No. 94

Irenaeus Wolff

On the Salience-Based Level-k Model

Research Paper Series
Thurgau Institute of Economics and Department of Economics
at the University of Konstanz

On the Salience-Based Level-k Model[§]

Irenaeus Wolff

Thurgau Institute of Economics (TWI) / University of Konstanz
Hauptstrasse 90, 8280 Kreuzlingen, Switzerland
wolff@twi-kreuzlingen.ch

Abstract:
Crawford and Iriberri (AER, 2007) show how a level-k model can be based on salience to explain behaviour in games with distinctive action labels, taking hide-and-seek games as an example. This study presents four different experiments designed to measure salience. When based on any of these empirical salience measures, their model does not explain behaviour. Modifying the model such that players follow salience when payoffs are equal, the model fits hide-and-seek data well. However, neither the original nor the modified model account for data from a discoordination game. This holds true even when incorporating the heterogeneity in measured salience perceptions.

Keywords: ABAA, hide and seek, cognitive hierarchy, strategic reasoning, saliency.
JEL: C72, C91

1 Introduction

Matching-pennies games and their generalisation to multiple actions, dubbed hide-and-seek games, have been well-studied games in game theory from its

[§]I am grateful to my co-authors Lisa Bruttel, David Dohmen, Timo Heinrich, Konstantin Hesler, and Simeon Schudy, for their cooperation on projects that produced some of the data I am using here; the latter four also contributed substantially to this paper through uncountable discussions on the level-k explanation for behaviour in hide-and-seek games. I would like to thank the lively research group at the Thurgau Institute of Economics (TWI) for helpful comments all along the way, as well as the participants of the 2013 GfeW meeting for the fruitful discussions. I am indebted to Marie-Claire Villeval, Dirk Sliwka, and in particular to Roberto Weber and Urs Fischbacher for their encouragement to go yet another step further to make the paper complete. I thank Vincent Crawford and Nagore Iriberri for rapidly answering any questions with respect to their paper, as well as for their comments on an earlier version of this paper. Last but not least, I am grateful to Shaun Hargreaves Heap, David Rojo Arjona and Robert Sugden for a lively discussion on the paper. Financial support by the University of Konstanz' Young Scholar Fund is gratefully acknowledged.

very beginning (cf., e.g., von Neumann, 1953). Real-life examples abound, from markets in which the brand leader will continue to have the largest revenues as long as it can match the rival products' features, to obvious applications in military, police, and intelligence work. While the standard game-theoretic solution to the generic games is straightforward, experiment participants do not seem to act according to this prediction (see Eliaz and Rubinstein, 2011, for a repeated matching-pennies game, and Rubinstein and Tversky, 1993, as well as Rubinstein, Tversky and Heller, 1996, for hide-and-seek games).

In a well-noted paper, Crawford and Iriberri (2007; henceforth CI) show how a salience-based level-k model can account for the observed patterns in hide-and-seek games if we assume a specific salience pattern. Current work by Hargreaves Heap, Rojo Arjona and Sugden (2014) shows that a level-k model cannot account simultaneously for data from hide-and-seek games, coordination games, and discoordination games all played on the same action-set frame if we assume the same salience pattern for all games. I provide empiciral measures of the salience pattern and argue that both assumptions are empirically inadequate. None of four salience measures is in line with CI's assumption on salience, and the evidence suggests that the game description changes participants' salience perceptions.

The important question ensuing from the findings on participants' salience measures is what they mean for the proposed level-k model. I show that the proposed model no longer predicts behaviour well when based on empirical salience measures. A simple and plausible modification restitutes the remarkable fit of CI's level-k model. However, neither the original model nor its modification can explain data from discoordination games. Most importantly, I show that a level-k model based on empirical salience does not account for the discoordination-game data even when we account for the fact that there is heterogeneity in participants' elicited salience perceptions (so that every participant may have their own level-0).

Before I say more about the model and its modification, let me briefly present the hide-and-seek game in its archetype version. A "hider" possesses a "treasure" she can hide in one of four boxes, labelled "A", "B", "A", and "A". A "seeker" may open one of these boxes. If he chooses the same box as the hider, the seeker gains the treasure, otherwise the hider keeps it. This multiple-action matching-pennies game obviously has a unique Nash equilibrium in mixed strategies, with both the hider and the seeker choosing each box with 25% probability. The typical distribution observed in experimental implementations of the game, on the other hand, has a strong mode on "central A" for both roles, being even more pronounced for seekers than for hiders (which leads to a substantial seeker-advantage relative to equilibrium).

Let us now turn to how a level-k model may account for the above pattern.

Level-k models have a very simple structure. Each k-type, $k > 0$, believes all her opponents are of level-$(k-1)$ and best-responds to this belief.[1] The two crucial elements to close the model are the specifications of level-0—which is assumed to exist in the players' minds only—and of the type distribution. For the latter, CI argue that "[t]he estimated distribution tends to be stable across games and hump-shaped," (p. 1734) while the level-0 specification is the central innovation of their paper. Instead of assuming the traditional uniform mixture over all possible actions, CI 'translate' Rubinstein and co-authors' statements on salience into (latent) numeric variables to use them as level-0 in their model. In CI's words,

> "[t]he 'B' location is distinguished by its label, and so is salient in one of Thomas Schelling's (1960) senses. And the two 'end A' locations, though not distinguished by their labels, may be inherently salient, as RT [Rubinstein and Tversky, 1993] and RTH [Rubinstein, Tversky, and Heller, 1996] 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.' " (p. 1732).

CI translate the last sentence as implying that "central A" really *is* "the least salient location," thus being chosen by a level-0 player least often. I argue that this need not be true. If "central A" has "its own brand of uniqueness", it is not clear *a priori* how it should be ranked in terms of salience. The evidence presented in this paper suggests "central A" is in fact more salient than "final A", whereas it is unclear how it compares to "first A" in terms of salience.

CI deliberately keep the model clear of salience influences except for its role in determining level-0. However, one could argue that *when players are indifferent* between various actions, they will act in the same way as if they were given no incentives at all. Following the argument of Mehta, Starmer and Sugden (1994), we should expect those players' actions to be shaped by salience. In section 3.5, I introduce this small but important twist and show that the revised model has a fit of comparable order as the best models in Crawford and Iriberri (2007). At the same time, it leads to a more plausible estimation of the level-k distribution compared to CI's model when constrained to an empirically-elicited

[1] This type of model was introduced by Stahl and Wilson (1994, 1995) and Nagel (1995), and later adapted by Costa-Gomes, Crawford and Broseta (2001). It is closely-related to other *cognitive-hierarchy* models like that proposed by Ho, Camerer and Weigelt (1998) and refined in Camerer, Ho and Chong (2004). For a discussion of both approaches, cf. Crawford, Costa-Gomes and Iriberri (2013). Note that Crawford and Iriberri (2007) allow for errors in their model. However, given the estimated error rate for the models under the assumption of uniform errors is zero, I abstract from errors for the time being. None of the findings hinges on this simplification.

level-0. Nonetheless, neither the modified nor CI's original model can account for data from a discoordination game played on an 'A-B-A-A landscape'. This continues to hold true when I relax the assumption that all participants act on the same salience perception and allow for a heterogeneous level-0.

This paper contributes to a growing literature that finds empirical support for level-k-like thinking in a variety of games.[2] It also contributes to a small but growing literature on how salience shapes behaviour and how this can be incorporated into game-theoretic models.[3] Crawford and Iriberri (2007) do a remarkable job in joining these two branches of the literature. What the present paper shows is that salience influences on behaviour remain a phenomenon that is not as straightforward as it may seem. It is not obvious *a priori* what is salient in the eyes of experimental participants, and it remains to be understood how exactly salience shapes participant behaviour.

The remainder of this paper is organised as follows: in Section 2, I present the four salience-elicitation experiments. Section 3.1 presents the results of these experiments. On the basis of these results, I modify the equilibrium model with salience-based payoff-perturbations CI use as a benchmark in Section 3.3. In Section 3.4, I present the predictions of Crawford and Iriberri's model and report on the resulting model fits for all models presented in their paper when the respective models are based on an experimentally-elicited salience-pattern. Section 3.5 presents two variants of a potential modification of Crawford and Iriberri's level-k model and evaluates them in terms of their fit to the data. In Sections 3.6 and 3.7, I analyse the predictive power of the different variants in out-of-sample and out-of-game predictions. Section 3.8 incorporates a heterogeneous level-0 and evaluates the resulting models' data fit. Section 4 summarises the data and discusses the findings. An explanation of the model denotations used throughout the paper can be found in Section 3.2.

2 Design of the salience-elicitation experiments

The purpose of the salience-elicitation exercise is to provide a clearer understanding of what may constitute an adequate level-0 specification for the model. I argue that there are multiple ways of how salience could determine level-0 that are associated with distinct empirical measures. On the level of beliefs involved, I follow three approaches: a first approach is to define level-0 directly in terms of the available actions' salience (*primary salience* in Bardsley et al., 2010, referring to Lewis, 1969). This corresponds most closely to CI's proposed *model.* A second

[2]E.g., Burchardi and Penczynski (2014) or the many papers cited in Crawford, Costa-Gomes and Iriberri (2013).

[3]E.g, Mehta, Starmer and Sugden (1994) or Bardsley et al. (2010).

approach is to ask what people think will be salient for other people (*secondary salience* in Bardsley et al., 2010, also referring to Lewis, 1969). This corresponds more closely to CI's *general reasoning* about level-0, given level-0 is meant to exist only in the players' minds. Finally, we may be tempted to argue that the truly relevant aspect would be to ask what people think others will think everybody will hold to be salient, and so on, *ad infinitum*.[4,5]

Having looked at the above 'levels of salience', I also want to test whether the game description will shape the salience of the available actions. More precisely, players may assess an action's salience differently, depending on whether they look at the actions *per se*, or whether they look at the actions taking into account the game they will be playing.[6] In the latter case, it would be plausible also to assume that players' roles may affect their salience assessment.[7]

In this study, I examine four experimental measures of a salience-based level-0. The first three measures are a full variation along the belief dimension, keeping the game description out. The fourth measure uses the *secondary-salience* measure to explore the effect of introducing the game story (and whether an asymmetry follows from that). To be precise, I look at the following experiments:

Picking Task. Elicitation of the different boxes' salience by asking people to choose one of four boxes labelled "A", "B", "A", and "A", and click on the chosen box, on a separate page of a post-experimental questionnaire after an unrelated experiment.[8] This is the "picking task" Bardsley et al. (2010) use to elicit *primary salience*. As a crucial complementary measure to assess salience, I record response times for this task.

[4]Bardsley et al. (2010) point out there may be higher 'levels of salience' but argue that they are likely to coincide with *secondary salience*. My results would support this conjecture.

[5]It could be argued that only primary salience should be a candidate for level-0, given secondary and higher-order salience involve strategic thinking in the sense of guessing about others' perceptions or even about others' reasoning about these perceptions. Because strategic thinking is what the model should explain, so the argument, we should not include strategic thinking as an input into the theory *via* level-0. I agree that a model based on an empirical measure of what players think others will regard as salient is *incomplete*. At the same time, I disagree that we should be able to discard level-k theory already by showing that primary and secondary salience are different. Rather, I would see a model able to explain choices based on empirically-measured secondary salience as a fruitful first step to a more complete model that gets rid of the problem. Because of this, I include also secondary and 'infinite-level' salience as candidates for level-0. I thank Shaun Hargreaves Heap, David Rojo Arjona and Robert Sugden for raising this point.

[6]In fact, the study by Hargreaves Heap, Rojo Arjona and Sugden (2014) suggests this may be the case.

[7]CI partially incorporate this latter aspect by presenting different model specifications, e.g., including a salience-seeking level-0 seeker and a salience-avoiding level-0 hider.

[8]The post-experimental questionnaire mainly contains questions from the 16PF personality inventory. Participants have not participated in any hide-and-seek experiment before.

GUESSING TASK. Elicitation of what people think will be salient for other people. For this purpose, we ask participants to estimate the relative click frequencies from the answers elicited by the PICKING TASK.[9] This is the "guessing task" Bardsley et al. (2010) use to elicit *secondary salience*.

BEAUTY CONTEST. A beauty contest on the consensus on what is salient. The contest anchored in the question "which is the most salient box, which are the second, third, and fourth most salient boxes?" and was conducted as a classroom experiment in the Experimental Methods course.[10]

POST-GAME GUESSING. Elicitation of what people who know the game think will be salient for others. For this purpose, we asked participants to estimate the relative click frequencies of the PICKING TASK responses. This was done *after* they had played the hide-and-seek game but before they got any feedback. This measure serves as a benchmark for how the game—and possibly, the role—changes salience-perceptions.[11]

None of the participants participated in more than one of the four experiments. All experiments were run at the University of Konstanz, the PICKING TASK, the GUESSING TASK, and the POST-GAME GUESSING at its *Lakelab*.

As an additional measure of primary salience, I include the predictions of a salience-based model of visual attention (ALGORITHM). This model has been extended from Itti, Koch and Niebur (1998) by EyeQuant Attention Analytics (www.eyequant.com) based on eye-tracking studies and psychophysics experiments (for another successful application, cf. Towal, Mormann and Koch, 2013).

3 Results

This section is organised as follows: first, I report on the outcomes of the salience-elicitation exercises. In Section 3.2, I introduce the denominations of all models that appear in the paper. Then, I briefly present what the elicited salience-measures mean for Crawford and Iriberri's (2007) benchmark model of an equilibrium with payoff perturbations, in Section 3.3. Following that, I look at what the salience-elicitation exercises would mean for the model variant proposed by

[9]The task was incentivised in the following way: if no frequency differed from the true value by more than 5% (10%/20%), participants could earn an additional 50 (25/10) Euro cents, otherwise, they did not earn anything. The task was the first task participants faced in the experiment, they knew there would be further tasks, but they did not know what those tasks would be.

[10]Amongst those stating the modal ordering, a prize of 12 Euros (about USD 15.60 at the time) was raffled off.

[11]Incentives as in the GUESSING TASK.

Crawford and Iriberri (2007) in Section 3.4, by replicating their model-fitting exercise using the elicited salience patterns as level-0. In Section 3.5, I propose a modification of the model and redo the fitting exercise for this modified version. In Sections 3.6 and 3.7, I evaluate the best-fitting models by their ability to predict out-of-sample and out-of-game. Finally, in Section 3.8 I analyse whether a level-k model based on a heterogeneous level-0 can account for coordination-game and discoordination-game data from Section 3.7. For ease of notation, in the remainder of this article I will describe the locations "A", "B", "A", and "A" by $A_{(1)}$, $B_{(2)}$, $A_{(3)}$, and $A_{(4)}$, respectively.

3.1 Salience in the ABAA hide-and-seek game

The results of the four salience-elicitation experiments and the additional ALGORITHM prediction are reported in Table 1, together with the respective numbers of independent observations (where applicable).[12]

Observation 1. $B_{(2)}$ is the most salient alternative.

This can be easily seen by looking at the second data column in Table 1. $B_{(2)}$ is the alternative chosen most often in the PICKING TASK, where it also is the fastest choice (The p-values of Wilcoxon-Mann-Whitney-tests for response times are 0.085, 0.061, and 0.001, for the comparisons with $A_{(1)}$, $A_{(3)}$, and $A_{(4)}$);[13] it is predicted to obtain the most attention by the ALGORITHM; in the GUESSING TASK and the POST-GAME GUESSING, it on average is estimated to be clicked on the most by a margin of 19% and 15%, respectively; and it ranks first in the BEAUTY CONTEST, no matter whether one looks at the winning ordering or at average ranks.[14]

Observation 2. $A_{(4)}$ rather than $A_{(3)}$ is the least salient alternative, possibly in conjunction with $A_{(1)}$.

[12] www.eyequant.com offers two versions of the model, one for the first impression of new visitors, and one for "engaged visitors". Here, we report the figures for new visitors. The normalised predicted relative attention for $B_{(2)}$ (38%) and $A_{(3)}$ (25%) under the "engaged-visitors" model differs only marginally, while $A_{(1)}$ (16%) and $A_{(4)}$ (21%) shift positions. This fluctuation in the bordering As could be read as indicating that they are similarly salient after all. The analysis can be found on http://www.wiwi.uni-konstanz.de/fischbacher/home/staff/dr-irenaeus-wolff/.

[13] In order not to favour the options I expected to be seen as most salient, the cursor was placed at the bottom right of the screen before the PICKING TASK, and therefore, closest to "final A".

[14] As the crudest-possible statistical measure, assume that the next-salient candidate has an equal chance of coming out as the most salient alternative on each of the six measures (frequency and response time in the PICKING TASK, ALGORITHM prediction, estimate by GUESSING-TASK participants and by hiders and by seekers in POST-GAME GUESSING, and any ranking measure in the BEAUTY CONTEST), and then compute the p-value of the according binomial test to be $p = 1/64$.

	A(1)	B(2)	A(3)	A(4)
Picking Task (405 participants)				
relative click frequencies (in %)	21	38	35	6
response times (in sec)				
mean	8.8	7.7	8.5	11.9
median	8.0	7.1	7.5	9.4
Algorithm				
predicted relative attention (in %)	20	37	24	18
Guessing Task (72 participants)				
average estimated relative click frequency	21	41	22	15
Beauty Contest (30 participants)				
rank in beauty contest				
winning order (chosen by 14 participants)	2	1	3	4
mean ranks	2.3	1.5	2.5	3.6
Post-Game Guessing (156 participants)				
average estimated relative click frequency	19	39	24	18
...by hiders (78 obs.)	19	38	24	19
...by seekers (78 obs.)	19	40	25	17

Table 1: Salience assessments of the four boxes denoted by “A”, “B”, “A”, and “A”.

Looking at the final column of Table 1, we see that all measures indicate $A_{(4)}$ is the least salient option.[15] In Post-Game Guessing, $A_{(1)}$ and $A_{(4)}$ may be considered to be jointly the least salient locations. $A_{(3)}$ in all measures is elicited to be the second-most (Picking Task, Algorithm, Guessing Task, and Post-Game Guessing) or third-most (Beauty Contest) salient location.

Observation 3. The description of the game alters the qualitative pattern of participants' assessment of others' salience perceptions. At the same time, participants' roles in the hide-and-seek game do not seem to influence their estimates of other people's salience perceptions.

For the first part of Observation 3, compare the average estimated relative click frequencies of the Guessing Task and the Post-Game Guessing. While the largest quantitative difference between the corresponding average estimates in the Guessing Task and the Post-Game Guessing is a mere 3%, there seems to be a clear difference in the qualitative pattern. In the Guessing Task, there is a substantial difference between the average estimated relative click frequency of $A_{(1)}$ and $A_{(4)}$ (and none between $A_{(1)}$ and $A_{(3)}$), while in the Post-Game Guessing, the average estimates for $A_{(1)}$ and $A_{(4)}$ are virtually identical (and there is

[15] In this case, all response time comparisons are associated with p-values below 0.02.

a clear difference with respect to $A_{(3)}$).[16] For the second part of Observation 3, note that within the POST-GAME GUESSING the qualitative pattern clearly is the same for hiders and seekers, and the quantitative difference between the average estimates is 2% at most.

3.2 Preliminaries: model denominations

Throughout this paper, I will work with a variety of models to account for behaviour. Table 2 is meant to systemize them sufficiently so that it is easier to refer to the different models in the text. There are two main aspects on which the models differ: the salience-pattern on which the model is based, and on the hypothesized strategic thinking given this salience-pattern. In terms of the latter, I will refer to three kinds of models: NAÏVE responses driven by salience that do not require any strategic thinking at all; equilibrium (EQM) models, potentially including salience-based payoff perturbations; and level-k (Lk)-models. In terms of the salience-pattern used, I will refer to the hypothesised pattern in Crawford and Iriberri (2007; HYP-$\left[A_{(1)}A_{(4)}\right]B_{(2)}A_{(3)}$), the PICKING-TASK data (PICK-$B_{(2)}\left[A_{(3)}A_{(1)}\right]A_{(4)}$; representing also the measures from the GUESSING TASK and the BEAUTY CONTEST), and different POST-GAME GUESSING experiments (POSTX, where X is a wildcard referring to the respective game). In Section 3.8, I will relax the assumption that all players have the same salience perception and allow for a heterogeneous level-0 (INDL0). To give an example, Lk_{SOPH}-POSTDISCOORD-INDL0 denotes a level-k model that is based on individuals' salience perceptions as elicited in a guessing task played after a discoordination game, where players follow salience when indifferent for payoff reasons and where higher-level players are aware of lower-level players' randomisation 'technique'.

3.3 Crawford and Iriberri's (benchmark) equilibrium model with payoff perturbations

In the following section, I briefly present the model of an equilibrium with hardwired payoff perturbations CI use as a benchmark. CI start with the normal form game and posit that players will have a preference for some locations which depends on those locations' salience. Hiders are assumed to dislike choosing salient locations, while seekers are assumed to favour them. Here, I use the salience

[16]Wilcoxon matched-pairs signed-ranks tests support these observations: in the GUESSING TASK, they yield $p \leq 0.001$ for the comparisons of a participant's $A_{(4)}$-estimate with both her $A_{(1)}$-estimate and her $A_{(3)}$-estimate, while for the comparison of her $A_{(1)}$-estimate with her $A_{(3)}$-estimate, the test yields $p = 0.883$. In POST-GAME GUESSING, the same test yields $p = 0.133$ for the comparison between $A_{(1)}$ and $A_{(4)}$, and $p < 0.001$ for the comparisons between $A_{(3)}$ and both $A_{(1)}$ and $A_{(4)}$.

Model of strategic thinking	
NAÏVE	Players choose according to salience, no strategic thinking involved.
EQM$_0$	Assumes rationality and common knowledge thereof; no payoff perturbations.
EQM$_+$	Assumes rationality and common knowledge thereof; there are payoff perturbations that follow salience.
Lk	Each level-k player best-responds to a level-$(k-1)$ player; when best-response is not unique, players randomise uniformly over all best-responses.
Lk_{UNSOPH}	Each level-k player best-responds to a level-$(k-1)$ player; when best-response is not unique, players randomise according to the best-responses' salience; when a level-k' player randomises, a level-$k'+1$ player best-responds to a uniformly-randomising level-k' player.
Lk_{SOPH}	Each level-k player best-responds to a level-$(k-1)$ player; when best-response is not unique, players randomise according to the best-responses' salience; when a level-k' player randomises, a level-$k'+1$ player best-responds to a level-k' player's true mix.
Salience-pattern origin	
HYP	The salience-pattern is inferred by model-fitting.
PICK	The salience pattern used stems from the PICKING-TASK data. Participants do not know anything about any of the games.
POSTX	The salience pattern used stems from POST-GAME GUESSING after participants have played game X. X can be H&S for the hide-and-seek, COORD for the coordination, and DISCOORD for the discoordination game.
Salience-pattern used	
W[XY]Z	This postfix repeats the salience ranking used in the model. Locations W to Z are ordered by decreasing salience, square brackets indicate indifference. In the example, location W is the most, and Z the least salient location, while X and Y are equally salient locations.
INDL0	This postfix means the model predictions use participants' individual salience measurements as their respective level-0.
AVOID	This additional postfix indicates that players are assumed to use a salience-avoiding level-0. It is used only in two specifications provided in the appendix for completeness (see ftn. 19).
ASYM	This additional postfix indicates that hiders (seekers) are assumed to use a salience-avoiding(-loving) level-0. It is used only in two specifications provided in the appendix for completeness (see ftn. 19).

Table 2: Systemization of the models used in this paper.

measure POSTH&S-$B_{(2)}A_{(3)}\left[A_{(1)}A_{(4)}\right]$: assume hiders will obtain an extra benefit (seekers incur a cost) of e when they choose one of the end locations and a cost (a benefit) of f when they choose $B_{(2)}$. If $A_{(1)}$ and $A_{(4)}$ are jointly least salient (as in POSTH&S-$B_{(2)}A_{(3)}\left[A_{(1)}A_{(4)}\right]$), we should expect $e > 0$ and $f > 0$.[17]

[17]Given CI posit that $A_{(1)}$ and $A_{(4)}$ are jointly *most* salient, they write down the model using $e' \equiv -e$ and expect the maximum-likelihood estimation to yield $e' > 0$. However, given CI do

Table 3 shows the resulting normal form.

		Seeker			
		$\mathbf{A}_{(1)}$	$\mathbf{B}_{(2)}$	$\mathbf{A}_{(3)}$	$\mathbf{A}_{(4)}$
	$\mathbf{A}_{(1)}$	$1-e$ $0+e$	$0+f$ $1+e$	0 $1+e$	$0-e$ $1+e$
Hider	$\mathbf{B}_{(2)}$	$0-e$ $1-f$	$1+f$ $0-f$	0 $1-f$	$0-e$ $1-f$
	$\mathbf{A}_{(3)}$	$0-e$ 1	$0+f$ 1	1 0	$0-e$ 1
	$\mathbf{A}_{(4)}$	$0-e$ $1+e$	$0+f$ $1+e$	0 $1+e$	$1-e$ $0+e$

Table 3: The hide-and-seek game with payoff perturbations when $A_{(1)}$ and $A_{(4)}$ are equally salient (adapted from Crawford and Iriberri, 2007, Figure 2).

Alternatively, we can base the game with payoff perturbations also on PICK-$B_{(2)}\left[A_{(3)}A_{(1)}\right]A_{(4)}$. This specification fits the data worse and is therefore relegated to Table A.1 in the Appendix.

3.4 Level-k model fit under the elicited salience patterns

Basing CI's level-k model on the elicited salience measures PICK-$B_{(2)}\left[A_{(3)}A_{(1)}\right]A_{(4)}$ and POSTH&S-$B_{(2)}A_{(3)}\left[A_{(1)}A_{(4)}\right]$, Table 4 presents players' predicted choices depending on their k-level.

Using the same data as CI, I perform a complete grid search over all possible type-distributions (at the percent level), to find the level-k distribution under which the data has the highest log-likelihood, using equation (2) in CI.[18] Table 5 presents the results. The focal models in Table 5 are CI's preferred level-k model using the empirically-elicited salience patterns as level-0, Lk-PICK-$B_{(2)}\left[A_{(3)}A_{(1)}\right]A_{(4)}$ and Lk-POSTH&S-$B_{(2)}A_{(3)}\left[A_{(1)}A_{(4)}\right]$.[19] For comparison, I include also estimates of the following six benchmark models: choice according to

not put restrictions on the signs of e and f (EQM$_+$-HYP-$\left[A_{(1)}A_{(4)}\right]$), both games are equivalent.

[18]Note that I present this analysis primarily for comparability. By the logic of this paper, I should restrict myself to a small subset of the data: CI use data from 6 different treatments conducted by Rubinstein and co-authors; to use all 6, CI have to make assumptions of how to convert the data from some treatments to make them comparable to the data from others. E.g., data from a treatment using "A", "A", "B", and "A" is adapted by simply switching the two locations in the middle. Whether this is appropriate based on the underlying salience structure is, again, an empirical question. I refer to the analysis of Heinrich and Wolff's (2012) data in the right-hand part of Table 5 for an analysis that does not rely on comparable assumptions.

[19] CI's alternative level-k specifications with an asymmetric level-0 (favouring salience for

k-level (frequency)	box	PICK-$B_{(2)}\left[A_{(3)}A_{(1)}\right]A_{(4)}$ Hider	Seeker	POSTH&S-$B_{(2)}A_{(3)}\left[A_{(1)}A_{(4)}\right]$ Hider	Seeker
L0 ($\pi_0 \equiv 0$)	A$_{(1)}$	0.21	0.21	0.19	0.19
	B$_{(2)}$	0.38	0.38	0.39	0.39
	A$_{(3)}$	0.35	0.35	0.24	0.24
	A$_{(4)}$	0.06	0.06	0.18	0.18
L1 (π_1)	A$_{(1)}$	0	0	1/2	0
	B$_{(2)}$	0	1	0	1
	A$_{(3)}$	0	0	0	0
	A$_{(4)}$	1	0	1/2	0
L2 (π_2)	A$_{(1)}$	1/3	0	1/3	1/2
	B$_{(2)}$	0	0	0	0
	A$_{(3)}$	1/3	0	1/3	0
	A$_{(4)}$	1/3	1	1/3	1/2
L3 (π_3)	A$_{(1)}$	1/3	1/3	0	1/3
	B$_{(2)}$	1/3	0	1/2	0
	A$_{(3)}$	1/3	1/3	1/2	1/3
	A$_{(4)}$	0	1/3	0	1/3
L4 (π_4)	A$_{(1)}$	0	1/3	0	0
	B$_{(2)}$	1	1/3	1	1/2
	A$_{(3)}$	0	1/3	0	1/2
	A$_{(4)}$	0	0	0	0
Total	A$_{(1)}$	$\frac{\pi_2+\pi_3}{3}$	$\frac{\pi_3+\pi_4}{3}$	$\frac{\pi_1}{2}+\frac{\pi_2}{3}$	$\frac{\pi_2}{2}+\frac{\pi_3}{3}$
	B$_{(2)}$	$\frac{\pi_3}{3}+\pi_4$	$\pi_1+\frac{\pi_4}{3}$	$\frac{\pi_3}{2}+\pi_4$	$\pi_1+\frac{\pi_4}{2}$
	A$_{(3)}$	$\frac{\pi_2+\pi_3}{3}$	$\frac{\pi_3+\pi_4}{3}$	$\frac{\pi_2}{3}+\frac{\pi_3}{2}$	$\frac{\pi_3}{3}+\frac{\pi_4}{2}$
	A$_{(4)}$	$\pi_1+\frac{\pi_2}{3}$	$\pi_2+\frac{\pi_3}{3}$	$\frac{\pi_1}{2}+\frac{\pi_2}{3}$	$\frac{\pi_2}{2}+\frac{\pi_3}{3}$

Table 4: Level-k players' hide-and-seek choice probabilities under PICK-$B_{(2)}\left[A_{(3)}A_{(1)}\right]A_{(4)}$ and POSTH&S-$B_{(2)}A_{(3)}\left[A_{(1)}A_{(4)}\right]$. The fraction of players of level i is denoted by π_i.

the empirically-elicited salience patterns (NAÏVE-PICK-$B_{(2)}\left[A_{(3)}A_{(1)}\right]A_{(4)}$, NAÏVE-POSTH&S-$B_{(2)}A_{(3)}\left[A_{(1)}A_{(4)}\right]$); the standard mixed-strategy Nash-equilibrium (EQM$_0$); CI's equilibrium with "unrestricted" payoff perturbations based on $A_{(1)}$ and $A_{(4)}$ being equally salient (EQM$_+$-HYP-$\left[A_{(1)}A_{(4)}\right]$), as well as with 'partially restricted' perturbations (so as to match the elicited salience pattern; EQM$_+$-POSTH&S-$B_{(2)}A_{(3)}\left[A_{(1)}A_{(4)}\right]$);[20] CI's preferred level-k model under their salience assump-

seekers and avoiding it for hiders, Lk-X-ASYM) and with a salience-avoiding level-0 (Lk-X-AVOID) are included in Table A.2 in the Appendix.

[20]The estimates of EQM$_+$-HYP-$\left[A_{(4)}\right]$ (no restriction on the sign of e) and EQM$_+$-PICK-$B_{(2)}\left[A_{(3)}A_{(1)}\right]A_{(4)}$) are also included in Table A.2 in the Appendix. For all EQM$_+$-estimations, I use a two-step procedure: I first do a complete grid search over all four parameters for $-1 \leq e_H, f_H, e_S, f_S \leq 1$ at the five-percent level, and then another one at the percent level for the parameter space $[e_H-0.1, e_H+0.1]\times[f_H-0.1, f_H+0.1]\times[e_S-0.1, e_S+0.1]\times[f_S-0.1, f_S+0.1]$.

Specification	RTH's data logL	RTH's data MSE	HW's data logL	HW's data MSE
Choices follow salience				
NAÏVE-PICK-$B_{(2)}\left[A_{(3)}A_{(1)}\right]A_{(4)}$	-1724	0.01271	-521	0.01654
NAÏVE-POSTH&S-$B_{(2)}A_{(3)}\left[A_{(1)}A_{(4)}\right]$	-1687	0.01647	-487	0.01662
Equilibrium models				
EQM$_0$	-1641†	0.00967†	-484	0.01436
EQM$_+$-HYP-$\left[A_{(1)}A_{(4)}\right]$ ($e_H = -0.29, f_H = 0.25, e_S = -0.15, f_S = 0.15$)†	-1562†	0.00006†	-456	0.00109
EQM$_+$-POSTH&S-$B_{(2)}A_{(3)}\left[A_{(1)}A_{(4)}\right]$ ($e_H = 0.00, f_H = 0.06, e_S = 0.00, f_S = 0.05$)	-1636	0.00909	-483	0.01467
CI's preferred model				
Lk-HYP-$\left[A_{(1)}A_{(4)}\right]B_{(2)}A_{(3)}$	-1564†	0.00027†	-456	0.00110
Lk-PICK-$\mathbf{B_{(2)}\left[A_{(3)}A_{(1)}\right]A_{(4)}}$	**-1635**	**0.00903**	**-482**	**0.01358**
Lk-POSTH&S-$\mathbf{B_{(2)}A_{(3)}\left[A_{(1)}A_{(4)}\right]}$	**-1664**	**0.01202**	**-485**	**0.01538**

Table 5: Log-likelihoods and mean squared errors of the maximum-likelihood estimates of the indicated models. The first two data columns use the data from Rubinstein, Tversky, and Heller's collected studies ("RTH"), reproduced in Table 3 of Crawford and Iriberri (2007; "CI"). Columns three and four replicate the findings using Heinrich and Wolff's (2012; "HW") data. † indicates the estimate is taken from CI's paper.

tion (Lk-HYP-$\left[A_{(1)}A_{(4)}\right]B_{(2)}A_{(3)}$).

Readers may object that the salience-elicitation experiments where all conducted in Germany, and therefore, in a different cultural environment than the actual games. Furthermore, if people from different cultures have different perceptions in terms of salience or if their salience-based strategic reasoning is shaped culturally, we cannot conclude much from elicitating salience in one part of the world to explain behaviour in another. To respond to this valid objection, I also include the model estimates for German hide-and-seek data, taken from a study by Heinrich and Wolff (2012).[21]

Observation 4. 'Random clicking' as shaped by salience patterns does not explain the data well.

Observation 4 rests on the fact that both the log-likelihoods and the mean squared errors of specifications NAÏVE-PICK-$B_{(2)}\left[A_{(3)}A_{(1)}\right]A_{(4)}$ and NAÏVE-POSTH&S-$B_{(2)}A_{(3)}\left[A_{(1)}A_{(4)}\right]$ indicate a fit that is even worse than the equilibrium prediction without payoff perturbations. This is important because naïve, unstrategic responses are one of two explanations for hide-and-seek data in the literature.

[21] For comparability, I include only the data obtained under the original instructions. As pointed out in footnote 18, this data has the additional advantage that it was obtained exclusively under the ABAA-protocol, so that no further assumptions are needed of how to translate salience patterns from other setups, such as the AABA-protocol.

Specification	RTH's data				HW's data			
	***L*1**	***L*2**	***L*3**	***L*4**	***L*1**	***L*2**	***L*3**	***L*4**
Lk-HYP-$\left[A_{(1)}A_{(4)}\right]B_{(2)}A_{(3)}$	0.19†	0.32†	0.24†	0.25†	0.12	0.37	0.29	0.22
Lk-PICK-$B_{(2)}\left[A_{(3)}A_{(1)}\right]A_{(4)}$	0.21	0.00	0.79	0.00	0.23	0.00	0.70	0.07
Lk-POSTH&S-$B_{(2)}A_{(3)}\left[A_{(1)}A_{(4)}\right]$	0.31	0.00	0.69	0.00	0.32	0.00	0.59	0.09

Table 6: Level-k distributions of the maximum-likelihood estimates in Table 5. The first four data columns use the data from Rubinstein, Tversky, and Heller's collected studies ("RTH"), reproduced in Table 3 of Crawford and Iriberri (2007; "CI"). Columns five to eight replicate the findings using Heinrich and Wolff's (2012; "HW") data. † indicates the estimate is taken from CI's paper.

Main Result 1. Using the same data as Crawford and Iriberri (2007), measured-salience-based estimates for their preferred level-k model fit the data clearly worse than the estimates they derive for an equilibrium model with 'unrestricted' payoff perturbations. The better-fitting Specification Lk-PICK-$B_{(2)}\left[A_{(3)}A_{(1)}\right]A_{(4)}$ hardly outperforms even the mixed-strategy Nash-equilibrium prediction, despite its higher number of free parameters.

Main Result 1 can be verified by a look at the Table-5 columns reporting the log-likelihoods, comparing specification Lk-PICK-$B_{(2)}\left[A_{(3)}A_{(1)}\right]A_{(4)}$ to specifications EQM$_+$-HYP-$\left[A_{(1)}A_{(4)}\right]$ and EQM$_0$, respectively.[22] Note that it does not depend on the level-k distribution we use—that is, it holds even for the estimates yielding the highest likelihoods. These distributions are depicted in Table 6.

Observation 5. Maximum-likelihood estimates of both elicited-salience-based variants of Crawford and Iriberri's (2007) preferred model are implausible, exhibiting a zero fraction of Level-2 players in conjunction with fractions of Level-1 and Level-3 players that (virtually) sum up to 1.

Observation 5 follows from the rows corresponding to specifications Lk-PICK-$B_{(2)}\left[A_{(3)}A_{(1)}\right]A_{(4)}$ and Lk-POSTH&S-$B_{(2)}A_{(3)}\left[A_{(1)}A_{(4)}\right]$ in Table 6. As was already stated, "[t]he estimated distribution tends to be stable across games and *hump-shaped*," (CI, p. 1734, emphasis added) which renders the estimates implausible.

[22]The first part holds also for all other level-k variants presented in CI: both Lk-POSTH&S-$B_{(2)}A_{(3)}\left[A_{(1)}A_{(4)}\right]$-ASYM and -AVOID exhibit log-likelihoods of -1603 (RTH's data) and -465 (HW's data), cf. Table A.2 in the Appendix. Further, it holds for an alternative Lk-PICK-$\left[B_{(2)}A_{(3)}\right]A_{(1)}A_{(4)}$ specification that Shaun Hargreaves Heap, David Rojo Arjona and Robert Sugden have suggested would fit the picking-task data better (logL of -1643, RTH's data, and of -471, HW's data). I am not presenting this alternative in the main text because the measured reaction times in my view suggest $B_{(2)}$ and $A_{(3)}$ are salient to different degrees, as does the ALGORITHM data.

Finally, note that CI's estimation also of the equilibrium with perturbations ($\textsc{eqm}_{+}$-$\textsc{hyp}$-$\left[A_{(1)}A_{(4)}\right]$) suggests a salience pattern that exhibits $A_{(3)}$ as the least salient alternative ($e_H, e_S < 0, f_H, f_E > 0$). This would imply that either all of our empirical estimates of salience are wrong or the benchmark equilibrium model rests on implausible mechanics, too.

Observation 6. Estimates for the equilibrium models with payoff perturbations under the constraint that the payoff perturbations follow one of the elicited salience-pattern candidates have a similarly bad model fit as the re-estimated CI models.

This observation follows from looking at the row in Table 5 pertaining to specification $\textsc{eqm}_{+}$-$\textsc{post}$H&S-$B_{(2)}A_{(3)}\left[A_{(1)}A_{(4)}\right]$, and comparing the log-likelihood to those of the specifications Lk-$\textsc{pick}$-$B_{(2)}\left[A_{(3)}A_{(1)}\right]A_{(4)}$ and Lk-$\textsc{post}$H&S-$B_{(2)}A_{(3)}\left[A_{(1)}A_{(4)}\right]$.[23] Note that Observations 4-6 also hold for the data from Heinrich and Wolff (2012), as can be verified by looking at the respective columns in the right-hand part of Tables 5 and 6. The similarity of the estimated parameters and of the models' relative likelihoods suggest that the hide-and-seek game is played in a similar fashion in Stanford, Tel Aviv, and Konstanz.

3.5 A modified level-k model based on elicited salience patterns

In section 3.1, I reported empirical measures of a salience-based level-0; in section 3.4, I established that basing the model proposed by CI on these empirically-elicited salience patterns leads to implausible model estimates with a poor data fit—an assertion that holds true also for their benchmark equilibrium model with payoff perturbations. In this section, I will argue that a simple modification of CI's model restitutes the notably good fit to the data reported in their paper. This modification assumes that a player who should be indifferent between multiple actions randomises over these actions according to their relative salience rather than randomising uniformly. This may happen for a variety of different reasons: different people might be inherently attracted to different locations when no compelling economic force acts on them; they might decide to choose 'just anything' from among the options they are indifferent about, in a similar fashion as participants in our Picking Task will have chosen one of the boxes when there was no reason to favour any box over the other; or they might try to randomise uniformly, but the attraction exerted by salience might unconsciously interfere with their randomisation attempts.

[23] The same holds true for $\textsc{eqm}_{+}$-$\textsc{pick}$-$B_{(2)}\left[A_{(3)}A_{(1)}\right]A_{(4)}$, cf. Table A.2 in the Appendix.

Given the above argument, there are two different ways to implement the idea. The simpler variant (Lk_{UNSOPH}) corresponds to the idea that higher-level players are unaware of salience-influences on randomisation by lower-level players. Hence, if a level-i player randomises, a level-$(i+1)$ player (wrongly) assumes the level-i player is randomising uniformly. The obvious alternative is that the level-$(i+1)$ player is aware of the randomising level-i player's inability not to be attracted by salience, and best-responds to the resulting probability distribution (Lk_{SOPH}). Table 7 presents the resulting predictions.

On the basis of the predictions from Table 7, I estimate the modified level-k model using level-0 specifications PICK-$B_{(2)}\left[A_{(3)}A_{(1)}\right]A_{(4)}$ and POSTH&S-$B_{(2)}A_{(3)}\left[A_{(1)}A_{(4)}\right]$. This time, I do not restrict the fraction of level-0 players to be 0, for two reasons. First, I want to show for which model variants the assumption is binding. And second, to be able to estimate specification Lk_{SOPH}-PICK-$B_{(2)}\left[A_{(3)}A_{(1)}\right]A_{(4)}$, we need to include errors of some form.[24] Arguably, if errors correspond to randomly picking an action, the salience pattern—which corresponds to level-0 in the model—constitutes a plausible error specification. In this view, the estimated fraction of level-0 players is a measure for the frequency of errors.[25] Table 8 reports the resulting model fits to the data.

Main Result 2. The modified level-k model is able to fit the data substantially better compared to Crawford and Iriberri's (2007) level-k variant when the latter also is based on empirically-elicited salience patterns. The best fit to the data—being almost as good as the fit of the equilibrium with unrestricted perturbations—is achieved by the model in which higher levels are aware of the influence of salience on randomising players and level-0 is given by the data from the POST-GAME GUESSING task.

The first claim rests on a comparison of the log-likelihoods of Lk_{SOPH}-POSTH&S-$B_{(2)}A_{(3)}\left[A_{(1)}A_{(4)}\right]$ in Table 8 to those of Lk-PICK-$B_{(2)}\left[A_{(3)}A_{(1)}\right]A_{(4)}$ and Lk-POSTH&S-$B_{(2)}A_{(3)}\left[A_{(1)}A_{(4)}\right]$ in Table 5.[26] For the second, note that the log-likelihood of Lk_{SOPH}-POSTH&S-$B_{(2)}A_{(3)}\left[A_{(1)}A_{(4)}\right]$ in Table 8 is very close to that of EQM$_{+}$-HYP-$\left[A_{(1)}A_{(4)}\right]$ in the same table and that while Lk_{SOPH}-PICK-$B_{(2)}\left[A_{(3)}A_{(1)}\right]A_{(4)}$ (with one additional parameter) has a slightly better fit to Heinrich and Wolff's

[24]Otherwise, the log-likelihood function would always yield $-\infty$, making it impossible to detect the best-fitting level-k distributions.

[25]One obvious alternative error specification would be to assume players choose any location with equal probabilities whenever they make an error. However, it is completely unclear to me what kind of errors would lead to a uniform error structure: e.g., mis-clicks should be more likely to end up at the immediately adjacent locations, in which case the error structure should be hump-shaped.

[26]The claim also holds true with respect to CI's other level-k variants, cf. Table A.2 in the Appendix.

k-level (frequency)	box	Lk_{unsoph}-pick-$B_{(2)}\left[A_{(3)}A_{(1)}\right]A_{(4)}$ Hider	Seeker	Lk_{unsoph}-postH&S-$B_{(2)}A_{(3)}\left[A_{(1)}A_{(4)}\right]$ Hider	Seeker
L0 (π_0)	A	0.21	0.21	0.19	0.19
	B	0.38	0.38	0.39	0.39
	A	0.35	0.35	0.24	0.24
	A	0.06	0.06	0.18	0.18
L1 (π_1)	A	0.00	0.00	0.51	0.00
	B	0.00	1.00	0.00	1.00
	A	0.00	0.00	0.00	0.00
	A	1.00	0.00	0.49	0.00
L2 (π_2)	A	0.35	0.00	0.31	0.51
	B	0.00	0.00	0.00	0.00
	A	0.56	0.00	0.39	0.00
	A	0.10	1.00	0.30	0.49
L3 (π_3)	A	0.23	0.35	0.00	0.31
	B	0.40	0.00	0.62	0.00
	A	0.37	0.56	0.38	0.39
	A	0.00	0.10	0.00	0.30
L4 (π_4)	A	0.00	0.23	0.00	0.00
	B	1.00	0.40	1.00	0.62
	A	0.00	0.37	0.00	0.38
	A	0.00	0.00	0.00	0.00
Total	A	$0.21\pi_0 + 0.35\pi_2 + 0.23\pi_3$	$0.21r + 0.35\pi_3 + 0.23\pi_4$	$0.19\pi_0 + 0.51\pi_1 + 0.31\pi_2$	$0.19\pi_0 + 0.51\pi_2 + 0.31\pi_3$
	B	$0.38\pi_0 + 0.4\pi_3 + \pi_4$	$0.38\pi_0 + 0.4v + \pi_1$	$0.39\pi_0 + 0.62\pi_3 + \pi_4$	$0.39\pi_0 + 0.62v + \pi_1$
	A	$0.35\pi_0 + 0.56\pi_2 + 0.37\pi_3$	$0.35\pi_0 + 0.56\pi_3 + 0.37\pi_4$	$0.24\pi_0 + 0.39\pi_2 + 0.38\pi_3$	$0.24\pi_0 + 0.39\pi_3 + 0.38\pi_4$
	A	$0.06\pi_0 + \pi_1 + 0.1\pi_2$	$0.06\pi_0 + \pi_2 + 0.1\pi_3$	$0.18\pi_0 + 0.49\pi_1 + 0.3\pi_2$	$0.18\pi_0 + 0.49\pi_2 + 0.3\pi_3$
k-level (frequency)	box	Lk_{soph}-pick-$B_{(2)}\left[A_{(3)}A_{(1)}\right]A_{(4)}$ Hider	Seeker	Lk_{soph}-postH&S-$B_{(2)}A_{(3)}\left[A_{(1)}A_{(4)}\right]$ Hider	Seeker
L0 (π_0)	A	0.21	0.21	0.19	0.19
	B	0.38	0.38	0.39	0.39
	A	0.35	0.35	0.24	0.24
	A	0.06	0.06	0.18	0.18
L1 (π_1)	A	0.00	0.00	0.51	0.00
	B	0.00	1.00	0.00	1.00
	A	0.00	0.00	0.00	0.00
	A	1.00	0.00	0.49	0.00
L2 (π_2)	A	0.35	0.00	0.31	0.51
	B	0.00	0.00	0.00	0.00
	A	0.56	0.00	0.39	0.00
	A	0.10	1.00	0.30	0.49
L3 (π_3)	A	0.23	0.00	0.00	0.00
	B	0.40	0.00	0.62	0.00
	A	0.37	1.00	0.38	1.00
	A	0.00	0.00	0.00	0.00
L4 (π_4)	A	0.33	0.00	0.25	0.00
	B	0.58	1.00	0.51	1.00
	A	0.00	0.00	0.00	0.00
	A	0.09	0.00	0.24	0.00
Total	A	$0.21\pi_0 + 0.35\pi_2 + 0.23\pi_3 + 0.33\pi_4$	$0.21\pi_0$	$0.19\pi_0 + 0.51\pi_1 + 0.31\pi_2 + 0.25\pi_4$	$0.19\pi_0 + 0.51\pi_2$
	B	$0.38\pi_0 + 0.4\pi_3 + 0.58\pi_4$	$0.38\pi_0 + \pi_1 + \pi_4$	$0.39\pi_0 + 0.62\pi_3 + 0.51\pi_4$	$0.39\pi_0 + \pi_1 + \pi_4$
	A	$0.35\pi_0 + 0.56\pi_2 + 0.37\pi_3$	$0.35\pi_0 + \pi_3$	$0.24\pi_0 + 0.39\pi_2 + 0.38\pi_3$	$0.24\pi_0 + \pi_3$
	A	$0.06\pi_0 + \pi_1 + 0.1\pi_2 + 0.09\pi_4$	$0.06\pi_0 + \pi_2$	$0.18\pi_0 + 0.49\pi_1 + 0.3\pi_2 + 0.24\pi_4$	$0.18\pi_0 + 0.49\pi_2$

Table 7: Types' hide-and-seek choice probabilities when salience follows pick-$B_{(2)}\left[A_{(3)}A_{(1)}\right]A_{(4)}$ and postH&S-$B_{(2)}A_{(3)}\left[A_{(1)}A_{(4)}\right]$ (left- and right-hand halves of the Table, respectively), and when higher-level players are unsophisticated and sophisticated (upper and lower half of the Table, respectively).

Specification	RTH's data logL	RTH's data MSE	HW's data logL	HW's data MSE
Equilibrium models				
EQM$_0$	-1641*	0.00967*	-484	0.01436
EQM$_+$-HYP-$\left[A_{(1)}A_{(4)}\right]$ ($e_H = 0.29$, $f_H = 0.25$, $e_S = 0.15$, $f_S = 0.15$)	-1562*	0.00006*	-456	0.00109
CI's preferred model				
Lk-HYP-$\left[A_{(1)}A_{(4)}\right]B_{(2)}A_{(3)}$	-1564*	0.00027*	-456	0.00110
Modified level-k models				
Lk_{UNSOPH}-PICK-$B_{(2)}\left[A_{(3)}A_{(1)}\right]A_{(4)}$	-1578	0.00201	-463	0.00509
best 'hump-shaped' type distribution	-1584	0.00263	-466	0.00642
Lk_{UNSOPH}-POSTH&S-$B_{(2)}A_{(3)}\left[A_{(1)}A_{(4)}\right]$	-1666	0.01286	-484	0.01437
best 'hump-shaped' type distribution‡	-1668	0.01616	-485	0.01735
Lk_{SOPH}-PICK-$B_{(2)}\left[A_{(3)}A_{(1)}\right]A_{(4)}$	-1597	0.00310	-457	0.00152
best 'hump-shaped' type distribution			-458	0.00166
Lk_{SOPH}-POSTH&S-$B_{(2)}A_{(3)}\left[A_{(1)}A_{(4)}\right]$	-1570	0.00097	-458	0.00143

†As CI do not rely on any specific $L0$-pattern, it is not clear to me how to endogenise π_0 without including yet another two parameters (p and q in their paper). As I am reluctant to do so, I maintain $\pi_0 \equiv 0$ for their preferred model. ‡I do not require $\pi_0 < \pi_1$ for a 'hump-shaped' pattern, given 'random-clicking' without further strategic deliberation is one of the two explanations for hide-and-seek behaviour in the literature, and thus may co-exist with level-k behaviour. *indicates figures taken from CI's paper.

Table 8: Log-likelihoods for the leading models in CI (first three specifications) and the modified level-k models assuming that players follow salience when indifferent. Variants indicated by "best hump-shaped distribution" are estimated under the constraint that there may not be fewer level-2 players than level-1 players. Columns three and four replicate the findings using Heinrich and Wolff's (2012; "HW") data.

data, it has a substantially worse fit to Rubinstein, Tversky and Heller's original data. Furthermore, the estimated level-distributions in Table 9 indicates that the best-fitting model does not exhibit the U-shaped levels distribution of the original level-k variants based on PICK-$B_{(2)}\left[A_{(3)}A_{(1)}\right]A_{(4)}$ or POSTH&S-$B_{(2)}A_{(3)}\left[A_{(1)}A_{(4)}\right]$, nor Crawford and Iriberri's uncommonly high estimates of level-4 prevalence.

Observation 7. The model estimations do not indicate a large fraction of non-strategic behaviour.

For an indicator of non-strategic behaviour in the sense of 'random-clicking', I use the estimated proportions of level-0 play given in Table 9, as the former will be shaped by salience and the latter is defined as following salience. Among the models not restricted to a hump-shaped level distribution, only Lk_{SOPH}-PICK-$B_{(2)}\left[A_{(3)}A_{(1)}\right]A_{(4)}$ indicates a substantial fraction of level-0 play. However, as pointed out above, this model fits the original data of Rubinstein, Tversky and Heller only unsatisfactorily. The same applies to the restricted model Lk_{UNSOPH}-POSTH&S-$B_{(2)}A_{(3)}\left[A_{(1)}A_{(4)}\right]$ (which does not fit Heinrich and Wolff's data well,

Specification	RTH's data $L0$	$L1$	$L2$	$L3$	$L4$	HW's data $L0$	$L1$	$L2$	$L3$	$L4$
Lk-HYP-$[A_{(1)}A_{(4)}]B_{(2)}A_{(3)}$	0.00†	0.19†	0.32†	0.24†	0.25†	–	0.12	0.37	0.29	0.22
Lk_{UNSOPH}-PICK-$B_{(2)}[A_{(3)}A_{(1)}]A_{(4)}$	0.00	0.20	0.11	0.69	0.00	0.00	0.23	0.11	0.63	0.03
best 'hump-shaped' distribution	0.10	0.16	0.16	0.58	0.00	0.02	0.18	0.18	0.53	0.09
Lk_{UNSOPH}-POSTH&S-$B_{(2)}A_{(3)}[A_{(1)}A_{(4)}]$	0.00	0.28	0.10	0.62	0.00	0.08	0.30	0.00	0.62	0.00
best 'hump-shaped' distribution	0.40	0.12	0.12	0.36	0.07	0.78	0.02	0.02	0.18	0.07
Lk_{SOPH}-PICK-$B_{(2)}[A_{(3)}A_{(1)}]A_{(4)}$	0.48	0.12	0.14	0.26	0.00	0.30	0.18	0.14	0.38	0.00
best 'hump-shaped' distribution						0.26	0.18	0.18	0.38	0.00
Lk_{SOPH}-POSTH&S-$B_{(2)}A_{(3)}[A_{(1)}A_{(4)}]$	0.00	0.20	0.37	0.43	0.00	0.00	0.23	0.28	0.49	0.00

Table 9: Level-k distributions of the maximum-likelihood estimates in Table 8. The first four data columns use the data from Rubinstein, Tversky, and Heller's collected studies ("RTH"), reproduced in CI's Table 3. Columns five to eight replicate the findings using Heinrich and Wolff's (2012; "HW") data. † indicates the estimate is taken from CI's paper.

either). Hence, the models either do not indicate a large fraction of non-strategic behaviour or they fit the data poorly, in which case we should doubt their validity as an indicator of behaviour.

3.6 Overfitting

Like Crawford and Iriberri (2007), I want to assess the models' ability to predict out of sample rather than their flexibility in fitting the data. Following their procedure, I estimate each model on each study to 'predict' the data of the other studies, using the resulting average MSEs as a criterion.[27] However, by the logic of this paper, I cannot use data from all of Rubinstein, Tversky, and Heller's treatments as done by Crawford and Iriberri: a hide-and-seek game using "1", "2", "3", and "4" as locations may have a completely different salience structure than the archetype ABAA-protocol I have focused on. Hence, I only use data from the three studies employing the ABAA-protocol: the data from the corresponding treatment in Rubinstein, Tversky and Heller (1996), the data from Rubinstein (1999), and the data from Heinrich and Wolff (2012). Furthermore, I only include five models in the analysis. These are EQM$_+$-HYP-$[A_{(1)}A_{(4)}]$, CI's preferred model Lk-HYP-$[A_{(1)}A_{(4)}]B_{(2)}A_{(3)}$, and the three level-$k$ models with the best model fit under a salience-seeking $L0$ that follows an empirically-elicited salience pattern: Lk-PICK-$B_{(2)}[A_{(3)}A_{(1)}]A_{(4)}$, L$k_{\text{UNSOPH}}$-PICK-$B_{(2)}[A_{(3)}A_{(1)}]A_{(4)}$, and L$k_{\text{SOPH}}$-POSTH&S-$B_{(2)}A_{(3)}[A_{(1)}A_{(4)}]$. Table 10 presents the results.

[27]Note that the above MSEs are not comparable to those in Tables 5 and 8: the former are prediction errors of an out-of-sample prediction, the latter were merely a goodness-of-fit measure.

	$\mathbf{e_H}$ resp. **L1**	$\mathbf{f_H}$ resp. **L2**	$\mathbf{e_S}$ resp. **L3**	$\mathbf{f_S}$ resp. **L4**	**RTH-4**	**R-ABAA**	**HW**	**Average MSE**
EQM$_+$-HYP-$\left[A_{(1)}A_{(4)}\right]$: **overall MSE = 0.00637**								
RTH-4	0.33	0.14	0.27	0.03	–	0.00904	0.00598	0.00751
R-ABAA	0.41	0.36	0.25	0.26	0.00856	–	0.00526	0.00691
HW	0.38	0.29	0.1	0.02	0.00501	0.00438	–	0.00469
Lk-HYP-$\left[A_{(1)}A_{(4)}\right]B_{(2)}A_{(3)}$: **overall MSE = 0.00364**								
RTH-4	0.25	0.27	0.48	0	–	0.00315	0.00250	0.00283
R-ABAA	0.2	0.37	0.26	0.17	0.00418	–	0.00251	0.00334
HW	0.12	0.37	0.29	0.22	0.00527	0.00425	–	0.00476
Lk-PICK-$B_{(2)}\left[A_{(3)}A_{(1)}\right]A_{(4)}$: **overall MSE = 0.01873**								
RTH-4	0.19	0	0.61	0.2	–	0.02648	0.01516	0.02082
R-ABAA	0.2	0	0.8	0	0.01818	–	0.01448	0.01633
HW	0.23	0	0.69	0.08	0.01564	0.02245	–	0.01905
Lk_{UNSOPH}-PICK-$B_{(2)}\left[A_{(3)}A_{(1)}\right]A_{(4)}$: **overall MSE = 0.00814**								
RTH-4	0.19	0.07	0.6	0.14	–	0.01361	0.00671	0.01016
R-ABAA	0.2	0.07	0.73	0	0.00666	–	0.00519	0.00593
HW	0.23	0.11	0.63	0.03	0.00699	0.00969	–	0.00834
Lk_{SOPH}-POSTH&S-$B_{(2)}A_{(3)}\left[A_{(1)}A_{(4)}\right]$: **overall MSE = 0.00566**								
RTH-4	0.01	0.26	0.48	0.25	–	0.01264	0.00487	0.00876
R-ABAA	0.19	0.31	0.5	0	0.00484	–	0.00167	0.00326
HW	0.23	0.28	0.49	0	0.00440	0.00554	–	0.00497

Table 10: Mean squared prediction errors (MSE) for the studies indicated in the columns that result when parameters are fitted to the study indicated in the row. RTH-4 refers to Rubinstein, Tversky and Heller's (1996), R-ABAA to Rubinstein's (1999) and HW to Heinrich and Wolff's (2012) ABAA-treatment.

Observation 8. The best-fitting model that is based on an empirically-elicited salience pattern yields better out-of-sample predictions than the equilibrium model with hard-wired payoff perturbations as well as the other level-k models based on empirically-elicited salience patterns. Surprisingly, it is outperformed clearly by Crawford and Iriberri's preferred level-k model Lk-HYP-$\left[A_{(1)}A_{(4)}\right]B_{(2)}A_{(3)}$ that relies on a salience pattern that is not in accordance with any of the empirical measures of salience.

Observation 8 rests on a comparison of the overall mean squared prediction errors reported in bold in Table 10.

3.7 Portability

Much like Crawford and Iriberri (2007), I next look at whether we can use the models estimated before to predict behaviour in another game. However, following the critique by Hargreaves Heap, Rojo Arjona and Sugden (2014), I use a coordination and a discoordination game played on an 'A-B-A-A' action set that was presented exactly like the one in the hide-and-seek game. Hargreaves Heap, Rojo Arjona and Sugden show that it is not possible to find a common level-0 that would allow to predict behaviour in these three games simultaneously. Given that the game description seems to influence participants' perceptions of salience (Observation 3), allowing for a different level-0 in each game seems to be acceptable—as long as this is done in an objective manner. For this purpose, we repeat the GUESSING TASK after participants have played the coordination (POSTCOORD TASK) or discoordination game (POSTDISCOORD TASK), respectively, without feedback.[28] Table 11 shows the results.

	$A_{(1)}$	$B_{(2)}$	$A_{(3)}$	$A_{(4)}$
POSTCOORD TASK (72 participants)				
average estimated relative click frequency	19	50	18	14
POSTDISCOORD TASK (72 participants)				
average estimated relative click frequency	20	37	24	19

Table 11: Salience assessments of the four boxes denoted by "A", "B", "A", and "A". The POSTCOORD TASK is the GUESSING TASK after participants played the coordination game, the POSTDISCOORD TASK the same task after participants played the discoordination game, again without feedback.

Observation 9. The POSTCOORD and POSTDISCOORD TASKS confirm Observation 3, in that the game description changes participants' estimate of what people in a non-strategic situation will regard as salient.

Observation 9 becomes obvious from Table 11 by focusing on the estimates for locations $B_{(2)}$ and $A_{(4)}$. While in the POSTDISCOORD TASK, the latter is virtually identical to players' estimate on $A_{(1)}$, there is a clear difference in the POSTCOORD TASK.[29] At the same time, the average estimate on $B_{(2)}$ is clearly higher in the POSTCOORD TASK compared to the POSTDISCOORD TASK.[30]

Table 12 reports the choices in the coordination and discoordination games proper, next to the predictions of Crawford and Iriberri's preferred model Lk-HYP-$\left[A_{(1)}A_{(4)}\right]B_{(2)}A_{(3)}$ and the modified models with 'sophisticated' higher-level players based on the POSTCOORD-TASK and POSTDISCOORD-TASK data, respectively. With respect to the POSTDISCOORD TASK, I interpret the data such that

[28] All procedures as in POST-GAME GUESSING. No participant had participated in any other of the experiments described in this paper.

[29] Two-sided Wilcoxon matched-pairs signed-ranks tests yield $p = 0.174$ and $p = 0.001$, respectively.

[30] A two-sided Wilcoxon Mann-Whitney test yields $p < 0.001$.

	$A_{(1)}$	$B_{(2)}$	$A_{(3)}$	$A_{(4)}$	MSE
COORDINATION GAME (72 participants)					
Choices (in %)	18	69	11	1	
Lk-HYP-$\left[A_{(1)}A_{(4)}\right]B_{(2)}A_{(3)}$					
Prediction without errors	50	0	0	50	0.20768
L$k_{\text{(UN)SOPH}}$-POSTCOORD-$B_{(2)}\left[A_{(1)}A_{(3)}\right]A_{(4)}$					
Prediction without errors	0	100	0	0	0.03518
Lk_{SOPH}-POSTCOORD-INDL0 (ML estimate)					
Prediction without errors	10	78	9	2	0.00381
DISCOORDINATION GAME (72 participants)					
Choices (in %)	15	26	40	18	
Lk-HYP-$\left[A_{(1)}A_{(4)}\right]B_{(2)}A_{(3)}$					
Prediction without errors	28	22	22	28	0.01523
L$k_{\text{(UN)SOPH}}$-POSTDISCOORD-$B_{(2)}A_{(3)}\left[A_{(1)}A_{(4)}\right]$					
Prediction without errors	14	44	28	14	0.01213
Lk_{SOPH}-POSTDISCOORD-INDL0 (ML estimate)					
Prediction without errors	20	37	23	19	0.01076

Table 12: Choices in and predictions for the coordination and discoordination games. The predictions rest on the estimated fractions of level-k types reported in Table 8. For the modified model allowing for a heterogeneous level-0 (-INDL0), I report the maximum-likelihood estimate.

$A_{(1)}$ and $A_{(4)}$ are equally salient. Both predictions use the corresponding estimated level-k distribution from the hide-and-seek game.[31]

Main Result 3. Neither Crawford and Iriberri's (2007) preferred model nor the modified variant proposed in this paper predicts well the coordination-game and discoordination-game data. The former predicts poorly even qualitatively in both the coordination and the discoordination games, while the latter predicts the qualitative pattern in the coordination game but not in the discoordination game.

To see this, note that the modal choice in the coordination game is $B_{(2)}$, while CI's preferred model Lk-HYP-$\left[A_{(1)}A_{(4)}\right]B_{(2)}A_{(3)}$ would predict that only players making an error would choose this location. Similarly, Lk-HYP-$\left[A_{(1)}A_{(4)}\right]B_{(2)}A_{(3)}$ with a level-k distribution as estimated in the hide-and-seek game predicts that in the discoordination game, most participants choose $A_{(1)}$ and $A_{(4)}$, when in fact, the modal choice is $A_{(3)}$ (and $A_{(1)}$ and $A_{(4)}$ are chosen least often). The

[31]For the modified level-k model, I use the estimate based on Heinrich and Wolff's (2012) data because that data comes exclusively from the A-B-A-A setup.

modified model predicts too concentrated a choice distribution in the coordination game (all choices vs. 69% on $B_{(2)}$), and the wrong modal choice in the discoordination game (44% on $B_{(2)}$ vs. 40% on $A_{(3)}$, while correctly predicting low choice-frequencies for $A_{(1)}$ and $A_{(4)}$).

Main Result 3 establishes that none of the level-k models can account for the data of all three experiments even when we acknowledge that the game descriptions change participants' salience perceptions. Note, however, that up to now, we stuck to the assumption that all players have the same idea of what is salient. This is empirically wrong. Looking at the postDiscoord Task as an example, only 60% of the participants estimate that $B_{(2)}$ is clicked on most often in the Picking Task, followed by 19% for $A_{(3)}$ and 10% for each $A_{(1)}$ and $A_{(4)}$. Taking this seriously calls for a model that allows every player to have their own level-0. I look at this possibility exemplarily for the coordination and discoordination games in the following section.

3.8 Salience-based level-k with a heterogeneous level-0

In the final lines of each part in Table 12, I report maximum-likelihood estimates for the modified level-k model when participants use their respective own individual postCoord/postDiscoord Task responses as level-0.[32] To give an example, assume that a participant in the postDiscoord Task estimates responses in the Picking Task to follow the distribution 10%, 45%, 15%, and 30% for $A_{(1)}$, $B_{(2)}$, $A_{(3)}$, and $A_{(4)}$, respectively. In that case, the prediction for the participant's behaviour in the discoordination game would be that she chooses $A_{(1)}$ for certain in case she is level-1 or level-3, and that she chooses $B_{(2)}$, $A_{(3)}$, and $A_{(4)}$ with probabilities $1/2$, $1/6$, and $1/3$, respectively, in case she is level-2 or level-4. The maximum-likelihood estimate for the discoordination game yields a combined 13% of level-1 and level-3 players and a combined 87% of level-2 and level-4 players.[33]

Main Result 4. Basing the level-k model on individual post-game guessing-task estimates as $L0$ improves the model's fit to the coordination-game data. Yet, the maximum-likelihood estimate even of this model does not produce a prediction that would capture the essential features of the discoordination-game data.

The first part of Main Result 4 results from comparison of mean squared prediction errors in the last column of Table 12. Notwithstanding, a χ^2-test on the

[32]Note that for the coordination game, L$k_{(\text{UN})\text{SOPH}}$-postCoord-indL0 makes the same predictions as a Lk-postCoord-indL0 model. For the discoordination game, the two differ only slightly (and not at all in terms of the estimated level-k distributions). The 'sophisticated' and the 'unconscious' modified models make the same predictions in both games.

[33]In the coordination game, all level-k distributions yield the same likelihood given all levels act in the same way.

data under the hypothesis that the data stems from the model's predicted distribution still yields $p \approx 0.14$. The second part results from the fact that the predicted choice distribution also of the model Lk_{SOPH}-POSTCOORD-INDL0 predicts the modal choice to be $B_{(2)}$ rather than $A_{(3)}$. The according χ^2-test yields $p = 0.008$.

4 Summary and discussion

The data gathered by Rubinstein, Tversky and Heller in their hide-and-seek experiments pose a serious challenge to Nash-equilibrium as a descriptive theory of behaviour. Up to today, two explanations have been proposed. Rubinstein and Tversky (1993; p. 402) claim that participants would fail to reason strategically and "[employ] a naïve strategy (avoiding the endpoints), that is not guided by valid strategic reasoning." If that were so, we would expect them to choose similarly to what they would pick if they had to click on any of the boxes without there being a game: according to the options' salience. **Observation 4** establishes that salience-clicking is a bad predictor for aggregate behaviour in hide-and-seek games, and **Observation 7** suggests that this is likely to be true also for virtually all participants individually.

A second explanation has been proposed by Crawford and Iriberri (2007). They propose a level-k model that is based on salience and convert Rubinstein, Tversky and Heller's account of what is salient into latent model parameters. They estimate the qualitative salience pattern to be such that "central A" is the least salient location, followed by "B", leaving the two "end As" as the most salient locations. Based on this salience pattern, Crawford and Iriberri (2007) present a model that fits the data almost as well as a benchmark model based on hard-wired payoff perturbations, and that outperforms any other model they study in terms of out-of-sample predictions. However, in all five empirical salience measures for the ABAA-setup I elicited, "B" turned out as the most salient (**Observation 1**) and "final A" as the least salient location (possibly in conjunction with "first A"; **Observation 2**). The natural question to be answered was then whether a level-k model based on the empirically-elicited salience patterns would do equally well as a descriptor/predictor of behaviour. **Main Result 1** and **Observation 5** clearly show this is not the case, and that the best-fitting estimates, on top of having a poor fit to the data, exhibit level-k type distributions that are implausible. Coincidentally, the benchmark equilibrium models with salience-based payoff perturbations fit the data similarly badly when based on the empirically-elicited salience patterns (**Observation 6**).

Does this mean a level-k model cannot be used to account for behaviour in hide-and-seek games at all? The answer is no. By modifying the model such that a level-k player will choose according to the salience of a location (consciously or not) whenever the player would be indifferent under pure payoff considerations, I obtain a model that has a fit of similar order as the best-performing models presented by Crawford and Iriberri (2007; **Main Result 2**). Furthermore, this model outperforms all other empirical-salience-based models in terms of its fit to the data.

Beyond any doubt, a model's usefulness is determined by how well the model can predict data rather than how well it does in *ex-post* fitting exercises. Because of that, Crawford and Iriberri (2007) perform tests of how well their model does in out-of-sample as well as in out-of-game predictions. For the latter, they use data from two additional games that have different action sets—and therefore, potentially also a different salience structure. Hence, they need additional assumptions on what the salience structure will be in those games. Hargreaves Heap, Rojo Arjona and Sugden (2014) avoid this problem by using data from a coordination and a discoordination games on the same action-set frames as in the hide-and-seek game. Using this approach, they show at a general level that calibrating salience and the type distributions on the hide-and-seek game data does not allow to predict behaviour in their coordination and discoordination games.

Do Hargreaves Heap, Rojo Arjona and Sugden's findings mean level-k cannot explain behaviour from the different laboratory experiments? Not necessarily, as **Observations 3** and **9** suggest that players' salience perception may change across different games with the same action-set frames. Interpreted in this way, my results suggest that not being able to account for data from three different games when assuming the same level-0 could constitute a solvable problem for the theory. However, **Main Result 3** establishes that the conclusion Hargreaves Heap, Rojo Arjona and Sugden draw is correct even if we do account for a changing level-0: in out-of-game predictions, Crawford and Iriberri's preferred model does badly. When I subject the modified level-k model based on empirical salience measures to the same test, it does slightly better, but it clearly fails to explain the data in the discoordination game (**Main Result 3**).

Up to this point, all authors including myself have assumed that at least within each game, there is a unique level-0 that is the same for all players.[34] A closer look at the data from the experiments presented in this study reveals this assumption is unwarranted, too. Data from guessing-task experiments conducted after the different games with no feedback in between are by no means homogeneous in terms of what participants expect others to choose in a Picking Task. Therefore, we need to analyse behaviour taking into account het-

[34] A notable exception is Burchardi and Penczynski (2014), who allow for different guesses about the behavior of non-strategically-acting players in a beauty-contest game.

erogeneous salience perceptions. I do this exemplarily using the coordination-game and discoordination-game data. **Main Result 4** shows that basing a level-k model on individual post-game guessing-task estimates as level-0 can improve the model's fit to the data. Unfortunately, it also shows that a heterogeneous level-0 still does not allow to understand the data from the discoordination game.

Beyond a doubt, the results presented in this paper pose a serious challenge to level-k theory. If level-k theory is to be considered as a candidate for explaining behaviour in hide-and-seek data also in the future, it has to be modified in a way that provides an explanation also for the results presented here. And yet, we need to bear in mind that there is no other model at hand that can explain the recurrent features of the hide-and-seek data. Also, note that studies like Burchardi and Penczynski (2014) find empirical support for level-k-like reasoning in a clever design that allows to observe participants' *reasoning* rather than only their choices. What do the results mean, then? They may mean that only a subset of participants really follow level-k reasoning. For the remaining (majority of the) participants, we may have to look for different models to understand their behaviour.

Technical acknowledgements

All experiments were computerised using z-Tree (Fischbacher, 2007), participants were recruited using ORSEE (Greiner, 2004) with Mozilla Firefox. The statistical analyses were done using R (R Development Core Team 2001, 2012; Ihaka 1998) in combination with RKWard (Rödiger et al., 2012). All this was done on a computer running on KDE-based (KDE eV, 2012) Kubuntu, which required the use of wine for the programming of the experiments. The article was written using Kile.

References

Bardsley, Nicholas, Judith Mehta, Chris Starmer, and Robert Sugden. 2010. "Explaining Focal Points: Cognitive Hierarchy Theory *Versus* Team Reasoning." Economic Journal, 120: 40–79.

Burchardi, Konrad B., and Stefan P. Penczynski. 2014. "Out of Your Mind: Estimating Individual Reasoning in One Shot Games." Games and Economic Behavior, 84: 39–57.

Camerer, Colin F., Teck-Hua Ho, and Juin-Kuan Chong. 2004. "A Cognitive Hierarchy Model of Games." Quarterly Journal of Economics, 119(3): p861 – 898.

Christenfeld, Nicholas. 1995. "Choices From Identical Options." Psychological Science, 6(1): 50–55.

Costa-Gomes, Miguel A., Vincent P. Crawford, and Bruno Broseta. 2001. "Cognition and Behavior in Normal-Form Games: An Experimental Study." Econometrica, 69(5): 1193–1235.

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

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.

Eliaz, Kfir, and Ariel Rubinstein. 2011. "Edgar Allan Poe's Riddle: Framing Effects in Repeated Matching Pennies Games." Games and Economic Behavior, 71(1): 88–99.

Hargreaves Heap, Shaun, 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.

Heinrich, Timo, and Irenaeus Wolff. 2012. "Strategic Reasoning in Hide-and-Seek Games: A Note." Thurgau Institute of Economics Research Paper 74.

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.

Itti, Laurent, Christof Koch, and Ernst Niebur. 1998. “A Model of Saliency-Based Visual Attention for Rapid Scene Analysis.” IEEE Transactions on Pattern Analysis and Machine Intelligence, 20(11): 1254–1259.

Lewis, David K. 1969. Convention: A Philosophical Study. Harvard University Press.

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.

Nagel, Rosemarie. 1995. “Unraveling in Guessing Games: An Experimental Study.” American Economic Review, 85(5): 1313–1326.

Rubinstein, Ariel. 1999. “Experience from a Course in Game Theory: Pre- and Postclass Problem Sets as a Didactic Device.” Games and Economic Behavior, 28(1): 155–170.

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, Benny Moldovanu and Eric van Damme, 394–402. Springer-Verlag.

Rubinstein, Ariel, and Amos Tversky. 1993. “Naive Strategies in Zero-Sum Games.” Working Paper 17-93, The Sackler Institute of Economic Studies.

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

Stahl, Dale O., and Paul W. Wilson. 1994. “Experimental Evidence on Players’ Models of Other Players.” Journal of Economic Behavior & Organization, 25(3): 309–327.

Stahl, Dale O., and Paul W. Wilson. 1995. “On Players’ Models of Other Players: Theory and Experimental Evidence.” Games and Economic Behavior, 10(1): 218–254.

Towal, R. Blythe, Milica Mormann, and Christof Koch. 2013. “Simultaneous ModeModel of Visual Saliency and Value Computation Improves Predictions of Economic Choice.” PNAS, 110(40): E3858–E3867.

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. Harold W. Kuhn and Albert W. Tucker, 5–12. Princeton University Press.

Appendix

Alternative payoff-perturbed game

Here, I present the game with hard-wired payoff perturbations when $A_{(4)}$ is least salient without $A_{(1)}$. Under the additional simplifying assumption that $A_{(1)}$ and $A_{(3)}$ are equally salient, we obtain the game shown in Table A.1.[35] Again, we should expect $e > 0$ and $f > 0$.

		Seeker			
		$\mathbf{A}_{(1)}$	$\mathbf{B}_{(2)}$	$\mathbf{A}_{(3)}$	$\mathbf{A}_{(4)}$
Hider	$\mathbf{A}_{(1)}$	0, 1	1, $0+f$	1, 0	1, $0-e$
	$\mathbf{B}_{(2)}$	$1-f$, 0	$0-f$, $1+f$	$1-f$, 0	$1-f$, $0-e$
	$\mathbf{A}_{(3)}$	1, 0	1, $0+f$	0, 1	1, $0-e$
	$\mathbf{A}_{(4)}$	$1+e$, 0	$1+e$, $0+f$	$1+e$, 0	$0+e$, $1-e$

Table A.1: The hide-and-seek game with payoff perturbations when $A_{(4)}$ is the single least salient location and $A_{(1)}$ and $A_{(3)}$ are equally salient.

Full version of estimation-result Tables 5 and 6

On the next page, I include the Table-5 equivalent containing all estimated models, as well as the corresponding level-k distributions, including those reported in Table 6.

[35] This assumption can be based on the observations from the Guessing Task and, arguably, from the locations' average ranks in the Beauty Contest.

Specification	Data	$L1$	$L2$	$L3$	$L4$	logL	MSE
Choices follow salience							
naïve-pick-$B_{(2)}\left[A_{(1)}A_{(3)}\right]A_{(4)}$	rth	–	–	–	–	-1724	0.01271
naïve-postH&S-$B_{(2)}A_{(1)}\left[A_{(3)}A_{(4)}\right]$	rth	–	–	–	–	-1687	0.01647
Equilibrium models							
eqm$_0$	rth	–	–	–	–	-1641†	0.00967†
eqm$_+$-hyp-$\left[A_{(1)}A_{(4)}\right]$	rth	–	–	–	–	-1562†	0.00006†
$(e_H = 0.29, f_H = 0.25, e_S = 0.15, f_S = 0.15)$							
eqm$_+$-postH&S-$B_{(2)}A_{(3)}\left[A_{(1)}A_{(4)}\right]$	rth	–	–	–	–	-1636	0.00909
$(e_H = 0.00, f_H = 0.06, e_S = 0.00, f_S = 0.05)$							
eqm$_+$-hyp-$\left[A_{(1)}A_{(3)}\right]$	rth	–	–	–	–	-1608	0.00744
$(e_H = 0.08, f_H = 0.08, e_S = 0.17, f_S = 0.12)$							
eqm$_+$-pick-$B_{(2)}\left[A_{(1)}A_{(3)}\right]A_{(4)}$	rth	–	–	–	–	-1636	0.00909
$(e_H = 0.00, f_H = 0.06, e_S = 0.00, f_S = 0.05)$							
CI's preferred model							
Lk-hyp-$\left[A_{(1)}A_{(4)}\right]B_{(2)}A_{(3)}$	rth	0.19†	0.32†	0.24†	0.25†	-1564†	0.00027†
Lk-pick-$B_{(2)}\left[A_{(1)}A_{(3)}\right]A_{(4)}$	rth	0.21	0.00	0.79	0.00	-1635	0.00903
Lk-postH&S-$B_{(2)}A_{(3)}\left[A_{(1)}A_{(4)}\right]$	rth	0.31	0.00	0.69	0.00	-1664	0.01202
CI's model with asymmetric $L0$							
Lk-pick-$B_{(2)}\left[A_{(1)}A_{(3)}\right]A_{(4)}$-asym	rth	0.00	0.15	0.64	0.21	-1632	0.00782
Lk-postH&S-$B_{(2)}A_{(3)}\left[A_{(1)}A_{(4)}\right]$-asym	rth	0.08	0.26	0.32	0.34	-1603	0.00556
CI's model with salience-avoiding $L0$							
Lk-pick-$B_{(2)}\left[A_{(1)}A_{(3)}\right]A_{(4)}$-avoid	rth	0.00	0.79	0.06	0.15	-1632	0.00782
Lk-postH&S-$B_{(2)}A_{(3)}\left[A_{(1)}A_{(4)}\right]$-avoid	rth	0.12	0.46	0.20	0.22	-1603	0.00556
Choices follow salience							
naïve-pick-$B_{(2)}\left[A_{(1)}A_{(3)}\right]A_{(4)}$	hw	–	–	–	–	-521	0.01654
naïve-postH&S-$B_{(2)}A_{(3)}\left[A_{(1)}A_{(4)}\right]$	hw	–	–	–	–	-487	0.01662
Equilibrium models							
eqm$_0$	hw	–	–	–	–	-484	0.01436
eqm$_+$-hyp-$\left[A_{(1)}A_{(4)}\right]$	hw	–	–	–	–	-456	0.00109
$(e_H = 0.38, f_H = 0.29, e_S = 0.10, f_S = 0.02)$							
eqm$_+$-postH&S-$B_{(2)}A_{(3)}\left[A_{(1)}A_{(4)}\right]$	hw	–	–	–	–	-483	0.01467
$(e_H = 0.00, f_H = 0.04, e_S = 0.00, f_S = -0.05)$							
eqm$_+$-hyp-$\left[A_{(1)}A_{(3)}\right]$	hw	–	–	–	–	-480	0.01712
$(e_H = 0.12, f_H = 0.08, e_S = -0.03, f_S = -0.06)$							
eqm$_+$-pick-$B_{(2)}\left[A_{(1)}A_{(3)}\right]A_{(4)}$	hw	–	–	–	–	-482	0.01485
$(e_H = 0.00, f_H = 0.04, e_S = -0.03, f_S = -0.06)$							
CI's preferred model							
Lk-hyp-$\left[A_{(1)}A_{(4)}\right]B_{(2)}A_{(3)}$	hw	0.12	0.37	0.29	0.22	-456	0.00110
Lk-pick-$B_{(2)}\left[A_{(1)}A_{(3)}\right]A_{(4)}$	hw	0.23	0.00	0.70	0.07	-482	0.01358
Lk-postH&S-$B_{(2)}A_{(3)}\left[A_{(1)}A_{(4)}\right]$	hw	0.32	0.00	0.59	0.09	-485	0.01538
CI's model with asymmetric $L0$							
Lk-pick-$B_{(2)}\left[A_{(1)}A_{(3)}\right]A_{(4)}$-asym	hw	0.08	0.10	0.65	0.17	-484	0.01320
Lk-postH&S-$B_{(2)}A_{(3)}\left[A_{(1)}A_{(4)}\right]$-asym	hw	0.19	0.09	0.50	0.23	-465	0.00640
CI's model with salience-avoiding $L0$							
Lk-pick-$B_{(2)}\left[A_{(1)}A_{(3)}\right]A_{(4)}$-avoid	hw	0.00	0.75	0.07	0.18	-484	0.01320
Lk-postH&S-$B_{(2)}A_{(3)}\left[A_{(1)}A_{(4)}\right]$-avoid	hw	0.03	0.56	0.17	0.24	-465	0.00639

Table A.2: Full version of Tables 5 and 6 combined: maximum-likelihood estimates, log-likelihoods, and mean squared errors of the fit for the different models, using the data from Rubinstein, Tversky, and Heller's collected studies ("rth"), reproduced in Table 3 of Crawford and Iriberri (2007; "ci"). †indicates the estimate is taken from ci's paper. The table's lower half replicates the upper-half findings using Heinrich and Wolff's (2012; "hw") data.

THURGAU INSTITUTE
OF ECONOMICS
at the University of Konstanz

Hauptstr. 90
CH-8280 Kreuzlingen 2

Telefon: +41 (0)71 677 05 10
Telefax: +41 (0)71 677 05 11

info@twi-kreuzlingen.ch
www.twi-kreuzlingen.ch