"Fatal Attraction" and Level-*k* Thinking in Games with Non-neutral Frames

Vincent P. Crawford[1]

17 September 2018; revised 17 October 2018

Abstract: Traditional game theory assumes that if framing does not affect a game's payoffs, it will not influence behavior. However, Rubinstein and Tversky (1993), Rubinstein, Tversky, and Heller (1996), and Rubinstein (1999) 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) proposed a level-*k* explanation of Rubinstein et al.'s results for hide-and-seek games. Heap, Rojo-Arjona, and Sugden's (2014) criticized Crawford and Iriberri's model on grounds of portability. This paper clarifies Heap et al/'s interpretation of their results and responds to their criticism, suggesting a way forward.

Keywords: behavioral game theory; experimental game theory; strategic thinking; level-*k* models; coordination; salience. Declarations of interest: None

[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, Daniel Houser, and especially Nagore Iriberri for their valuable 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.

1. Introduction

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

The experiments of Rubinstein and Tversky (1993), Rubinstein, Tversky, and Heller (1996), and Rubinstein (1999) (henceforth collectively "RTH") provide some of the most powerful evidence on the effects of non-neutral framing in games. RTH elicited subjects' initial responses to two-person hide-and-seek, discoordination, and pure coordination games.[2] Their games had publicly known action labelings. Some labelings (summarized in Crawford and Iriberri, 2007a, Table 1; "CI") were abstract, such as A-B-A-A for each player's actions, without connotations that might directly influence payoffs; but with the pattern making some actions more and others less salient, in one of Schelling's (1960) senses. Others (CI, 2007b, Table A1) had positive or negative connotations, as in Frown-Smile-Frown-Frown, with the pattern again making some actions more, or less, salient.

In each case RTH's subjects deviated systematically from equilibrium in ways that were sensitive to the action labelings. Their results for hide-and-seek games with abstract

[2] In RTH's pure coordination games, players who choose actions with the same label receive payoff 1; otherwise both receive 0. In their discoordination games, players who choose actions with different labels receive payoff 1; otherwise both receive 0. In their hide-and-seek games, if players choose actions with the same label, the seeker receives payoff 1 and the hider receives 0; and vice versa if their actions have different labels.

labelings were particularly informative. These zero-sum two-person games have unique equilibria in mixed strategies, in which players randomize uniformly over their four actions. Even so, hiders and seekers both tended to favor the action the labeling RTH argued made least salient (like the "central-A" action in the A-B-A-A framing), with seekers favoring the least salient action even more than hiders and thereby obtaining significantly higher average payoffs than in the equilibrium.[3] Heap, Rojo-Arjona, and Sugden (2014) ("HRS") called this pattern of deviations from equilibrium the "fatal attraction" pattern, after CI's (2007a) title. That pattern, properly interpreted, persisted across RTH's six hide-and-seek treatments with abstract labels and, less clearly, in their five treatments in which the labels had connotations.

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 (2007ab) suggested that such systematic deviations from equilibrium in a game where the equilibrium is so obvious and its rationale is so strong are unlikely to have a

[3] 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.

nonstrategic explanation. Moreover, hide-and-seek's payoff structure, despite its role-*a*symmetry between hiders and seekers, makes equilibrium and its noisy generalizations like McKelvey and Palfrey's (1995) quantal response equilibrium ("QRE") predict role-symmetric behavior: QRE action probabilities coincide with the mixed-strategy equilibrium action probabilities for any well-behaved symmetric noise distributions, including logit (CI, 2007b). Thus the fatal attraction pattern cleanly separates traditional notions based on fixed points from non-equilibrium notions based on iterated best responses, which to my knowledge are the only models of behavior that map hide-and-seek's role-asymmetric structure into role-asymmetric patterns.

Motivated by these observations, CI (2007ab) proposed a level-*k* model to explain the fatal attraction pattern and the other aspects of RTH's results for hide-and-seek games.[4] In a level-*k* model, players anchor beliefs in a nonstrategic initial assessment of others' likely responses to the game, called "*L0*", and adjust them via iterated best responses, with *L1* best responding to *L0*, *L2* to *L1*, and so on. Players' levels are heterogeneous, drawn from a distribution concentrated on the lowest levels but excluding *L0*, which most evidence suggests exists only as the starting point for higher levels' thinking (Crawford, Costa-Gomes, and Iriberri, 2013, Section 2.4).

Even if *L0*'s frequency is zero, its specification is crucial. It is usually taken to be uniform random over a player's feasible actions. However, in games with non-neutral

[4] The only plausible alternative to a level-*k* model is Camerer, Ho, and Chong's (2004) cognitive hierarchy model, which differs from a level-*k* model in details that are not very relevant here (Crawford, Costa-Gomes, and Iriberri, 2013, p. 15). HRS, like CI, focus on the level-*k* version, as this paper will.

framing that is behaviorally implausible; and in RTH's hide-and-seek games it would make *L1* (and, with neutral decision errors, higher *Lk*s) mimic the equilibrium mixed strategies. Instead CI allowed *L0* to probabilistically favor salient actions. To avoid begging the question of role-asymmetric behavior, they constrained *L0* and the frequencies of higher levels to be the same for hiders and seekers.

CI's level-*k* model, estimated using RTH's data for their six hide-and-seek games with abstract action labels, tracks the main patterns in RTH's results for those games, including the fatal attraction pattern. CI's analysis shows that the pattern is by no means an inevitable consequence of a level-*k* structure; but observing it is powerful evidence for a level-*k* structure, given the lack of alternative models that can generate such a pattern.

CI (2007a, pp. 1734-5, 1743-8) tested their model for portability across games, but considered only games fairly similar to hide-and-seek. HRS (2014) report new experiments designed to assess the portability of CI's level-*k* model across a wider range of hide-and-seek, discoordination, and pure coordination games. HRS's games had structures like RTH's (footnote 1), although sometimes with more actions; and their treatments had groups including the three kinds of game, with uniform action labels across a group's games. Some sets of labels, including both abstract ones and those with connotations, were exactly like RTH's; others were similar in spirit but new to HRS's study (HRS, 2014, Figure 1, pp. 1140-42). HRS's test rests on strong assumptions about how a level-k model should port across the three kinds of game, based on a particular interpretation of CI's (2007a) motivation of L0. HRS show that no level-k model that

satisfies their assumptions about portability can explain their results. They also criticize other aspects of CI's analysis, and question the existence of the fatal attraction pattern.

The rest of this paper discusses CI's and HRS's analyses in more detail, clarifying CI's analysis and some criticisms that HRS make of it. Section 2 reviews CI's analysis in more detail and addresses some criticisms. Section 3 critiques HRS's analysis. Section 4 clarifies CI's identification of the fatal attraction pattern. Section 5 is the conclusion.

2. CI's Analysis of RTH's Results

Recall that in CI's (2007ab) level-*k* model, players anchor beliefs in an *L0* and adjust them via iterated best responses, with *L1* best responding to *L0*, *L2* to *L1*, and so on. Players' levels are heterogeneous, drawn from a distribution concentrated on the lowest levels, but excluding *L0*. *L0* is allowed to probabilistically favor salient actions, and its action probabilities and the level frequencies are constrained equal for both player roles.

CI used RTH's data from their six treatments with abstract action labels, omitting the five treatments where the labels had connotations as likely to induce uncontrolled payoff variation. However, the patterns in the data for the omitted treatments (CI 2007b, Table A1) are close enough to those for the treatments with abstract action labels to suggest that estimates for them would have been similar to those CI reported, though less clear.

CI (2007a, pp. 1738-40 and Table 3) estimated *L0* and the level frequencies under alternative constraints on how *L0* responds to salience for hiders and/or seekers. Given *L0*'s motivation, CI proposed a model in which it favors salience for both player roles. The only other important 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.[5] The resulting model reproduces the fatal attraction pattern (CI, 2007a, pp. 1738-9). Two alternative level-k models, which don't favor salience for both roles, also reproduce the fatal attraction pattern and fit slightly better; but do poorly in overfitting tests (CI, 2007a, pp. 1734).

In CI's proposed model, the fatal attraction pattern stems from interactions between iterated best responses to an *L0* that favors salience, and hide-and-seek's role-asymmetric payoff structure. Specifically, it follows from a hump-shaped estimated level distribution and an *L0* that makes *L1* hiders choose the least salient central-A action and *L1* seekers avoid it; *L2* hiders choose it with positive probability (due to payoff ties) and *L2* seekers choose it; *L3* hiders avoid it and *L3* seekers choose it with positive probability. However, the pattern is by no means an inevitable consequence of a level-*k* structure: For it to emerge, the heterogeneity of levels is crucial, and the frequencies of *L2* and *L3* must be large enough to make seekers choose central-A even more often than hiders. Conversely, however, observing the pattern in a setting like hide-and-seek is powerful evidence for a level-*k* structure, given the lack of alternative models that can generate such a pattern.

CI (2007a, pp. 1736-37) also considered alternative, equilibrium-based models of players' responses to salience, with "hard-wired" payoff perturbations that reflect a

[5] HRS (pp. 1133-1134) imply that this aspect of CI's estimation makes their model prone to overfitting. However, all that matters about the continuously variable probabilities that define *L0* in CI's model is the best-response *L1* 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 price to pay for explaining a non-zero response to salience.

plausible instinctive attraction to salience for seekers and an aversion to salience for hiders.[6] CI (2007b) 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) also showed that logit QRE versions of the models with perturbations do no better, and generally worse, than equilibrium-with-perturbations. Such models even predict that hiders favor the least salient action more than seekers, the opposite of the fatal attraction pattern. In fact, maximum likelihood estimates of the QRE models' noise parameters are 0, reducing them to equilibrium-with-perturbations models.

Although CI did not explicitly consider RTH's treatments in which the labels had connotations, payoff perturbations like those in CI's equilibrium-based analyses could be used to model their effects directly, in either the level-*k* or the equilibrium version of the model. RTH's data (CI 2007b, Table A1) suggest that the results of such analyses would not differ greatly from the results for RTH's treatments with abstract action labels.

Finally, CI, aware that their estimation gives their level-*k* model more freedom than equilibrium or QRE, tested their model's portability by adapting it to the fairly similar

[6] Superficially, this role asymmetry differs from the symmetric assumption in CI's proposed level-*k* model that *L0* favors salience for both hiders and seekers. However, that *L0* is meant to describe a strategically naive initial reaction, while the perturbations in CI's equilibrium-with-perturbations models directly determine actual strategic behavior. An equilibrium-with-perturbations model with role-symmetric attractions to salience would fit poorly.

structures and salience patterns of O'Neill's (1987) and Rapoport and Boebel's (1992) games, using the level frequencies estimated from RTH's data and defining *L0* via the same principles as for RTH's games. The adapted model describes O'Neill's subjects' early responses well and Rapoport and Boebel's subjects' responses fairly well. By contrast, equilibrium- or logit QRE-with-perturbations models describe O'Neill's and Rapoport and Boebel's results no better than RTH's results (CI 2007a, pp. 1743-8).

HRS (footnote 23) criticize CI's portability analysis, saying "…in effect CI use a new *L0* specification for each of the games". But as CI (2007a, p. 1746) explained, they 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 face cards and end locations. 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. These identifications stop short of the general theory of salience HRS seem to wish for, but they are hardly controversial, and I suspect that a general theory of salience is not possible.

3. HRS's Analysis

As already noted, HRS (2014) reported new experiments specifically designed to test the portability of CI's level-*k* model across games with widely different structures. HRS's games had structures like RTH's hide-and-seek, discoordination, and pure coordination games (footnote 1), though some had more actions. Unlike RTH's subjects, HRS's played the games within subjects across groups of games including all three kinds of game, with the numbers and labelings of actions constant within groups and across

player roles. Some sets of labels, some abstract ones and some with connotations, were exactly like RTH's; others were similar but new to HRS (2014, Figure 1, pp. 1140-42).

HRS used their results to test a level-*k* model like CI's. 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 the same for both player roles. They report results for treatments where action labels have connotations as well as those with abstract labels; but their main conclusions hold even if the analysis is restricted to the latter treatments.

HRS's test rests on assumptions about how a level-*k* model should port across hide-and-seek, discoordination, and pure coordination games. Crucially, HRS argued that because *L0* 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 the three kinds of game are strategic, a level-*k* model should hold equally across each of their groups of hide-and-seek, coordination, and discoordination games with the same actions and labelings, and with constant *L0* and level frequencies.

HRS's constant-*L0* assumption implies cross-game restrictions on behavior. In each group of games, the action labelings make one label, which they call the "oddity", uniquely different from the others. In their pure coordination games, the odd action (as I will call it) is subjects' majority choice for almost all groups and labelings. Those games are symmetric across player roles, and in a level-*k* model the odd action must then be *L0*'s modal choice. With *L0* constant across games and player roles, the odd action's average frequency across discoordinators, hiders, and seekers must be disproportionately

high (HRS, p. 1138, Implication 1). However, HRS's subjects chose the odd action 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 by implication that level-*k* models lack the portability needed to be useful.

Experiments that test the portability of models of strategic behavior across games with wider ranges of strategic structures are welcome.[7] However, HRS's cross-game restrictions are neither theoretically justified nor behaviorally credible. CI's (2007ab) and Crawford, Costa-Gomes, and Iriberri's (2013) motivations of *L0* as strategically naïve meant only 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. As intuition suggests, even strategically naïve people can distinguish pure coordination from other kinds of games; to suggest otherwise is akin to a play on words.

Further, there is ample evidence 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;

[7] 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.

Crawford, Gneezy, and Rottenstreich, 2008, p. 1448). Team reasoning involves fixed-point reasoning, which is different in kind from level-*k* thinking. To expect a level-*k* model (with *L0* constant or not) to track subjects' behavior across hide-and-seek, discoordination, and pure coordination games is a behavioral non-starter, based on a questionable interpretation of "strategically naïve". The failure of a level-*k* 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.

HRS (p. 1135) acknowledge that a level-*k* model with 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 alternatives; nor do they acknowledge that the empirically most promising model for pure coordination games is not even a level-*k* model, but rather team reasoning.

In motivating their approach, HRS argue that a useful general model must be well-defined in advance for any game, and must be evaluated via ex ante hypothesis testing, not ex post model fitting (HRS, p. 1135). This poses a significant obstacle in settings like those RTH or HRS study, because to my knowledge team reasoning has been defined only for pure coordination games. I agree that ex ante definition and hypothesis testing is preferable, 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.

4. The Fatal Attraction Pattern

This section clarifies CI's identification of the fatal attraction pattern. HRS (2014, p. 1148) effectively ignore the pattern in RTH's hide-and-seek games, which persists in treatments whose labels have connotations, and the similar but not identical patterns in HRS's own data for hide-and-seek games with abstract labelings.[8] As argued above, the fatal attraction pattern's role asymmetry is a key indicator of iterated-best-response as opposed to fixed-point reasoning, an important clue that any general model will need to have a level-*k* component. Yet HRS make no attempt to explain why RTH's and their own subjects showed such patterns in hide-and-seek and related games.

Instead HRS's main response is to express doubts about whether CI (2007ab) really identified a fatal attraction pattern. So readers can judge for themselves, I now explain CI's arguments in more detail. CI (2007a, 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 (Table I, p. 1735) called "RTH-4". There, subjects' actions are ordered left to right and labeled "A", "B", "A", "A". RTH and CI identified the "B" action as salient via the

[8] 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.

uniqueness of its label; and the "end-A" actions as also salient, on which RTH cited Christenfeld (1995).[9] Given these saliencies, RTH and CI viewed "central-A" as the "least salient" action. In this landscape, the least salient action was modal for both hiders and seekers, and even more frequent for seekers: the fatal attraction pattern. CI (2007b) argued that RTH's five other hide-and-seek treatments with abstract labeling had analogous salience landscapes, and that the pattern extends qualitatively to them.

In each case CI's arguments are based on RTH's conjectures about the least salient decisions in their labelings; on RTH's (1993) observation that their "treasure" and "mine" treatments have the same normal form with player roles interchanged and evoked roughly the same responses; and on the assumption that behavior is determined by the normal rather than the extensive form.[10] With regard to RTH's conjectures: 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 the A-A-B-A or A-B-

[9] 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 actually 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, sought to favor salient decisions rather than avoid them. RTH (1993, p. 4) made clear that in their view, the end-A locations are salient.

[10] 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.

A-A treatments because it is based on position alone, but it is still plausible.[11] CI's assumption that behavior is determined by the normal rather than the extensive form extends a common assumption in traditional game theory (Kohlberg and Mertens, 1986) to nonequilibrium models, and is consistent with RTH's findings (CI, 2007a, p. 1736).

Overall, CI's analogies use simplifying assumptions to group observations and identify a model to make sense of otherwise puzzling patterns, a long-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 hide-and-seek 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 pattern is real, CI's level-*k* model is sufficiently flexible that it is not surprising that the model can account for it. Specifically, HRS note that some plausible form of a level-*k* model can account for seven of the 18 possible qualitative patterns of hiders' and seekers' action frequencies. However, 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

[11] 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.

alternative theories, none of which but models based on iterated best responses can explain the pattern. The fatal attraction pattern is just the kind of surprising regularity that discriminates among alternative models that empirical economists find informative.

5. Conclusion

If a level-*k* model is not suitable as a fully general model of strategic responses to framing, what might a suitable model look like? I have argued that 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 of people's responses to framing across games with a variety of structures is not via "completing level-*k* theory" as HRS (pp. 1149-50) suggest, but rather via constructive experiments with the goal of identifying a robust hybrid of team reasoning, level-*k* thinking, and possibly other kinds of strategic thinking that people use in different games; and of mapping the boundaries between the kinds of games 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, a general model 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.[12]

Finally, although labels with connotations are not the main issue, future experiments in this area would be more useful if they avoided action labelings with connotations like some of HRS's, some as emotionally loaded as "Hitler".[13] It is a commonplace in 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.[14]

[12] For example, 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.

[13] 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.

[14] Those influences would need to be modelled to draw useful inferences about theories of strategic behavior. CI (2007a) 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 *L1*'s choice. This does not happen in HRS's data, but ruling out such responses prevents 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

Bardsley, Nicholas, Judith Mehta, Chris Starmer, and Robert Sugden (2009): "Explaining Focal Points: Cognitive Hierarchy Theory *versus* Team Reasoning." *Economic Journal*, 120(2): 40-79.

Camerer, Colin F., Teck-Hua Ho, and Juin-Kuan Chong (2004): "A Cognitive Hierarchy Model of Games." *Quarterly Journal of Economics*, 119(3): 861–98.

Christenfeld, Nicholas (1995): "Choices from Identical Options." *Psychological Science*, 6(1): 50-55.

Costa-Gomes, Miguel A., and Vincent P. Crawford (2006): "Cognition and Behavior in Two-Person Guessing Games: An Experimental Study." *American Economic Review*, 96(5): 1737-1768.

Crawford, Vincent P., Miguel A. Costa-Gomes, and Nagore Iriberri (2013): "Structural Models of Nonequilibrium Strategic Thinking: Theory, Evidence, and Applications." *Journal of Economic Literature*, 51(1): 5-62.

Crawford, Vincent P., Uri Gneezy, and Yuval Rottenstreich (2008): "The Power of Focal Points is Limited: Even Minute Payoff Asymmetry May Yield Large Coordination Failures." *American Economic Review*, 98(4): 1443-1458.

Crawford, Vincent P., and Nagore Iriberri (2007a): "Fatal Attraction: Salience, Naivete, and Sophistication in Experimental Hide-and-Seek Games." *American Economic Review*, 97(5): 1731-1750.

Crawford, Vincent P., and Nagore Iriberri (2007b): Web appendix for "Fatal Attraction: Salience, Naivete, and Sophistication in Experimental Hide-and-Seek Games." http://www.aeaweb.org/aer/data/dec07/20050133_app.pdf.

Georganas, Sotiris, Paul J. Healy, and Roberto A. Weber (2015): "On the Persistence of Strategic Sophistication", *Journal of Economic Theory* 159A(1): 369-400.

Heap, Shaun Hargreaves, David Rojo-Arjona, and Robert Sugden (2014): "How Portable is Level-0 Behavior? A Test of Level-k Theory in Games with Non-neutral Frames." *Econometrica*, 82(3): 1133–1151.

Kohlberg, Elon, and Jean-François Mertens (1986): "On the Strategic Stability of Equilibria." *Econometrica*, 54(5): 1003–1037.

McKelvey, Richard, and Thomas Palfrey (1995): "Quantal Response Equilibria for Normal-Form Games." *Games and Economic Behavior*, 10(1): 6-38.

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.

Penczynski, Stefan (2016): "Strategic Thinking: The Influence of the Game." *Journal of Economic Behavior and Organization*, 128: 72-84.

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.

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; "Second Edition" posted at http://arielrubinstein.tau.ac.il/99/gt100.html .

Rubinstein, Ariel, and Amos Tversky (1993): "Naïve Strategies in Zero-Sum Games." Working Paper 17-93, Sackler Institute of Economic Studies, Tel Aviv University.

Rubinstein, Ariel, Amos Tversky, and Dana Heller (1996): "Naïve Strategies in Competitive Games." In *Understanding Strategic Interaction: Essays in Honor of Reinhard Selten*, ed. Wulf Albers, Werner Güth, Peter Hammerstein, Benny Moldovanu, and Eric van Damme, 394–402. Berlin: Springer-Verlag.

Schelling, Thomas. 1960. *The Strategy of Conflict*. Cambridge: Harvard University Press.