0023 |
| RT-AABA-Mine | 0.0085 | 0.0028 | 0.0017 | 0.0017 | 0.0007 | 0.0057 |
| RT-1234-Treasure | 0.0081 | 0.0009 | 0.0029 | 0.0005 | 0.0005 | 0.0026 |
| RT-1234-Mine | 0.0079 | 0.0014 | 0.0021 | 0.0007 | 0.0003 | 0.0034 |
| R-ABAA | 0.0096 | 0.0024 | 0.0068 | 0.0025 | 0.0031 | 0.0006 |

Table 5. Comparison of the Leading Models in O’Neill’s Game

| Model | Parameter estimates | Observed or predicted choice frequencies | | | | | MSE |
|--|---|--|--------|--------|--------|--------|--------|
| | | Player | A | 2 | 3 | J | |
| Observed frequencies (25 Player 1s, 25 Player 2s) | | 1 | 0.0800 | 0.2400 | 0.1200 | 0.5600 | - |
| | | 2 | 0.1600 | 0.1200 | 0.0800 | 0.6400 | - |
| Equilibrium without perturbations | | 1 | 0.2000 | 0.2000 | 0.2000 | 0.4000 | 0.0120 |
| | | 2 | 0.2000 | 0.2000 | 0.2000 | 0.4000 | 0.0200 |
| Level- k with a role-symmetric $L0$ that favors salience | $a > 1/4$ and $j > 1/4$ | 1 | 0.0824 | 0.1772 | 0.1772 | 0.5631 | 0.0018 |
| | $3j - a < 1, a + 2j < 1$ | 2 | 0.1640 | 0.1640 | 0.1640 | 0.5081 | 0.0066 |
| Level- k with a role-symmetric $L0$ that favors salience | $a > 1/4$ and $j > 1/4$ | 1 | 0.0000 | 0.2541 | 0.2541 | 0.4919 | 0.0073 |
| | $3j - a < 1, a + 2j > 1$ | 2 | 0.2720 | 0.0824 | 0.0824 | 0.5631 | 0.0050 |
| Level- k with a role-symmetric $L0$ that avoids salience | $a < 1/4$ and $j < 1/4$ | 1 | 0.4245 | 0.1807 | 0.1807 | 0.2142 | 0.0614 |
| | | 2 | 0.1670 | 0.1807 | 0.1807 | 0.4717 | 0.0105 |
| Level- k with a role-asymmetric $L0$ that favors salience for locations for which player is a seeker and avoids it for locations for which player is a hider | $a_1 < 1/4, j_1 > 1/4;$ $a_2 > 1/4, j_2 < 1/4$ | 1 | 0.1804 | 0.2729 | 0.2729 | 0.2739 | 0.0291 |
| | $3j_1 - a_1 < 1, a_1 + 2j_1 < 1,$ $3a_2 + j_2 > 1$ | 2 | 0.1804 | 0.1804 | 0.1804 | 0.4589 | 0.0117 |

Table 6. Comparison of the Leading Models in Rapoport and Boebel’s Game

| Model | Parameter estimates | Observed or predicted choice frequencies | | | | | | MSE, Tr. 1 | MSE, Tr. 2 |
|---|---------------------|--|--------|--------|--------|--------|--------|------------|------------|
| | | Pl. | C | L | F | I | O | | |
| Observed frequencies, Tr. 1 (10 Player 1s, 10 Player 2s) | | 1 | 0.1000 | 0.0000 | 0.2000 | 0.6000 | 0.1000 | - | - |
| | | 2 | 0.8000 | 0.0000 | 0.0000 | 0.1000 | 0.1000 | - | - |
| Observed frequencies, Tr. 2 (10 Player 1s, 10 Player 2s) | | 1 | 0.1000 | 0.1000 | 0.1000 | 0.6000 | 0.1000 | - | - |
| | | 2 | 0.2000 | 0.6000 | 0.2000 | 0.0000 | 0.0000 | - | - |
| Equilibrium without perturbations | | 1 | 0.3750 | 0.2500 | 0.1250 | 0.1250 | 0.1250 | 0.0740 | 0.0650 |
| | | 2 | 0.3750 | 0.2500 | 0.1250 | 0.1250 | 0.1250 | 0.0520 | 0.0380 |
| Level- k with a role-symmetric $L0$ that favors salience | $m > 2/5, n > 1/5$ | 1 | 0.3085 | 0.3488 | 0.0612 | 0.2204 | 0.0612 | 0.0660 | 0.0505 |
| | $3m/2 + n > 1$ | 2 | 0.4657 | 0.1593 | 0.0618 | 0.2514 | 0.0618 | 0.0331 | 0.0702 |
| Level- k with a role-symmetric $L0$ that favors salience | $m > 2/5, n > 1/5$ | 1 | 0.3796 | 0.4369 | 0.0612 | 0.0612 | 0.0612 | 0.1160 | 0.0970 |
| | $3m/2 + n < 1$ | 2 | 0.4107 | 0.2204 | 0.1230 | 0.1230 | 0.1230 | 0.0433 | 0.0449 |
| Level- k with a role-symmetric $L0$ that avoids salience | $m < 2/5, n < 1/5$ | 1 | 0.0944 | 0.5420 | 0 | 0.3636 | 0 | 0.0799 | 0.0543 |
| | $2m + 3n < 1$ | 2 | 0.4864 | 0.1812 | 0.1213 | 0.0898 | 0.1213 | 0.0293 | 0.0573 |
| Level- k with a role-symmetric $L0$ that avoids salience | $m < 2/5, n < 1/5$ | 1 | 0.1843 | 0.5462 | 0.0898 | 0.0898 | 0.0898 | 0.1156 | 0.0933 |
| | $2m + 3n > 1$ | 2 | 0.4565 | 0.1371 | 0.1355 | 0.1355 | 0.1355 | 0.0315 | 0.0642 |

| | | | | | |
|---------------------|---|----------|----------|----------|----------|
| Hider/Seeker | | A | B | A | A |
| A | 0 | 1 | 0 | 0 | 0 |
| B | 1 | 0 | 1 | 0 | 0 |
| A | 1 | 0 | 0 | 1 | 0 |
| A | 1 | 0 | 0 | 0 | 1 |

Figure 1. RTH's Hide-and-Seek Game

| | | | | | |
|---------------------|-----|----------|----------|----------|----------|
| Hider/Seeker | | A | B | A | A |
| A | 0-e | 1+e | 0+f | 1-e | 0+e |
| B | 1-f | 0+e | 1+f | 0 | 0+e |
| A | 1 | 0+e | 0+f | 1 | 0+e |
| A | 1-e | 0+e | 0+f | 0 | 1+e |

Figure 2. RTH's Hide-and-Seek Game with Payoff Perturbations

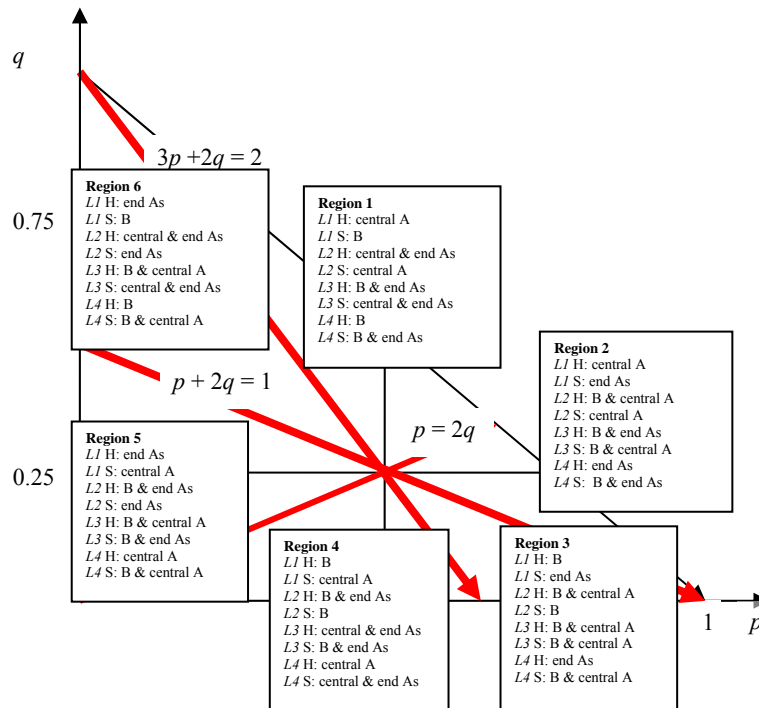


Figure 3. L1's Through L4's Choices in RTH's Games as Functions of L0's Choice Probabilities p and q

| Player 1/Player 2 | A | 2 | 3 | J |
|-------------------|-----|-----|-----|-----|
| A | 0 1 | 1 0 | 1 0 | 0 1 |
| 2 | 1 0 | 0 1 | 1 0 | 0 1 |
| 3 | 1 0 | 1 0 | 0 1 | 0 1 |
| J | 0 1 | 0 1 | 0 1 | 1 0 |

Figure 4. O'Neill's Card-Matching Game

| Player 1/Player 2 | C | L | F | I | O |
|-------------------|-----|-----|-----|-----|-----|
| C | 1 0 | 0 1 | 0 1 | 0 1 | 0 1 |
| L | 0 1 | 0 1 | 1 0 | 1 0 | 1 0 |
| F | 0 1 | 1 0 | 0 1 | 0 1 | 1 0 |
| I | 0 1 | 1 0 | 0 1 | 1 0 | 0 1 |
| O | 0 1 | 1 0 | 1 0 | 0 1 | 0 1 |

Figure 5. Rapoport and Boebel's Game

| Player 1/Player 2 | A | 2 | 3 | J |
|-------------------|-----------------------|--------------|--------------|----------------------|
| A | $0-\alpha$ $1+\alpha$ | $1-\alpha$ 0 | $1-\alpha$ 0 | $0-\alpha$ $1-\iota$ |
| 2 | 1 $0+\alpha$ | 0 1 | 1 0 | 0 $1-\iota$ |
| 3 | 1 $0+\alpha$ | 1 0 | 0 1 | 0 $1-\iota$ |
| J | $0+\iota$ $1+\alpha$ | $0+\iota$ 1 | $0+\iota$ 1 | $1+\iota$ $0-\iota$ |

Figure 6. O'Neill's Card-Matching Game with Payoff Perturbations

Web Appendix

This appendix provides background and more detail for “Fatal Attraction: Saliency, Naivety, and Sophistication in Experimental ‘Hide-and-Seek’ Games”.

Labels with Positive or Negative Connotations and/or Focally Labeled End Locations

Table A1 lists the choice frequencies from five additional RTH Treasure treatments with the same payoff structure as RTH-4 (Table 1), but labels with positive or negative connotations and/or focally labeled end locations. RTH-2 and RTH-5 are analogous to RTH-4 except for the connotations of the focal label. RTH-1 and RTH-3 are like RTH-4 except that the focal label is at an end position, and in RTH-3 it has a negative connotation. RTH-6 is like RTH-5 except that the focal label is in the third rather than second position; and is like RTH-2 and RTH-4 except for this difference in position and that the focal label has a positive connotation in RTH-6 but negative or neutral connotations in RTH-2 or RTH-4. The choice frequencies for these treatments echo those for the ones we analyze, with shifts in expected directions, and so provide additional evidence of the robustness of the patterns in RTH's data. It seems likely that our analysis could be extended to them by introducing and estimating payoff perturbations and/or new $L0$ choice probabilities.

Table A1. Aggregate Choice Frequencies in RTH's Experiments with Non-neutral Connotations

| RTH-1 | <i>Triangle</i> | Circle | Circle | Circle |
|--------------|-----------------|---------------|---------------|-----------------|
| Hider (53) | 23% | 23% | 43% | 11% |
| Seeker (62) | 29% | 24% | 42% | 5% |
| RTH-2 | Polite | <i>Rude</i> | Honest | Friendly |
| Hider (53) | 15% | 26% | 51% | 8% |
| Seeker (62) | 8% | 40% | 40% | 11% |
| RTH-3 | Smile | Smile | Smile | <i>Frown</i> |
| Hider (53) | 21% | 26% | 34% | 19% |
| Seeker (62) | 7% | 25% | 34% | 34% |
| RTH-5 | Frown | <i>Smile</i> | Frown | Frown |
| Hider (53) | 15% | 40% | 34% | 11% |
| Seeker (62) | 16% | 55% | 21% | 8% |
| RTH-6 | Hate | Detest | <i>Love</i> | Dislike |
| Hider (53) | 11% | 23% | 38% | 28% |
| Seeker (62) | 20% | 21% | 55% | 14% |

Sample sizes in parentheses; *salient* labels in italics; order of presentation of locations to subjects as shown.

Data Adjustments

In Table 1, we made minor adjustments to RTH's published data to reconcile reported frequencies and sample sizes. Hiders' choice frequencies in RT-AABA-Treasure and RT-1234-Treasure, and seekers' frequencies in RT-AABA-Mine, all sum to 101%; and hiders' frequencies in RT-AABA-Mine and hiders' and seekers' frequencies in RT-1234-Mine sum to 99%. We deal with this by translating the percentages into integer numbers of subjects and then rounding as needed. In RT-AABA-Treasure, for example, RT's reported percentages for hiders are 22%, 35%, 19%, and 25%, with reported sample size 189. Applying the rounded percentages to the sample size yields numbers of subjects 41.58, 66.15, 35.91, and 47.25, which round to 42, 66, 36, and 47, which sum to $191 > 189$. We rounded 41.58 down to 41 and 35.91 down to 35, which is the only way to reconcile the sample size and the rounded reported percentages. Similarly, in RT-AABA-Mine we rounded $0.39 \times 132 = 51.48$ up to 52, which allows us to reconcile the sample size and reported percentages. Finally, in R-ABAA the reported percentages for hiders are 16%, 18%, 45%, and 22%, which add to 101%, with a sample size of 50. This yields numbers of subjects 8, 9, 22.5, and 11, which add to 50.5. The only way to reconcile this with one typo is to adjust the 45% to 44% as in Table 1, yielding numbers of subjects 8, 9, 22, and 11.

Symmetry of Payoff Perturbations Across Roles for Equilibrium and QRE with Perturbations Models

Here we discuss, in more detail than in Section II, the issue of symmetry of payoff perturbations across player roles. We focus on the equilibrium with perturbations model, but our discussion also applies to the QRE with perturbations model.

There is no logical reason why a role-asymmetric payoff structure should not evoke instinctive aversions or attractions that differ in magnitudes as well as signs. The issue here is slightly different than for the symmetry across roles of $L0$ in our level- k model, discussed below, because we allow the signs of the perturbations to differ in keeping with Bacharach and Stahl's (1997ab) intuition, without which equilibrium with perturbations fits the data extremely poorly; the only question is whether to assume that their magnitudes are the same across roles. If anything the case for equal magnitudes is weaker than the case for a role-symmetric $L0$, because hard-wired payoff perturbations presumably stem from an evolutionary or learning process that is surely influenced by payoffs as well as framing. These a priori considerations are reinforced by our strong

rejection of the restriction to perturbations of equal magnitudes in RTH's dataset (Table 3) and by the qualitatively perverse behavior of the analogous QRE with restricted perturbations model (web appendix below).

The problem with equilibrium with unrestricted perturbations, as we see it, is that there is no theory that could explain (or even restrict the specification of) such differences in magnitudes. This makes “explaining” the role-asymmetry in RTH's subjects’ behavior (via an unexplained twofold role difference in magnitudes) an empirical dead end. Our overfitting and portability analyses make this subjective judgment more concrete, by showing that equilibrium with unrestricted perturbations, which has the best fit of all the models we consider in RTH's dataset (Table 3), does worst of all in our within-sample overfitting test (Table 4); does worse than equilibrium without perturbations under the natural restrictions on hiders’ and seekers’ instinctive reactions to salience in O’Neill’s game; and is not well-defined for Rapoport and Boebel's game.

Quantal Response Equilibrium (“QRE”) with and without Payoff Perturbations

Here we discuss the QRE with payoff perturbations models mentioned in Section II. In a QRE players’ choices are noisy, with the probability of each choice increasing in its expected payoff given the distribution of others’ choices; a QRE is thus a fixed point in the space of players’ choice distributions. QRE describes the patterns of deviations from equilibrium in some other experiments, and so has the potential to explain RTH’s results better than equilibrium does. Its specification is completed by a response distribution, whose noisiness is represented (inversely) by a precision parameter. Some of our results are independent of this distribution, but for others we adopt the standard assumption of logit responses and study the special case called “logit QRE”.

Because QRE responds only to the payoff structure, it ignores the framing of the Hide-and-Seek game without payoff perturbations. In that game, for any error distribution, there is a unique QRE, which yields the same choice probabilities as equilibrium. To see this, suppose to the contrary that in a QRE the most probable location for Hiders, call it P, has probability greater than 1/4. Because QRE choice probabilities increase with expected payoffs and the game is constant-sum, P must then have the highest expected payoff for Seekers, thus probability greater than 1/4 for them. But then some location other than P has higher expected payoff for Hiders, a contradiction.

We therefore consider explanations that combine logit QRE with payoff perturbations as in Section II’s equilibrium with perturbations analysis (Figure 2), which make QRE sensitive to the

framing and give it the potential to explain RTH's results by responding asymmetrically to the asymmetries in the perturbed game's payoff structure. As usual, such models must be solved computationally.

QRE with perturbations does no better than equilibrium with perturbations in explaining RTH's results. Figure A1 illustrates logit QRE with payoff perturbations restricted to be equal in magnitude but opposite in sign across player roles, as a function of the precision λ , with $e = 0.2187$ and $f = 0.2010$, the values that best fit RTH's data for the equilibrium with restricted perturbations model. (The maximum likelihood estimate of λ in the QRE with restricted perturbations model is effectively infinite, reducing the model to the analogous equilibrium model.) For all combinations of $e, f = 0.1, 0.2, 0.3, \text{ or } 0.4$ (all consistent with a totally mixed equilibrium), as in Figure A1, the logit QRE probability of central A dips below 0.25 for low values of λ for seekers but never for hiders; and it is always higher for hiders, reversing the patterns in RTH's data. There is enough structure to suggest that this result is symptomatic of a theorem, but we have been unable to prove it. Thus, logit QRE can explain the prevalence of central A for Hiders and Seekers with perturbations of equal magnitudes but opposite signs across player roles. But the main difficulty is explaining the greater prevalence of central A for seekers, and in this case logit QRE robustly predicts that central A is more prevalent for hiders.

Like equilibrium with payoff perturbations, logit QRE can only explain RTH's results by postulating large differences across player roles in the magnitudes of the perturbations e and f as well as their signs. But this again yields an effectively infinite estimate of λ , reducing logit QRE, in terms of its substantive implications, to Section II's equilibrium with unrestricted perturbations model. (With payoff perturbations restricted to have equal magnitudes across player roles, the estimated $\lambda \rightarrow \infty$. With perturbations allowed to differ in magnitude across roles, for any sufficiently large but finite λ QRE can adjust the perturbations to match the observed frequencies exactly. Thus λ and the perturbations are not identified, but all parameter values that maximize the likelihood are equivalent to those obtained when $\lambda \rightarrow \infty$. With finite λ the estimated perturbations for hiders (seekers) are higher (lower) than those estimated for equilibrium with perturbations.) Figure A2 illustrates logit QRE with $e_H = 0.2910, f_H = 0.2535$, and $e_S = f_S = 0.1539$, the values that give the best fit for this model.

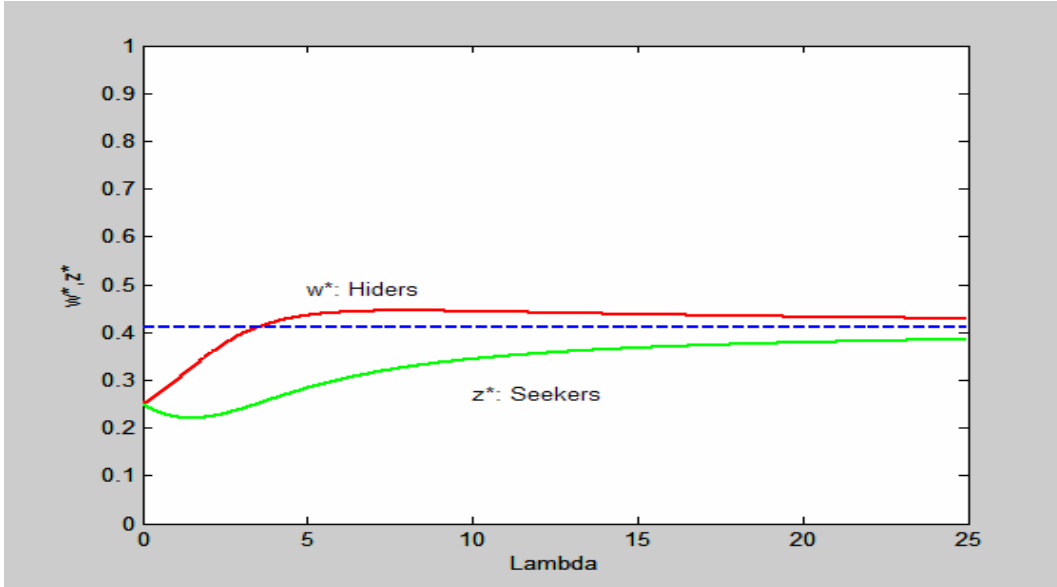


Figure A1. QRE with Payoff Perturbations of Equal Magnitudes Across Player Roles

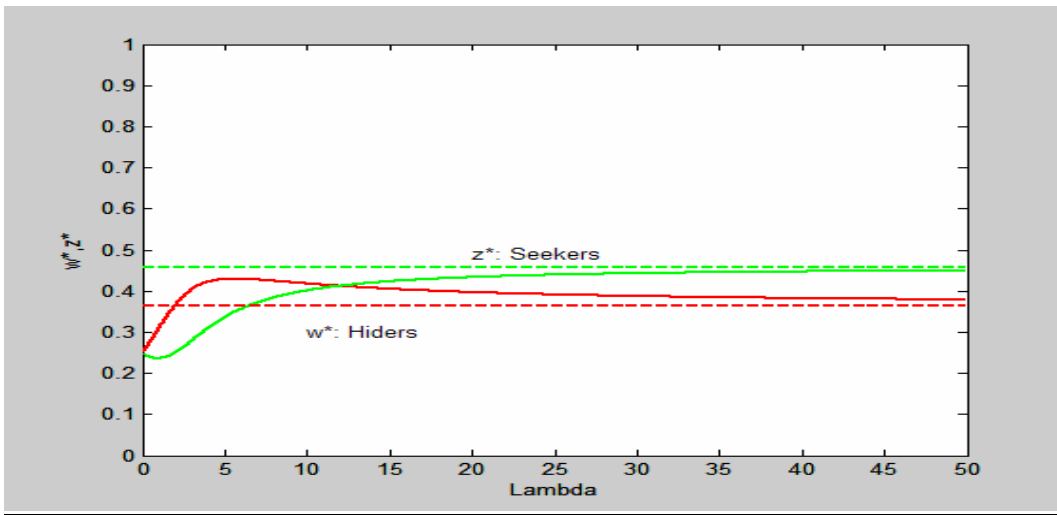


Figure A2. QRE with Payoff Perturbations of Differing Magnitudes Across Player Roles

Symmetry of $L0$ Across Roles for Level- k Models

Here we discuss, in more detail than in Section III, the issue of symmetry of $L0$ across player roles for level- k models.

In our view the case for a role-symmetric $L0$ is partly an empirical question and partly rests on behavioral plausibility. Assuming for the sake of argument that some kind of level- k model is correct, the issue is how best to describe a player’s strategic thinking with regard to the order in

which he processes information about the game. One can imagine a model in which a player first reacts, all at once, to the game's framing, feasible decisions, and payoffs, and so forms a payoff-sensitive, role-asymmetric (though perhaps nonstrategic) $L0$; and then reacts further to the payoffs and strategic structure via a best-responding $L1$, $L2$, etc. But one can equally well imagine a compartmentalization of the thought process in which a player reacts first to the framing and feasible decisions, which are arguably more primitive than payoffs even in games where the strategic structure is as transparent as in hide-and-seek; and to payoffs only in later stages.

The latter compartmentalization is more in keeping with the spirit of the nonstrategic $L0$ s in most of the previous level- k literature. See the uniform random $L0$ in Stahl and Wilson (1994, 1995) and several subsequent analyses; the *truthful* and *credulous* $L0$ s in Crawford (2003); and the *truthful* and *random* $L0$ s in Crawford and Iriberry (2005). The role-symmetric, non-uniform random $L0$ in Ho, Camerer, and Weigelt (1998) implicitly allows payoffs to matter by not imposing uniformity, but does so only for symmetric games where the distinction between nonstrategic and payoff-insensitive doesn't matter, and doesn't try to model payoff-sensitivity or connect $L0$ s across games. The one $L0$ in the literature that is clearly payoff-sensitive is Bacharach and Stahl's (1997ab), which favors salience for seekers and avoids it for hidiers.

A Bacharach and Stahl-style model fits slightly better than our favored model, and does 10% better within RTH's sample in our overfitting test (Table 4). But it does poorly beyond sample in O'Neill's game (Table 5), and it is not even well-defined for Rapoport and Boebel's game. The simplicity of a payoff-insensitive $L0$ and the fact that its specification can be based on decision-theoretic evidence about reactions to framing are (together with the simplicity and generality of the iterated best responses that define $L1$, etc.) the key to our level- k model's portability, because different games are much more likely to have compelling analogies between their sets of feasible decisions and framing than between strategic structures, as would be required to transport a Bacharach and Stahl-style $L0$.

Overfitting

This section gives more detail on the overfitting test discussed in Section V. Table A2 gives the treatment by treatment parameter estimates on which the overfitting comparisons in Table 4 are based. The estimated type frequencies for our proposed level- k model in Table A2, particularly those of $L4$, vary widely across RTH's six treatments, which is disturbing because level- k types are meant to be general strategic decision rules. The estimated type

frequencies for $L1$ and $L2$ are actually very stable. The type frequencies for $L3$ and $L4$ are less stable, but mainly for RTH-4 and RT-AABA-Mine. Even so, we cannot reject the constraint that in our proposed level- k model, the type frequencies are the same in all six RTH treatments, despite RTH's large samples (p -value 0.9873). This failure to reject despite the varying point estimates is due mainly to two factors: (i) The designs, with only one observation per subject in the games we study, are not well suited to identifying subjects' decision rules; and (ii) $L4$'s and to a lesser extent $L3$'s frequency estimates are weakly identified in RTH's data because $L4$ never chooses central A and $L3$ seldom does (Table 2), so they are not much involved in either of the major, robust patterns in the data.

As Table A2 shows, the parameter estimates for the other models also vary widely across treatments. Like the instability of the type frequencies for $L3$ and $L4$, this is probably due to differences in observed frequencies other than the larger ones involving central A that our analysis focuses on. The two treatments that differ the most from the others are RTH-4, in which B is chosen more than in any other treatment (Table 1, footnote 7) and RT-AABA-Mine, in which hiders' and seekers' frequencies of central A differ the least.

Table A2. Treatment by Treatment Parameter Estimates in RTH's Games

| Treatment | Level- k with symmetric $L0$ favoring salience | | | | | | Equilibrium with unrestricted perturbations | | | |
|-------------------------|--|--------|--------|--------|--------|-----|---|--------|--------|------------|
| | r | s | T | u | v | e | e_H | f_H | e_S | f_S |
| RTH-4 | 0 | 0.2499 | 0.2643 | 0.4858 | 0.0000 | 0 | 0.3307 | 0.1451 | 0.2736 | 0.0377 |
| RT-AABA-Treasure | 0 | 0.1577 | 0.3265 | 0.2257 | 0.2901 | 0 | 0.3648 | 0.2941 | 0.1164 | 0.1640 |
| RT-AABA-Mine | 0 | 0.1566 | 0.3393 | 0.0686 | 0.4355 | 0 | 0.1818 | 0.2121 | 0.1028 | 0.2192 |
| RT-1234-Treasure | 0 | 0.1572 | 0.3810 | 0.1421 | 0.3197 | 0 | 0.3035 | 0.2976 | 0.1471 | 0.1390 |
| RT-1234-Mine | 0 | 0.2066 | 0.3153 | 0.2603 | 0.2178 | 0 | 0.2669 | 0.2406 | 0.1667 | 0.1111 |
| R-ABAA | 0 | 0.1933 | 0.3743 | 0.2683 | 0.1641 | 0 | 0.4141 | 0.3594 | 0.2500 | 0.2600 |
| Treatment | Level- k with symmetric $L0$ avoiding salience | | | | | | Level- k with asymmetric $L0$ | | | |
| | r | s | T | u | v | e | r | s | t | ϵ |
| RTH-4 | 0 | 0.2897 | 0 | 0.4911 | 0.2192 | 0 | 0 | 0.7940 | 0.2060 | 0.7312 |
| RT-AABA-Treasure | 0 | 0.4184 | 0.0668 | 0.3265 | 0.1883 | 0 | 0 | 0.5408 | 0.4592 | 0.6588 |
| RT-AABA-Mine | 0 | 0.2176 | 0.4239 | 0 | 0.3585 | 0 | 0 | 0.8032 | 0.1968 | 0.8081 |
| RT-1234-Treasure | 0 | 0.3761 | 0.0822 | 0.3816 | 0.1601 | 0 | 0 | 0.6091 | 0.3909 | 0.6984 |
| RT-1234-Mine | 0 | 0.3797 | 0.0334 | 0.4745 | 0.1124 | 0 | 0 | 0.6804 | 0.3196 | 0.7419 |
| R-ABAA | 0 | 0.3925 | 0.0337 | 0.3326 | 0.2412 | 0 | 0 | 0.7300 | 0.2700 | 0.6042 |

Portability

This section gives background on Section VI's portability analysis. Figure A3 and Tables A3-A4 give the details of types' choices in the best-fitting regions for our proposed level- k model in O'Neill's game, just as Figure 3 and Table 2 did for RTH's games. Table A5 gives the details of types' choices in the best-fitting regions for our proposed model in Rapoport and Boebel's game.

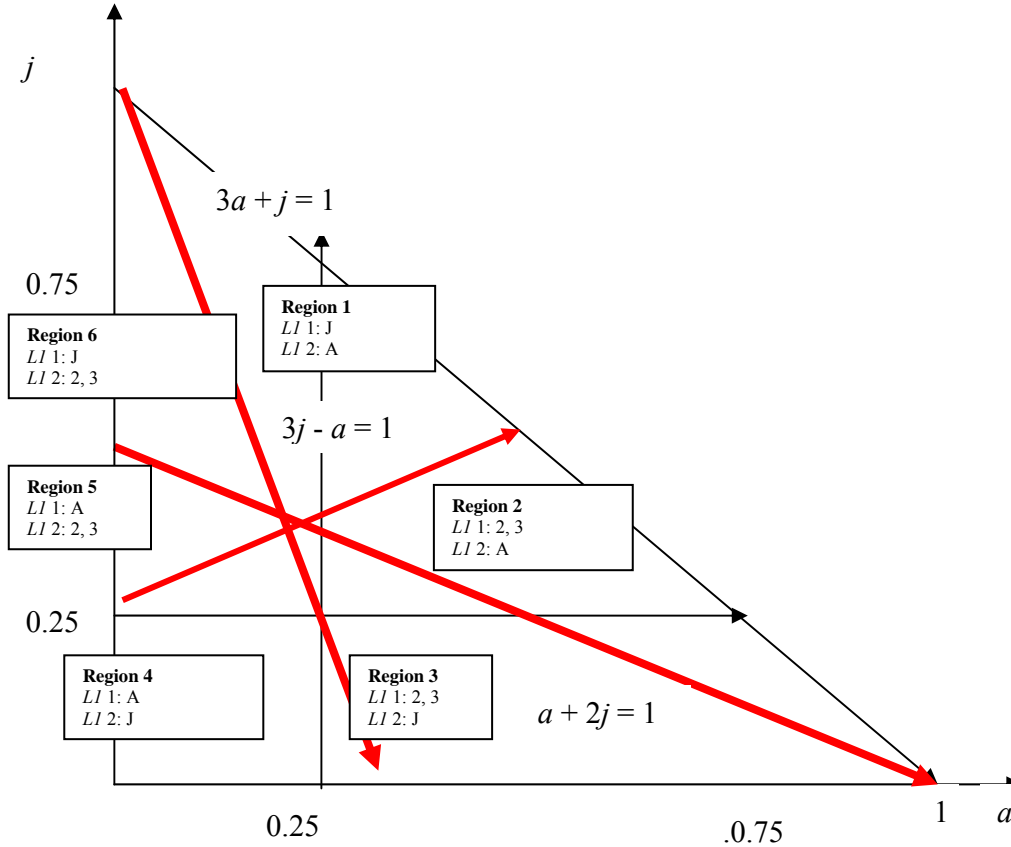


Figure A3. $L1$'s Choices in O'Neill's Game as Functions of $L0$'s Choice Probabilities a and j

Table A3. Types' Expected Payoffs and Choice Probabilities in O'Neill's Game when $3j - a < 1$

| Player 1 | Exp. Payoff $A+2j < 1$ | Choice Pr. $a+2j < 1$ | Exp. Payoff $a+2j > 1$ | Choice Pr. $a+2j > 1$ | Player 2 | Exp. Payoff $a+2j < 1$ | Choice Pr. $a+2j < 1$ | Exp. Payoff $a+2j > 1$ | Choice Pr. $a+2j > 1$ |
|-------------------|--|---------------------------------|--|---------------------------------|-------------------|---|---------------------------------|---|---------------------------------|
| L0 (Pr. R) | | | | | L0 (Pr. r) | | | | |
| A | - | a | - | A | A | - | a | - | a |
| 2 | - | $(1-a-j)/2$ | - | $(1-a-j)/2$ | 2 | - | $(1-a-j)/2$ | - | $(1-a-j)/2$ |
| 3 | - | $(1-a-j)/2$ | - | $(1-a-j)/2$ | 3 | - | $(1-a-j)/2$ | - | $(1-a-j)/2$ |
| J | - | j | - | J | J | - | j | - | j |
| L1 (Pr. s) | | | | | L1 (Pr. s) | | | | |
| A | $1-a-j$ | 0 | $1-a-j$ | 0 | A | $a+j$ | 0 | $a+j$ | 1 |
| 2 | $(1+a-j)/2$ | 1/2 | $(1+a-j)/2$ | 1/2 | 2 | $(1-a+j)/2$ | 0 | $(1-a+j)/2$ | 0 |
| 3 | $(1+a-j)/2$ | 1/2 | $(1+a-j)/2$ | 1/2 | 3 | $(1-a+j)/2$ | 0 | $(1-a+j)/2$ | 0 |
| J | J | 0 | J | 0 | J | $1-j$ | 1 | $1-j$ | 0 |
| L2 (Pr. t) | | | | | L2 (Pr. t) | | | | |
| A | 0 | 0 | 0 | 0 | A | 0 | 0 | 0 | 0 |
| 2 | 0 | 0 | 1 | 1/2 | 2 | 1/2 | 0 | 1/2 | 0 |
| 3 | 0 | 0 | 1 | 1/2 | 3 | 1/2 | 0 | 1/2 | 0 |
| J | 1 | 1 | 0 | 0 | J | 1 | 1 | 1 | 1 |
| L3 (Pr. u) | | | | | L3 (Pr. u) | | | | |
| A | 0 | 0 | 0 | 0 | A | 1 | 1/3 | 0 | 0 |
| 2 | 0 | 0 | 0 | 0 | 2 | 1 | 1/3 | 1/2 | 0 |
| 3 | 0 | 0 | 0 | 0 | 3 | 1 | 1/3 | 1/2 | 0 |
| J | 1 | 1 | 1 | 1 | J | 0 | 0 | 1 | 1 |
| L4 (Pr. v) | | | | | L4 (Pr. v) | | | | |
| A | 2/3 | 1/3 | 0 | 0 | A | 1 | 1/3 | 1 | 1/3 |
| 2 | 2/3 | 1/3 | 0 | 0 | 2 | 1 | 1/3 | 1 | 1/3 |
| 3 | 2/3 | 1/3 | 0 | 0 | 3 | 1 | 1/3 | 1 | 1/3 |
| J | 0 | 0 | 1 | 1 | J | 0 | 0 | 0 | 0 |
| Total | $a+2j < 1$ | | $a+2j > 1$ | | Total | $a+2j < 1$ | | $a+2j > 1$ | |
| A | $ra+(1-\varepsilon)[v/3] + (1-r) \varepsilon/4$ | | $ra+ (1-r) \varepsilon/4$ | | A | $ra+(1-\varepsilon) [u/3+v/3]+ (1-r) \varepsilon/4$ | | $ra+(1-\varepsilon) [s+v/3]+ (1-r) \varepsilon/4$ | |
| 2 | $r(1-a-j)/2+ (1-\varepsilon) [s/2+v/3]+ (1-r) \varepsilon/4$ | | $r(1-a-j)/2+ (1-\varepsilon) [s/2+t/2]+ (1-r) \varepsilon/4$ | | 2 | $r(1-a-j)/2+(1-\varepsilon) [u/3+v/3]+ (1-r) \varepsilon/4$ | | $r(1-a-j)/2+(1-\varepsilon) [v/3]+ (1-r) \varepsilon/4$ | |
| 3 | $r(1-a-j)/2+(1-\varepsilon) [s/3+v/3]+ (1-r) \varepsilon/4$ | | $r(1-a-j)/2+ (1-\varepsilon) [s/2+t/2]+ (1-r) \varepsilon/4$ | | 3 | $r(1-a-j)/2+(1-\varepsilon) [u/3+v/3]+ (1-r) \varepsilon/4$ | | $r(1-a-j)/2+(1-\varepsilon) [v/3]+ (1-r) \varepsilon/4$ | |
| J | $Rj+(1-\varepsilon) [t+u]+ (1-r) \varepsilon/4$ | | $rj+(1-\varepsilon) [u+v]+ (1-r) \varepsilon/4$ | | J | $rj+(1-\varepsilon) [s+t]+ (1-r) \varepsilon/4$ | | $rj+(1-\varepsilon) [t+u]+ (1-r) \varepsilon/4$ | |

Table A4. Types' Expected Payoffs and Choice Probabilities in O'Neill's Game when $3j - a > 1$

| Player 1 | Exp. Payoff | Choice Pr. | Player 2 | Exp. Payoff | Choice Pr. |
|--------------------------|--|-------------|--------------------------|--|-------------|
| <i>L0 (Pr. R)</i> | | | <i>L0 (Pr. r)</i> | | |
| A | - | a | A | - | a |
| 2 | - | $(1-a-j)/2$ | 2 | - | $(1-a-j)/2$ |
| 3 | - | $(1-a-j)/2$ | 3 | - | $(1-a-j)/2$ |
| J | - | j | J | - | j |
| <i>L1 (Pr. S)</i> | | | <i>L1 (Pr. s)</i> | | |
| A | $1-a-j$ | 0 | A | $a+j$ | 1 |
| 2 | $(1+a-j)/2$ | 0 | 2 | $(1-a+j)/2$ | 0 |
| 3 | $(1+a-j)/2$ | 0 | 3 | $(1-a+j)/2$ | 0 |
| J | j | 1 | J | $1-j$ | 0 |
| <i>L2 (Pr. T)</i> | | | <i>L2 (Pr. t)</i> | | |
| A | 0 | 0 | A | 1 | 1/3 |
| 2 | 1 | 1/2 | 2 | 1 | 1/3 |
| 3 | 1 | 1/2 | 3 | 1 | 1/3 |
| J | 0 | 0 | J | 0 | 0 |
| <i>L3 (Pr. U)</i> | | | <i>L3 (Pr. u)</i> | | |
| A | 2/3 | 1/3 | A | 0 | 0 |
| 2 | 2/3 | 1/3 | 2 | 1/2 | 0 |
| 3 | 2/3 | 1/3 | 3 | 1/2 | 0 |
| J | 0 | 0 | J | 1 | 1 |
| <i>L4 (Pr. V)</i> | | | <i>L4 (Pr. v)</i> | | |
| A | 0 | 0 | A | 1/3 | 0 |
| 2 | 0 | 0 | 2 | 1/3 | 0 |
| 3 | 0 | 0 | 3 | 1/3 | 0 |
| J | 1 | 1 | J | 1 | 1 |
| Total | | | Total | | |
| A | $Ra+(1-\varepsilon)[u/3]+(1-r)\varepsilon/4$ | | A | $ra+(1-\varepsilon)[s+t/3]+(1-r)\varepsilon/4$ | |
| 2 | $r(1-a-j)/2+(1-\varepsilon)[t/2+u/3]+(1-r)\varepsilon/4$ | | 2 | $r(1-a-j)/2+(1-\varepsilon)[t/3]+(1-r)\varepsilon/4$ | |
| 3 | $R(1-a-j)/2+(1-\varepsilon)[t/2+u/3]+(1-r)\varepsilon/4$ | | 3 | $r(1-a-j)/2+(1-\varepsilon)[t/3]+(1-r)\varepsilon/4$ | |
| J | $Rj+(1-\varepsilon)[s+v]+(1-r)\varepsilon/4$ | | J | $rj+(1-\varepsilon)[u+v]+(1-r)\varepsilon/4$ | |

Table A5. Types' Expected Payoffs and Choice Probabilities in Rapoport and Boebel's Game

| Player 1 | Exp. Payoff | Choice Pr. | Exp. Payoff | Choice Pr. | Player 2 | Exp. Payoff | Choice Pr. | Exp. Payoff | Choice Pr. |
|-------------------|--|-------------|--|-------------|-------------------|--|-------------|--|-------------|
| | $3m/2+n>1$ | $3m/2+n>1$ | $3m/2+n<1$ | $3m/2+n<1$ | | $3m/2+n>1$ | $3m/2+n>1$ | $3m/2+n<1$ | $3m/2+n<1$ |
| L0 (Pr. r) | | | | | L0 (Pr. r) | | | | |
| C | - | $m/2$ | - | $m/2$ | C | - | $m/2$ | - | $m/2$ |
| L | - | $(1-m-n)/2$ | - | $(1-m-n)/2$ | L | - | $(1-m-n)/2$ | - | $(1-m-n)/2$ |
| F | - | n | - | n | F | - | N | - | n |
| I | - | $(1-m-n)/2$ | - | $(1-m-n)/2$ | I | - | $(1-m-n)/2$ | - | $(1-m-n)/2$ |
| O | | $m/2$ | | $m/2$ | O | | $m/2$ | | $m/2$ |
| L1 (Pr. s) | | | | | L1 (Pr. s) | | | | |
| C | $m/2$ | 0 | $m/2$ | 0 | C | $1-m/2$ | 0 | $1-m/2$ | 1 |
| L | $1/2+n/2$ | 1 | $1/2+n/2$ | 1 | L | $1/2-n/2$ | 0 | $1/2-n/2$ | 0 |
| F | $1/2-n/2$ | 0 | $1/2-n/2$ | 0 | F | $1/2+n/2$ | 0 | $1/2+n/2$ | 0 |
| I | $1-m-n$ | 0 | $1-m-n$ | 0 | I | $m+n$ | 1 | $m+n$ | 0 |
| O | $(1-m+n)/2$ | 0 | $(1-m+n)/2$ | 0 | O | $(1+m-n)/2$ | 0 | $(1+m-n)/2$ | 0 |
| L2 (Pr. t) | | | | | L2 (Pr. t) | | | | |
| C | 0 | 0 | 1 | 1 | C | 1 | $1/2$ | 1 | $1/2$ |
| L | 1 | $1/2$ | 0 | 0 | L | 1 | $1/2$ | 1 | $1/2$ |
| F | 0 | 0 | 0 | 0 | F | 0 | 0 | 0 | 0 |
| I | 1 | $1/2$ | 0 | 0 | I | 0 | 0 | 0 | 0 |
| O | 0 | 0 | 0 | 0 | O | 0 | 0 | 0 | 0 |
| L3 (Pr. u) | | | | | L3 (Pr. u) | | | | |
| C | $1/2$ | $1/4$ | $1/2$ | $1/4$ | C | 1 | 1 | 0 | 0 |
| L | 0 | 0 | 0 | 0 | L | $1/2$ | 0 | 1 | $1/4$ |
| F | $1/2$ | $1/4$ | $1/2$ | $1/4$ | F | $1/2$ | 0 | 1 | $1/4$ |
| I | $1/2$ | $1/4$ | $1/2$ | $1/4$ | I | 0 | 0 | 1 | $1/4$ |
| O | $1/2$ | $1/4$ | $1/2$ | $1/4$ | O | $1/2$ | 0 | 1 | $1/4$ |
| L4 (Pr. v) | | | | | L4 (Pr. v) | | | | |
| C | 1 | 1 | 0 | 0 | C | $3/4$ | $1/4$ | $3/4$ | $1/4$ |
| L | 0 | 0 | $3/4$ | 1 | L | 0 | 0 | $1/4$ | 0 |
| F | 0 | 0 | $1/2$ | 0 | F | $3/4$ | $1/4$ | $3/4$ | $1/4$ |
| I | 0 | 0 | $1/2$ | 0 | I | $3/4$ | $1/4$ | $3/4$ | $1/4$ |
| O | 0 | 0 | $1/2$ | 0 | O | $3/4$ | $1/4$ | $3/4$ | $1/4$ |
| Total | $3m/2+n>1$ | | $3m/2+n<1$ | | Total | $3m/2+n>1$ | | $3m/2+n<1$ | |
| C | $rm/2+(1-\varepsilon)[u/4+v]+(1-r)\varepsilon/5$ | | $rm/2+(1-\varepsilon)[t+u/4]+(1-r)\varepsilon/5$ | | C | $rm/2+(1-\varepsilon)[t/2+u+v/4]+(1-r)\varepsilon/5$ | | $rm/2+(1-\varepsilon)[s+t/2+v/4]+(1-r)\varepsilon/5$ | |
| L | $r(1-m-n)/2+(1-\varepsilon)[s+t/2]+(1-r)\varepsilon/5$ | | $r(1-m-n)/2+(1-\varepsilon)[s+v]+(1-r)\varepsilon/5$ | | L | $r(1-m-n)/2+(1-\varepsilon)[t/2]+(1-r)\varepsilon/5$ | | $r(1-m-n)/2+(1-\varepsilon)[t/2+u/4]+(1-r)\varepsilon/5$ | |
| F | $rn+(1-\varepsilon)[u/4]+(1-r)\varepsilon/5$ | | $rn+(1-\varepsilon)[u/4]+(1-r)\varepsilon/5$ | | F | $rn+(1-\varepsilon)[v/4]+(1-r)\varepsilon/5$ | | $rn+(1-\varepsilon)[u/4+v/4]+(1-r)\varepsilon/5$ | |
| I | $r(1-m-n)/2+(1-\varepsilon)[t/2+u/4]+(1-r)\varepsilon/5$ | | $r(1-m-n)/2+(1-\varepsilon)[u/4]+(1-r)\varepsilon/5$ | | I | $r(1-m-n)/2+(1-\varepsilon)[s+v/4]+(1-r)\varepsilon/5$ | | $r(1-m-n)/2+(1-\varepsilon)[u/4+v/4]+(1-r)\varepsilon/5$ | |
| O | $rm/2+(1-\varepsilon)[u/4]+(1-r)\varepsilon/5$ | | $rm/2+(1-\varepsilon)[u/4]+(1-r)\varepsilon/5$ | | O | $rm/2+(1-\varepsilon)[v/4]+(1-r)\varepsilon/5$ | | $rm/2+(1-\varepsilon)[u/4+v/4]+(1-r)\varepsilon/5$ | |