Level-k Thinking in Games with Non-neutral Frames

Vincent P. Crawford

17 September 2018

Abstract: Traditional game theory assumes that if framing does not affect a game’s payoffs, it cannot influence behavior. However, Rubinstein and Tversky (1993), Rubinstein, Tversky, and Heller (1996), and Rubinstein (1999) (henceforth “RTH”) reported experiments eliciting initial responses to hide-and-seek and other types of game, in which subjects’ behavior responded systematically to non-neutral framing via decision labelings. Crawford and Iriberri (2007ab) (“CI”) proposed a level-k explanation of RTH’s results for hide-and-seek games. Heap, Rojo-Arjona, and Sugden’s (2014) (“HRS”) criticized CI’s model on grounds of portability. This paper clarifies HRS’s interpretation of their results and responds to their criticism, suggesting a way forward.

1 Department of Economics, University of Oxford, Manor Road Building, Oxford OX1 3UQ and All Souls College, Oxford OX1 4AL, UK, e-mail vincent.crawford@economics.ox.ac.uk; and Department of Economics, University of California, San Diego, 9500 Gilman Drive, La Jolla, California, 92093-0508, USA. This paper supersedes my paper “A Comment on ‘How Portable is Level-0 Behavior? A Test of Level-k Theory in Games with Non-neutral Frames’ by Heap, Rojo-Arjona, and Sugden”. I thank Miguel Costa-Gomes and Nagore Iriberri for helpful advice. The research leading to these results has received funding from the European Research Council under the European Union’s Seventh Framework Programme (FP7/2007-2013) / ERC grant agreement no. 339179. The contents reflect only the author’s views and not the views of the ERC or the European Commission, and the European Union is not liable for any use that may be made of the information contained therein. All Souls College and the University of California, San Diego also provided support.
Keywords: behavioral game theory; experimental game theory; strategic thinking; level-k models; coordination; salience

Declarations of interest: None
1. Introduction

Traditional noncooperative game theory assumes that if the framing of a game does not affect players’ payoffs, it cannot influence their behavior. Even so, it has been known since Schelling (1960) that framing, via labels for players’ actions, player roles, or the game itself, can significantly affect people’s behavior via the patterns of salience it creates, even if the labels are abstract and free of connotations that might affect payoffs.

RTH’s (1993, 1996, 1999) experiments provide some of the most powerful evidence on framing effects in games. RTH elicited subjects’ initial responses to two-person hide-and-seek, discoordination, and pure coordination games, with non-neutral framing via the labelings of actions. Their designs included some treatments with abstract decision labels and others whose labels had positive or negative connotations. In each case subjects deviated systematically from equilibrium in ways that were sensitive to the labeling, and which were inconsistent not only with equilibrium but also with other leading models.

RTH’s results for their hide-and-seek games with abstract decision labels were particularly informative. Those subjects deviated systematically from equilibrium, with hiders and seekers both tending to favor the action whose position and labeling RTH argued made it “least salient”, and with seekers favoring the least salient action even more than hiders. HRS (2014) call this the “fatal attraction” pattern, after CI’s title. Such systematic deviations from equilibrium are surprising in a game where the equilibrium strategy is so obvious and its rationale so strong. They are doubly surprising because equilibrium and its noisy generalizations such as quantal response equilibrium
(McKelvey and Palfrey 1995; “QRE”) preclude such role-asymmetric patterns of deviation in RTH’s hide-and-seek games, despite their role-asymmetric payoff structures.

CI (2007ab) proposed a level-$k$ model to explain the fatal attraction pattern and other aspects of RTH’s results for hide-and-seek games. In a level-$k$ model, players anchor their beliefs in a nonstrategic initial assessment of others’ likely responses to the game, called “$L_0$”, and then adjust them via iterated best responses, with $L_1$ best responding to $L_0$, $L_2$ to $L_1$, and so on. Players’ levels are heterogeneous, drawn from a distribution concentrated on the lowest levels but excluding $L_0$, which evidence suggests exists only in the minds of higher levels (Crawford, Costa-Gomes, and Iriberri, 2013, Section 2.4).

Even when $L_0$’s frequency is zero, its specification is crucial. Although $L_0$ is usually taken to be uniform random, that assumption is behaviorally implausible when the labeling creates non-neutral patterns of salience. And in RTH’s hide-and-seek games, where the mixed-strategy equilibrium is also uniform random, such an $L_0$ would make $L_k$ (with neutral decision errors) for higher levels mimic equilibrium decisions. CI instead allowed $L_0$ to probabilistically favor actions that the labeling makes salient. That allows the level-$k$ model to capture the effects of how the game is framed, with higher levels responding to salience indirectly, through iterated best responses. To avoid begging the question of hiders’ and seekers’ role-asymmetric behavior, CI constrained $L_0$ and the frequencies of higher levels to be equal in both player roles.

When estimated using RTH’s data, CI’s level-$k$ model explains the fatal attraction pattern, via interactions between hide-and-seek’s role-asymmetric payoff structure and players’ iterated best responses to an $L_0$ that favors salience. CI’s model also tracks
RTH’s other results for hide-and-seek games, and ports well to some games related to hide-and-seek. As CI noted, the fatal attraction pattern is a crucial test, in that other leading models, such as equilibrium and QRE, which lack an iterated-best-response structure, cannot explain role-asymmetric patterns like RTH’s hide-and-seek results.²

HRS (2014) report new experiments, designed to assess the portability of CI’s level-k model across games like RTH’s hide-and-seek, discoordination, and pure coordination games. HRS’s test turns on assumptions about how a level-k model and its $L_0$ should port across the different types of game. Taken literally, HRS argue that because $L_0$ is often motivated as “a strategically naïve initial assessment of others’ likely responses to the game” (Crawford, Costa-Gomes, and Iriberri, 2013, p. 14), but the differences across hide-and-seek, discoordination, and pure coordination games are strategic, a level-k model should hold equally across their hide-and-seek, coordination, and discoordination games, with $L_0$ constant across the games within each of HRS’s groups of games with analogous decisions and labelings. HRS then show that no level-k model with a constant $L_0$ can explain their results across all three kinds of game. They also criticize some other aspects of CI’s analysis, even questioning the existence of the fatal attraction pattern.

This paper responds to HRS’s criticisms. I argue first that the central assumption of their test, that a level-k model should port across hide-and-seek, discoordination, and

² CI (2007a, p. 1738) discuss in detail the models’ mechanism for explaining the fatal attraction pattern, which depends on the heterogeneity of levels’ decisions. Camerer, Ho, and Chong’s (2004) cognitive hierarchy model, in which iterated best responses are also central, but which differs from the level-k model in details that are not relevant to this topic, can also explain the fatal attraction pattern. HRS, like CI and this paper, focus on level-k specifications.
pure coordination games, is neither theoretically justified nor behaviorally credible. CI’s and Crawford, Costa-Gomes, and Iriberri’s (2013) motivations of L0 as strategically naïve meant that L0 doesn’t model the details of others’ responses to incentives, not that players cannot distinguish pure coordination games from discoordination or hide-and-seek games. As intuition suggests, even strategically naïve people can distinguish pure coordination from other kinds of games. And there is ample evidence that in pure coordination games subjects may employ “team reasoning”, whereby “each player chooses the decision rule which, if used by all players, would be optimal for each of them” (Bardsley, Mehta, Starmer, and Sugden, 2009, p. 40). Team reasoning involves fixed-point reasoning, which is different in kind from level-k thinking. It is a behavioral non-starter to expect a level-k model to track subjects’ behavior across hide-and-seek, discoordination, and pure coordination games. Rejecting such portability says nothing we didn’t already know about behavior in games with non-neutral frames.

If not a level-k model, what might a general model of strategic responses to framing look like? I argue that the most promising route is not via the blanket rejection of level-k models that HRS seem to advocate (p. 1135), but via constructive experiments designed to identify a robust hybrid of team reasoning, level-k thinking, and perhaps other kinds of thinking; and to map the boundaries between settings that evoke each kind of thinking.

The rest of the paper is organized as follows. Section 2 summarizes RTH’s results for hide-and-seek games. Section 3 reviews CI’s level-k analysis of RTH’s results. Section 4 summarizes HRS’s results on the portability of level-k models like CI’s. Section 5 responds to HRS’s other critiques of CI’s analysis. Section 6 is the conclusion.
2. RTH’s Results for Hide-and-Seek Games

Recall that RTH’s subjects played two-person hide-and-seek, discoordination, and pure coordination games. I focus here on RTH’s results for hide-and-seek games, whose zero-sum structures made their subjects’ systematic deviations from equilibrium especially informative. RTH’s action labels were publicly known and had non-neutral patterns, e.g. A-B-A-A for each player’s four actions, making some actions more and others less salient, in one of Schelling’s (1960) senses. In some treatments (CI, 2007a, Table 1) the action labels were abstract, like A-B-A-A, arguably without connotations that might directly influence payoffs; while in others (CI, 2007b, Table A1) they had connotations, like Frown-Smile-Frown-Frown. Section 5 below gives more detail.

RTH’s subjects deviated from equilibrium in systematic patterns that were closely related to their action labelings. Both hiders and seekers tended to favor the action whose position and labeling made it least salient (as RTH argued). Seekers favored the least salient action even more than hiders did, and so had substantially higher average payoffs than in the game’s unique mixed-strategy equilibrium. This fatal attraction pattern, properly interpreted (Section 5), persisted across RTH’s six different hide-and-seek games.

---

3 In RTH’s pure coordination games, as in HRS’s, players who choose actions with the same label receive payoff one; otherwise both receive zero. In their discoordination games players who choose actions with different labels receive payoff one; otherwise both receive zero. In their hide-and-seek games, if players choose actions with the same label, the seeker receives payoff one and the hider receives zero; and vice versa if their actions have different labels.

4 In hide-and-seek games, any pure or mixed strategy is a best response to the equilibrium beliefs. But systematic deviations of aggregate choice frequencies from equilibrium mixed-strategy probabilities must with high probability have a cause that is partly common across players, and so are indicative of systematic deviations from equilibrium.
seek treatments with abstract decision labels (CI, 2007a, Section I) and less clearly in RTH’s five treatments with labels with connotations (CI, 2007b, Table A1).

As already mentioned, the fatal attraction pattern is surprising because RTH’s hide-and-seek games are zero-sum games in which the equilibrium strategies are obvious and their rationale is strong; and those games have a special payoff structure that, despite its role-asymmetry between hiders and seekers, makes equilibrium and its noisy generalizations such as quantal response equilibrium (McKelvey and Palfrey 1995; “QRE”) predict role-symmetric responses. In RTH’s hide-and-seek games, QRE action probabilities exactly coincide with the mixed-strategy equilibrium action probabilities for any well-behaved symmetric noise distributions, including logit (CI, 2007b).

RTH attributed the fatal attraction pattern to subjects’ strategic naiveté. In Rubinstein and Tversky’s (1993) words, “The finding that both choosers and guessers selected the least salient alternative suggests little or no strategic thinking.” In Rubinstein, Tversky, and Heller’s (1996) 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.” CI’s analysis was motivated by the belief that systematic, robust deviations from equilibrium in games as plainly strategic as hide-and-seek are unlikely to have a nonstrategic explanation.

3. CI’s Level-k Analysis of RTH’s Results

CI (2007ab) proposed a strategic but nonequilibrium level-k model to explain RTH’s results for hide-and-seek games. Recall that in CI’s level-k model, players anchor their
beliefs in an $L_0$ and then adjust them via iterated best responses, with $L_1$ best responding to $L_0$, $L_2$ to $L_1$, and so on. Players’ levels are heterogeneous, drawn from a distribution concentrated on the lowest levels, but excluding $L_0$ (Crawford, Costa-Gomes, and Iriberri, 2013). $L_0$ is allowed to probabilistically favor salient actions, and its action probabilities and the level frequencies are constrained to be equal for both player roles.

In estimating the model, CI used RTH’s data from their six treatments with abstract decision labels, omitting RTH’s treatments where the labels had connotations as likely to induce uncontrolled variation in payoffs (CI 2007ab). However, the patterns in the data for treatments where labels had connotations suggest that estimates for those treatments would have been similar, though less clear. Section 5 gives more detail.

CI (2007a, pp. 1738-40 and Table 3) estimated $L_0$ and the level frequencies under alternative constraints on whether and how $L_0$ favors salience for hiders and/or seekers.\(^5\) For theoretical reasons, CI proposed a model in which $L_0$ favors salience for both hiders and seekers. That model reproduces the fatal attraction pattern, in a way that depends on the heterogeneity of levels (CI, 2007a, pp. 1738-9). Two alternative level-$k$ models, which don’t favor salience for both hiders and seekers, and also reproduce the fatal attraction pattern and fit slightly better; but they perform poorly in CI’s overfitting tests.

---

\(^5\) One issue is whether players treat the “end-A” actions in the A-B-A-A frame, or its analogs in other frames, as more or less salient than the “central-A” action. CI’s estimates favor the salience of the “end-A” actions. HRS (pp. 1133-1134) imply that this aspect of CI’s estimation makes their model prone to overfit. However, all that matters about the continuously variable probabilities that define $L_0$ in CI’s model is the best-response $L_1$ actions it induces for hiders and seekers. As a result, this aspect of CI’s estimation adds only a single binary-valued parameter: the smallest possible non-zero price to pay for explaining a non-zero response to salience.
The fatal attraction pattern is a crucial test of level-$k$ models: No other model with role-symmetric behavioral assumptions I am aware of generates its role-asymmetric behavior.

CI (2007a, pp. 1736-37) also considered equilibrium-based models of players’ responses to salience, with “hard-wired” payoff perturbations that reflect a plausible instinctive attraction to salience for seekers and an aversion to salience for hiders.\(^6\) CI showed that such models can track the fatal attraction pattern, but only with perturbations twice as large for hiders as for seekers, with the difference unexplained. An equilibrium-with-unrestricted-payoff-perturbations model that is flexible enough to fit almost perfectly, fits best; but it too performs poorly in CI’s tests for overfitting. CI (2007b, pp. 3-5) also showed that logit QRE versions of the models with perturbations do no better, and generally worse, than their equilibrium-with-perturbations counterparts. Such models even predict that hiders favor the least salient action more than seekers, the opposite of the fatal attraction pattern in the data. Maximum likelihood estimates of their noise parameters are 0, thus reducing them to equilibrium-with-perturbations models.

CI, aware that their estimation gives their level-$k$ model more freedom than equilibrium or QRE, tested the model’s portability by adapting it to the very different structures and salience patterns of O’Neill’s (1987) and Rapoport and Boebel’s (1992) hide-and-seek-like games, using the level frequencies estimated from RTH’s data and

\(^6\) This role asymmetry superficially seems to differ from the assumption in CI’s proposed level-$k$ model that $L0$ favors salience for both hiders and seekers. Note however that $L0$ is meant to describe a strategically naive initial reaction, while the perturbations in CI’s equilibrium-with-perturbations models are meant to directly determine strategic behavior. An equilibrium-with-perturbations model in which hiders were attracted to salience would fit very poorly.
defining $L_0$ via the same principles as for RTH’s games.\(^7\) The adapted model describes O’Neill’s subjects’ early responses well, and Rapoport and Boebel’s subjects’ fairly well.

Although CI did not consider RTH’s treatments in which the labels had positive or negative connotations, payoff perturbations, as in CI’s equilibrium-based analyses, could be used to model their effects directly, either in the level-$k$ or the equilibrium version of the model. I do not attempt that here, but the data (CI 2007b, Table A1) suggest that the results would not be very different from the results with abstract action labels.

4. HRS’s Analysis

HRS (2014) reported new experiments on hide-and-seek, discoordination, and pure coordination games, designed to test the portability of level-$k$ models like CI’s across games with different structures. They argue that a useful model must be well-defined in advance for any game,\(^8\) and must be evaluated via ex ante hypothesis testing, not ex post model fitting (HRS, p. 1135). I agree that ex ante hypothesis testing is preferable for model evaluation, other things equal. But it is seldom the most efficient way to discover and interpret new facts about behavior. For that and other reasons, ex post model fitting has had a long and useful tradition in experimental (as well as empirical) economics.

HRS’s subjects played series of hide-and-seek, coordination, and discoordination games, with structures close to those of RTH’s games, but different labelings, which

\(^7\) Equilibrium-with-perturbations or logit QRE-with-perturbations do no better in describing O’Neill’s (1987) and Rapoport and Boebel’s (1992) results than in describing RTH’s results (CI 2007a, pp. 1743-8).

\(^8\) Yet to my knowledge, team reasoning (Bardsley, Mehta, Starmer, and Sugden, 2009) has been clearly defined only for pure coordination games.
however created similar patterns of salience. Unlike RTH’s subjects, HRS’s played the games within subjects across games with different structures, but with the numbers and labelings of actions constant within groups and across player roles. Some of HRS’s treatments had abstract action labels and others had labels with connotations.

HRS use their results to test a level-$k$ model in the spirit of CI’s, focusing mainly on their treatments with abstract action labels. They followed CI in assuming that $L0$ has zero frequency, that higher levels respond to the labeling of actions only through their iterated best responses to a salience-sensitive $L0$, and in constraining $L0$ and the level frequencies to be constant across player roles in a given game.

Crucially, HRS argue that because $L0$ is motivated as strategically naïve (Crawford, Costa-Gomes, and Iriberri 2013, p. 14), while the differences across games within their groups of games are strategic, a level-$k$ model should hold across their hide-and-seek, coordination, and discoordination games, with $L0$ constant across the games within each of HRS’s groups with analogous decisions and labelings, and the level frequencies constant across games. (HRS (p. 1135) do acknowledge that a level-$k$ model with a constant $L0$ is not the only possibility: “Of course, we cannot claim that the portability property we test is implied by every possible general hypothesis about $L0$ behavior.” But they make no attempt to discuss non-level-$k$ alternatives; nor do they acknowledge that the empirically most promising model for pure coordination games is not even a level-$k$ model with a non-constant $L0$, but rather team reasoning.)

HRS’s constant-$L0$ level-$k$ assumption implies cross-game restrictions on behavior, which are the heart of their test. In each of their groups of games, the action labelings
make one label, which they call the “oddity”, uniquely different from the others. In their pure coordination games the oddity is subjects’ majority choice for almost all groups and labelings. Those games are symmetric across player roles, and in a level- \( k \) model the oddity must then be \( L0 \)’s modal choice. With \( L0 \) constant across games and player roles, the oddity’s average frequency across discoordinators, hiders, and seekers must be disproportionately high (HRS, p. 1138, Implication 1). However, HRS’s subjects chose the oddity far less frequently in discoordination and hide-and-seek games than in pure coordination games, and its frequency is usually too low to be consistent with their constant- \( L0 \) level- \( k \) model (HRS, pp. 1145-6, Table III). HRS conclude (p. 1135) that “…it would seem hard to be optimistic about finding…a general theory of \( L0 \) behavior”, and that level- \( k \) models lack the portability needed to be useful.

Experiments that test models of strategic behavior across games with wider ranges of strategic structures are welcome.\(^9\) However, neither theory nor behavioral plausibility supports HRS’s cross-game restrictions. CI’s and Crawford, Costa-Gomes, and Iriberri’s (2013) motivations of \( L0 \) as strategically naïve meant that \( L0 \) doesn’t model the details of others’ responses to incentives, not that players cannot distinguish pure coordination from discoordination or hide-and-seek games at all.

\(^9\) There are at least two other notable tests of the portability of level- \( k \) models. Georganas, Healy, and Weber (2015) study “undercutting” games and Costa-Gomes and Crawford’s (2006) two-person guessing games, with neutral framing. Penczynski (2016) studies the consistency of subjects’ levels across the hider’s and the seeker’s role in hide-and-seek games framed like RTH’s, in a design whose player roles are filled by two-person teams, whose communications are monitored to gain further insight into their thinking. Both find only limited portability. Crawford, Costa-Gomes, and Iriberri (2013, Section 3.5) discuss them in more detail.
It is well known that pure coordination games often trigger “team reasoning”, whereby “each player chooses the decision rule which, if used by all players, would be optimal for each of them” (Bardsley, Mehta, Starmer, and Sugden 2009, p. 40; Crawford, Gneezy, and Rottenstreich 2008, p. 1448). Team reasoning involves fixed-point reasoning, and is inherently different from level-k thinking. Asking a level-k model (with any specification of $L0$, constant or not) to track subjects’ behavior across hide-and-seek, discoordination, and coordination games is empirically a non-starter, based on a questionable interpretation of “strategically naïve”, essentially a play on words. The failure of such a model to pass HRS’s test tells us little that was not already known about how to model behavior in games with non-neutral action labelings.

If a level-k model is not suitable as a general model of strategic responses to framing, what might a suitable model look like? In my view evidence like RTH’s fatal attraction pattern, which is surprisingly robust across games whose structures resemble hide-and-seek’s, shows that a suitable model must accommodate something with the iterated-best-response structure of level-k thinking for games with a substantial range of structures. This, coupled with the prevalence of team reasoning in pure coordination games, suggests that the most promising route to a general model is not via the all-or-nothing rejection of level-k models that HRS seem to advocate (p. 1135). Rather, it is via constructive experiments with the goal of identifying a robust hybrid of team reasoning, level-k thinking, and perhaps other kinds of thinking; and of mapping the boundaries between the kinds of settings that evoke each kind of thinking. As Crawford, Gneezy, and Rottenstreich (2008, p. 1448) said of their experimental results for (pure and impure)
coordination games, “Overall, our results suggest a synthesis of level-\(k\) thinking and team reasoning in which team reasoning supplements or supplants level-\(k\) thinking in some settings.” Following HRS’s dictum that a useful model must be well-defined in advance for any game, this will require either a more general, empirically grounded definition of team reasoning, which to my knowledge has not yet been proposed for anything but pure coordination games; or an evidence-based restriction of its domain.

5. Response to HRS’s Other Claims Regarding CI’s Analysis

Section 4 responds to HRS’s main critique of CI’s analysis, and I have responded in passing to a couple of their minor criticisms. I now address their other criticisms.

Most importantly, HRS (2014, p. 1148) effectively ignore the fatal attraction pattern in RTH’s hide-and-seek games, which persists even in RTH’s treatments whose labels have connotations, and the similar (though not identical) patterns in HRS’s own data for hide-and-seek games with abstract labelings.\(^{10}\) As I have argued, the role asymmetry of the fatal attraction pattern is key indicator of the iterated-best-response reasoning characteristic of level-\(k\) as opposed to fixed-point reasoning: an important clue that any general model will need to have a level-\(k\) component. Yet HRS offer no model or conjecture about why RTH’s and their own subjects chose as they did. Instead their main response is to express doubts about whether CI really identified such a pattern.

---

\(^{10}\) In HRS’s treatments with games like those of RTH’s that CI studied, the analogue of central A is still modal for seekers and even more prevalent for seekers than hiders. But for hiders the analogues of the end As are now 11% more frequent than that of central A, which is still chosen above chance. For the reasons explained in CI (2007b), this pattern remains a puzzle for models other than a salience-sensitive level-\(k\) model.
I now explain CI’s (2007ab) arguments in more detail, so that readers can judge for themselves. Recall that CI focused on RTH’s hide-and-seek treatments with abstract labelings, arguably free of connotations that might affect preferences. Those treatments then induce zero-sum two-person games with unique equilibria, in which players randomize uniformly without regard to labeling. Despite the strong normative justification for such equilibria, RTH’s subjects’ action frequencies systematically deviated from equilibrium, in patterns that were sensitive to salience; and those patterns partly recurred in HRS’s own data for hide-and-seek treatments with abstract labeling.

CI (Section I) argued that all six of RTH’s hide-and-seek treatments with abstract labeling had analogous salience landscapes and the same qualitative pattern of deviations from equilibrium. Start with the treatment CI (2007a, Table I) called “RTH-4”. In that treatment subjects’ decisions were ordered left to right and labeled “A”, “B”, “A”, “A”. RTH and CI identified the “B” decision as salient via the uniqueness of its label; and the “end A” decisions as also salient, on which RTH cited Christenfeld (1995).11 Given these saliencies, RTH (and CI) viewed “central A” as the “least salient” decision. In this landscape, the least salient decision was modal for both hiders and seekers, and even more frequent for seekers: the so-called fatal attraction pattern.

11 HRS (footnote 21) criticize CI’s argument on this point: “The latter claim is supported by an unexplained citation of an experiment by Christenfeld (1995), which in fact found that when individuals pick from a row of identical items, they tend to avoid the end locations (CI, p. 1732).” As CI explained, this argument was first made by RTH (1996, p. 401). Further, HRS’s criticism rests on the implicit assumption that Christenfeld’s subjects, in a decision problem with no clear rewards, wished to favor salient decisions rather than avoid them. RTH (1993, p. 4) made clear that in their view, the “end A” locations are salient.
CI argued that RTH’s five other hide-and-seek treatments with abstract labeling had analogous salience landscapes, and that the fatal attraction pattern extends qualitatively to them. CI’s arguments are based on RTH’s intuitive conjectures about the least salient decisions in their various labelings; on RTH’s (1993) observation that their “treasure” and “mine” treatments have the same normal form with player roles interchanged and, mutatis mutandis, evoked roughly the same responses; and on the assumption that behavior is determined by the normal rather than the extensive form.12

With regard to RTH’s conjectures about the least salient decisions, no one doubts that uniquely labeled decisions are salient; and RTH’s belief that end locations are salient is at least plausible. RTH’s (1993) identification of “3” as least salient in their 1-2-3-4 treatments is less clear than that “central A” is least salient in A-A-B-A or A-B-A-A treatments because it is based on position alone, but it is still plausible.13 CI’s assumption that behavior is determined by the normal rather than the extensive form extends a common assumption in traditional game theory (e.g. Kohlberg and Mertens, 1986) to nonequilibrium models, in a way consistent with RTH’s findings (CI, 2007a, p. 1736).

Overall, CI’s analogies use simplifying assumptions to group observations in the process of identifying a model to make sense of otherwise puzzling patterns, a long-

---

12 RTH introduced their mine treatments to test whether the difference in the extensive form due to the fact that hiders must hide before seekers seek, might explain subjects’ deviations from equilibrium; and found that they did not.

13 HRS’s claim that CI’s assumed least salience of “3” was “[w]ithout further explanation” ignores CI’s (2007a, p. 1736) reference to RTH’s admittedly vague conjecture. Even omitting this treatment, the other treatments share a consistent pattern. HRS also critically mention CI’s assumption that subjects choose the end A locations with equal frequencies, which is an expository simplification suggested by the data on which nothing important turns.
standing and useful practice in empirical and experimental economies. Whether or not one agrees with every detail of CI’s arguments, RTH’s and HRS’s subjects deviated from equilibrium in role-asymmetric patterns that are indicative of iterated-best-response reasoning, not the fixed-point reasoning on which alternatives to level-k models rely.

Having raised these doubts about CI’s identification of the fatal attraction pattern, HRS (p. 1149) go on to argue that even if the fatal attraction pattern is real, CI’s level-k model is sufficiently flexible that it is not especially surprising that the model can account for the pattern. Specifically, HRS note that there are only 18 possible qualitative patterns of hiders’ and seekers’ action frequencies, and that some plausible form of a level-k model can account for seven of them. However, this way of judging flexibility by counting qualitative frequency patterns implicitly treats the patterns as random by giving them equal weight. As CI argued, their level-k model’s explanation of the fatal attraction pattern is surprising against the highly non-random background of alternative theories, none of which but their salience-sensitive level-k model (or its cognitive-hierarchy analogues) can explain the pattern. Put another way, if the fatal attraction pattern is real, it is just the kind of surprising regularity with the power to discriminate among alternative models that empirically oriented economists usually find informative.

Moving to less important criticisms, HRS (footnote 23) criticize CI’s portability analysis, saying “…in effect CI use a new \( L0 \) specification for each of the games”. As CI (2007a, p. 1746) explained, they in fact used plausible general principles to adapt \( L0 \) to the new hide-and-seek-like games with different patterns of salience, with no added degrees of freedom. O’Neill’s labeling, for instance, was A-2-3-J (playing cards), for
which CI took A and J (only) to be salient, as both face cards and end locations as the cards were presented to subjects. Rapoport and Boebel’s labeling was C-L-F-I-O, for which CI took C, F, and O to be salient because of their locations, on the assumption that no labels are salient per se. These identifications stop short of the general theory of salience HRS seem to wish for, but their implications are hardly controversial. I would be suspicious of any pure theory of salience claiming full generality across cultures.

6. Conclusion

I hope my arguments have made clear that the best chance for a general model of people’s responses to framing across games with a variety of structures is via constructive experiments with the goal of identifying a robust hybrid of team reasoning, level-\(k\) thinking, and perhaps other kinds of thinking, and of delineating which kinds of settings evoke which kinds of thinking. Bardsley, Mehta, Starmer, and Sugden (2009) find cross-country variations in the occurrence of team reasoning. Crawford, Gneezy, and Rottenstreich (2008) find evidence of something like team reasoning in some of their treatments with impure as well as pure coordination games. Progress along these lines will depend on developing either a more general, empirically grounded definition of team reasoning; or an evidence-based restriction of its domain.

Finally, although labels with connotations are not the main issue, future experiments in this area would be more useful if they avoided decision labelings with connotations like some of HRS’s, some as emotionally loaded as “Hitler”.\(^{14}\) It is a commonplace in

\(^{14}\) HRS (p. 1141) state that they used labels with connotations to “maintain subjects’ interest and attention”. I suggest that subjects’ interest and attention are better maintained by making the experimental tasks engaging.
marketing that labels with connotations influence subjects’ actions, and they plainly did so in HRS’s experiments. Because such influences are not controllable or observable, they sacrifice control of preferences while gaining little of interest in return.\footnote{Those influences would need to be modelled to draw useful inferences about theories of strategic behaviour. CI did not take a position on how to model the influence of connotations, because they focused on RTH’s treatments with abstract labels. But CI (2007b) showed how to use payoff perturbations to model similar influences. Such influences might even change the strategic structure, undermining the cross-game implications on which HRS’s tests depend. HRS (footnote 7) seek to address this criticism by allowing L0 to respond to salience, while restricting higher levels to respond only through their iterated best responses to such an L0 as in CI’s model. Because they assume, as CI did, that L0 players exist only in the minds of higher levels, this rules out any direct influence of connotations, unless connotations influence the results for their coordination games enough to change the discrete choice of L1. This does not happen in HRS’s data, but ruling out such responses prevents them from giving a full account of what their subjects are doing, much as would a version of consumer theory that ruled out any possible effect of brands.}

References


“Explaining Focal Points: Cognitive Hierarchy Theory \textit{versus} Team Reasoning.”


