No. 115

THURGAU INSTITUTE OF ECONOMICS
at the University of Konstanz

Irenaeus Wolff

Lucky Numbers in Simple Games

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

Lucky Numbers in Simple Games[§]

Irenaeus Wolff

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

Abstract:
Simple game structures like discoordination, hide-and-seek, or Colonel-Blotto games have been used to model a wide range of economically relevant situations. Yet, Nash-equilibrium and its alternatives notoriously fail to explain observed behaviour in these games when alternatives carry descriptive labels. This paper shows that throughout the different games, behavioural patterns resemble 'lucky-number' patterns: the choice patterns in related lotteries. Starting from this observation, I adjust standard models to account for the data. The adjusted models outperform the existing models, but they do not outperform a simple benchmark model. In the benchmark model, agents pick according to the 'lucky numbers' or, under certain circumstances, choose any of the other options with equal probabilities. Interestingly, this benchmark model predicts two additional general regularities that bear out on the existing data and new data from two additional games: hide-and-seek seekers rely on 'lucky numbers' more heavily than any other player role; and the stronger the 'lucky-number' pattern deviates from a uniform distribution, the more likely it is observed also in the game data.

Keywords: Bounded Rationality, Level-k, Salience, Heuristic, Hide & Seek, Discoordination, Rock-Paper-Scissors, Colonel Blotto, Representativeness.
JEL: C72, C90, D83.

1 Introduction

Alternatives in real-life situations are hardly ever abstract objects. Instead, they usually have non-neutral labels attached to them, and they often have a spatial

[§]I am thankful for the encouragement and helpful comments of Colin Camerer, Adrian Chadi, Vincent Crawford, Sebastian Fehrler, Urs Fischbacher, Susanne Goldlücke, Shaun Hargreaves Heap, Botond Köszegi, Jörg Oechssler, Stefan Penczynski, Ariel Rubinstein, Abdolkarim Sadrieh, Dirk Sliwka, Robert Sugden, Chris Starmer, Marie-Claire Villeval, Roberto Weber, and the lively research group at the Thurgau Institute of Economics (TWI), as well as to the participants of the World Congress of the Game Theory Society 2016 and the UEA Behavioural Game Theory Workshop 2016.

ordering. When driving from one city to another, we may choose between "the long but pretty route" and "the short but boring route" or between the "northern" and the "southern route", but we rarely choose between "option i and option j, where $i \neq j$".

Given the ubiquity of non-payoff-related characteristics like the above in our every-day life, it is important to understand how such descriptions of the available options affect behaviour also in strategic contexts. For example, when a mother and her child lose each other in the middle of a city (an obvious coordination game), "the place where they last talked to each other" may be more salient than "the next street on the right from wherever you are now".

The literature has long documented that such non-neutral decision frames do indeed have a strong influence on behaviour in games.[1] Some of these games have been at the heart of game theory for decades, capturing elements of important situations in every-day life: coordination games like that of the mother and her child looking for each other, when two firms simultaneously decide on market entry in two markets (potentially a discoordination game), or in military, police, and intelligence work (hide-and-seek games), just to name some examples. In contrast, Schelling's (1960) account of how an action's non-payoff-related characteristics may influence behaviour has re-entered the academic debate only rather recently.

This paper wants to understand behaviour in the general class of "strategy-isomorphic games" (Hargreaves Heap et al., 2014) that are not coordination games. Strategy-isomorphic games are all games in which the strategies are indistinguishable once we remove the labels of players' strategies. Examples include the discoordination and hide-and-seek games mentioned above. I do not focus on coordination games in this paper because I am convinced that psychologically, coordination games induce a completely different reasoning in players' minds. If the available options are denoted by "A", "B", "A", and "A", the question of "what option can we coordinate on?" will generate a different answer in most people than the question "where should I be hiding so that the other player doesn't find me"?[2] Running experiments on a number of these games, Rubinstein et al. (1997) show that participants' behaviour deviates systematically from the unique Nash-equilibrium (uniform randomisation).[3] For example, in a hide-and-seek game

[1] See, *e.g.*, Rubinstein et al. (1997), for coordination games, discoordination games, and hide-and-seek games; Scharlemann et al. (2001) for trust games where participants' interaction partners were "labeled" by photographs; or Mehta et al. (1994) and Bardsley et al. (2010), for coordination games with options carrying naturally occurring labels (following Schelling, 1960).

[2] I include coordination-game data and the corresponding analysis in Appendix B for reference. The analysis indeed suggests that reasoning is clearly different in coordination games than in the games I focus on in the main body of the paper.

[3] See also Rubinstein and Tversky (1993), Rubinstein (1999), or Penczynski (2016).

played on ("A", "B", "A", "A"), there is a clear mode on the 'central A' that is even more pronounced for seekers than for hiders. Uniform randomisation is also the quantal-response equilibrium given that strategies are not distinguishable by their payoffs. Finally, level-k or cognitive-hierarchy models yield the same solution if they rely on a uniformly-mixing level-0.[4]

There is only a single model in the literature that accounts for some of Rubinstein et al.'s data. In an important contribution, Crawford and Iriberri (2007) show that a level-k variant based on salience as level-0 can account for hide-and-seek data from a number of frames. However, Hargreaves Heap et al. (2014) argue that, had Crawford and Iriberri (2007) used all available games for the frames Crawford and Iriberri use (namely, the data also from the coordination and discoordination games), the model's explanatory power would have been small. In Wolff (2016), I use the central frame from Crawford and Iriberri (2007)—ABAA—and elicit salience in nine different ways. The elicited salience patterns all tend to be similar, but they do not allow to account for the data when used as level-0. In light of the above, it seems fair to say that we currently lack a model that would come anywhere close to describing data from a domain that is both at the very heart of game theory and relevant for every-day economic behaviour.

The first contribution of this paper is to establish that when people play different games on the same frame, the actual game being played has a surprisingly minor role: behaviour always is well-correlated with people's 'lucky numbers'—the options they would pick in a lottery.[5] Note that the way I use the term 'lucky numbers' in this paper differs from how it is used in everyday language. I do not assume that when I ask participants "what is your lucky number in ABAA?" they would reply by "the central A." Rather, I use the term 'lucky numbers' to describe what options participants choose in a lottery context. Section 4.6 relates these choices to other concepts like salience, representativeness, and valence.

Starting from the observations above, the paper sets out to understand the strategic behaviour of participants in a variety of games by enriching two of the most popular models—Nash equilibrium and level-k—in order to accommodate the findings. I then compare the models' ability to fit the data to four benchmarks: the standard versions of Nash equilibrium and level-k, salience-based level-k,

[4]The quantal-response equilibrium was suggested by McKelvey and Palfrey (1995) and applied successfully, *e.g.*, by Capra et al. (1999). Level-k was introduced by Nagel (1995) and Stahl and Wilson (1994). Building on these, studies like Costa-Gomes et al. (2001) or Crawford (2003) further developed level-k models, while Camerer et al. (2004) developed the cognitive-hierarchy model. For an overview, see Crawford et al. (2013).

[5]An earlier reader of the paper pointed out that this correlation might be due to the fact that seekers and coordinators have to *match* their opponents' action much like lottery players have to *match* the randomly selected option. However, this argument does not explain why 'lucky numbers' would also be so well-correlated with the choices of hiders, discoordinators, or players of the other two games I will be discussing in this paper.

and a simple heuristic. Under the heuristic, agents first consider playing their 'lucky numbers'. With a certain probability, $(1-p)$, they simply implement this choice. However, with the counter-probability, p, they reconsider their choice. Reconsidering agents only sticks to their choice if it is a best-response to a 'lucky-number'-playing opponent. Otherwise, they choose anything else with equal probabilities.

In principle, the probability with which agents reconsider could be different between people as well as between games—there could be certain situations in which people are primed to reconsider and others in which they are not. However, for the purpose of this paper, I assume the reconsidering probability to be the same for all players and all games. The implication of this assumption is quite radical: no matter which strategy-isomorphic game we look at, there are only two choice patterns that can be consistent with the heuristic. One choice pattern coincides with the 'lucky-number' pattern, and the other one with a convex combination of 'lucky numbers' (weighted by $(1-p)$) and uniform distribution amongst the 'non-lucky numbers' (weighted by p). Note that this implication makes it clear that the lexicographic nature of the heuristic is important: if the heuristic were just a convex combination of two behavioural rules, the prediction would be a single pattern. In contrast, the heuristic makes a different prediction for hiders and seekers of a hide-and-seek game (or for seekers and for discoordinators), for example.[6]

The 'lucky number'-based variants of level-k and Nash equilibrium generally outperform the pre-existing standard models. Surprisingly, however, none of the adapted models consistently outperforms the fourth benchmark, the two-step heuristic. This holds true even though the enriched models all have a higher number of free parameters, and although following the heuristic involves hardly any strategic reasoning: in most of the games in this paper, the 'best-response check' merely makes agents randomise among 'non-lucky numbers'.

This result means that we cannot anchor simply any behavioural model in 'lucky numbers' and expect a satisfactory explanation for the data: doing so for a level-k model does not strongly outperform the heuristic on hide-and-seek data

[6]Note also that the heuristic is not just 'salience plus randomisation': while 'lucky numbers' are influenced by salience, this is not the only component as my analysis in Section 4.6 shows. To take a prominent example, in ABAA, the most salient option is clearly the B, while the 'most lucky number' by our measure is 'the central A'. Finally, note the relationship of the heuristic with level-k: a non-reconsidering agent will behave like level-0 in a 'lucky number'-based level-k model; a reconsidering agent who finds it optimal to stick to her choice essentially moves to level-1 in the same model; yet, reconsidering agents who do not find it optimal to stick to their choices do something that is completely outside of that model (their choice is not a best-response). Thus, for seekers, the heuristic coincides with a specific level-k model that is based on a 'lucky number'-choosing level-0 and is truncated at level-1, while for hiders or discoordinators, it only shares the level-0 part.

and performs clearly worse in discoordination games. The reason for why the heuristic performs so well is that it captures two important aspects of the choice process. First, people are attracted by 'lucky numbers' in all games. This aspect is present also in the enriched variants of Nash equilibrium and level-k. But there is also the second aspect: people's strategic reaction to others favouring 'lucky numbers' seems to be weaker and less directed than the 'best-response' concept suggests.

The heuristic implies two further general regularities that bear out surprisingly well on both the existing data and new data from two additional games: (i) the stronger the 'lucky-number' pattern deviates from uniform randomisation, the more likely we will observe it also in the game data, and (ii) hide-and-seek-game seekers rely on 'lucky numbers' more than any other player role I focus on in this paper. Both regularities together also imply the seeker-advantage that has been documented in hide-and-seek games (*e.g.*, Rubinstein et al., 1997; Eliaz and Rubinstein, 2011). In other words, what was meant to be a simple benchmark model ends up accounting for a long-standing puzzle from the literature.

The present research contained three steps. The first step was the attempt to find an explanation for the data by amending two popular models of strategic behaviour. While fitting the amended models to the data, I noticed that a simplistic benchmark strategy was able to account for the data at least as well as the amended models. The second step was to derive some general implications from this benchmark strategy that I could use as testable *ex-ante* hypotheses for additional data. To do so, I enlarged the set of games by adding a Colonel-Blotto game (Borel, 1921) and a new game I call the 'to-your-right game'. In the 'to-your-right game', a player wins a prize if she chooses the action immediately to the right of her opponent's choice (in a circular fashion, so that the left-most action wins against the right-most action). The game is thus a variant of the rock-paper-scissors game.[7]

Up to this point, all games I looked at were four-action games. The Colonel-Blotto game, instead, is a multi-action game. In the variant I use, players have to allocate up to 40 troops to four locations (that correspond to the four actions in the other games); whoever has more troops on a location wins it, and whoever wins more locations wins the game (with random tie-breaking). I added this game as another example for a useful game: *inter alia*, Blotto games have been used to model allocations of funds towards different voter groups in electoral campaigns.[8] Note that I did not add the to-your-right and Blotto games to

[7]The game is also related to the 11-20 game (Arad and Rubinstein, 2012a), in which a player earns a bonus if she chooses exactly one unit less than the other player. An important difference is that prizes differ for different options, so that the game is not a strategy-isomorphic game.

[8]*E.g.*, Groseclose and Snyder (1996); for an earlier experimental implementation, see, *e.g.*, Chowdhury et al. (2013). Arad and Rubinstein (2012b) analyse participants' strategies in a Blotto

compare the predictive power of the heuristic against some equilibrium (which would not predict any systematic, and thus, exploitable non-uniformness in the observed choice pattern). Rather, I added them for a proof of concept (by testing the derived *ex-ante* hypotheses), and to suggest how to apply the heuristic to more general games.

The third step was to look at behaviour at the individual level. To this end, I went back to the laboratory to run the discoordination, hide-and-seek, and to-your-right games again using the *same* participants in all games as well as the lottery task to elicit the individuals' 'lucky numbers'. Does the heuristic successfully organise this data, too? The answer depends on the benchmark: behaviour in all games is still clearly correlated with participants' betting choice (on average, between 28.3 and 32.8 percent of choices in a game are identical to the participant's choice in the corresponding lottery task). On top, there is a strong correlation between games of how much a participant relies on 'lucky numbers' in the different games.

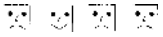
On the other hand, the frequency of 'lucky-number' choices amongst seekers (32 percent) is far from the heuristic's predicted 100 percent (which, however, relies on a deterministic interpretation of 'lucky numbers' that may be much closer to the literal meaning of 'lucky numbers' than the concept I use here. And still, discarding the explanation altogether based on this observation seems wrong: we would be left without an explanation for the four characteristics of the data described in this paper (see Section 4.5 for an additional argument). I am the first to acknowledge that the proposed heuristic may prove not to be the complete answer yet. What does become clear, however, is that participants favour 'lucky numbers', and that 'best-responding' seems to be too demanding as a concept to describe participants' strategic reaction to others favouring 'lucky numbers'.

2 The data

I use data of several papers on behaviour in games where actions are not distinguishable by their payoffs. All of the games are two-player games played on frames that have four locations each. The left-hand column of Table 1 provides an overview of the frames. I use all frames presented by Rubinstein and Tversky (1993) and Rubinstein et al. (1997), plus two obvious complements, BAAA and AAAB, as well as the Ace-2-3-Joker frame introduced by O'Neill (1987) and also referred to in Crawford and Iriberri (2007).

It can be argued that including all of Rubinstein et al.'s frames distorts the analysis because some of these frames use labels with positive or negative con-

game also with respect to locations (left *vs.* right, centre *vs.* borders).

Frame	Discoordination	Hiders	Seekers	To-your-right	Blotto
	RTH (49)	RTH (53)	RTH (62)	new (110)	new (94)
polite-rude-honest- -friendly	RTH (49)	RTH (53)	RTH (62)	new (110)	new (94)
	RTH (49)	RTH (53)	RTH (62)	new (110)	new (94)
ABAA	RTH+WB+BW (442)	RTH+HW+W (339)	RTH+HW+W (281)	new (110)	new (94)
	RTH (49)	RTH (53)	RTH (62)	new (110)	new (94)
hate-detest-love-dislike	RTH (49)	RTH (53)	RTH (62)	new (110)	new (94)
1-2-3-4 (292)	WB+BW (187)	RT (84)	RT (110)	new (94)	new
AABA	WB+BW (292)	RT (189)	RT (85)	new (110)	new (94)
Ace-2-3-Joker	WB+BW (292)			new (110)	new (94)
BAAA	WB+BW (292)			new (110)	new (94)
AAAB	WB+BW (292)			new (110)	new (94)

RTH: Rubinstein et al. (1997). RT: Rubinstein and Tversky (1993). HW: Heinrich and Wolff (2012). BW: Bauer and Wolff (2018). WB: Wolff and Bauer (2018). W: Wolff (2015).

Table 1: Origin of the data I use (numbers of observations in parentheses).

notations. Therefore, choosing the associated actions may increase or decrease utility on top of the utility associated with the resulting monetary outcome. I nevertheless include all of Rubinstein et al.'s frames, for three reasons: (i) in my view, understanding behaviour in non-neutral landscapes extends beyond 'neutral non-neutral' landscapes (and it would be difficult to draw the line if we accept the idea that people tend to have lucky numbers); (ii) at least the heuristic and the LuckyNoEqm Nash-equilibrium variant described below are meant to apply also under 'truly non-neutral' frames (much like the Nash-equilibrium with payoff perturbations considered in Crawford and Iriberri, 2007); and (iii) excluding frames with clearly positive or negative connotations in our sample does not change the results meaningfully but leaves us with less statistical power for the analysis.[9]

Table 1 presents the origin of the data I use in this study, together with the

[9] Appendix C presents the main tables from the data analysis for the reduced sample. Most importantly, the heuristic still has the lowest mean squared error for discoordination games and the second-lowest mean squared error in the hide-and-seek games; at the same time, the best-performing model for hide-and-seek games is among the worst-performing models for the other games.

number of observations (in parentheses). The data for the hide-and-seek games mostly comes from Rubinstein and Tversky's (1993) and Rubinstein et al.'s (1997) studies, only for the ABAA frame, I have additional observations from other studies for each game. For the discoordination games, Rubinstein et al. (1997) only have data for six of the frames. I complement this with data on a different subset of six frames collected for two studies run by Dominik Bauer and myself. Finally, I collected the data for the to-your-right and Blotto games specifically for this study, to see whether the predictions and model implications also apply to a new setting. A complete listing of all the data I use—old and new—can be found in Appendix A.

I ran the to-your-right and Blotto games as the first part of (different) sessions comprised of three parts, where I described each part only as it started. Only one part was paid. If the first part was payoff relevant, the roll of a die selected one of the to-your-right games (or Blotto games, in the Blotto sessions) for payment. Participants played under all eleven frames with a randomised order, random rematching, and without feedback between games. I described the to-your-right game as follows:

> There are four boxes. You and the other participant choose a box without knowing the decision of the respective other. One of you can obtain a prize of 12 Euros. Who wins depends on the relative position of the two chosen boxes. The participant wins whose box lies to the inmediate right of the box of the other participant. If a participant chooses the right-most box, then the other participant wins if he chooses the left-most box. Who does not win obtains a consolation prize of 4 Euros. Of course, it is possible that neither you nor the other participant wins.[10]

Participants had not participated in any other experiments using the same type of non-neutral frames.[11]

To inform my heuristic and some of the other models I use on the game data, I collected data from additional tasks in separate sessions. First, I conducted seven sessions with a total of 140 participants that had a BettingTask as the first of several parts (following the same procedures as with the to-your-right and Blotto games). In the BettingTask, participants faced the following task:

[10] Similarly, the instructions for the Blotto game read: There are four fields. You and the other participant have to assign 40 units to the four fields without knowing the decision of the respective other. One of you can obtain a prize of 12 Euros. Who wins depends on how many fields you can win. The participant who has placed more units on a field wins the field. The participant who wins most fields overall wins the prize of 12 Euros. The other participant obtains a consolation prize of 4 Euros. If both participants win the same number of fields, chance determines who obtains which prize.

[11] I used z-Tree (Fischbacher, 2007) and ORSEE (Greiner, 2015).

> In each decision situation you have to choose one out of several boxes. Subsequently, one of the boxes will be randomly selected by the cast of a die. In case the randomly-selected box coincides with the box you chose, you receive 12 Euros. If the two boxes do not coincide, you receive 4 Euros.

Next, I conducted three sessions with a total of 58 participants of a HiderBettingTask. This task differs from the BettingTask only in that participants receive the bigger prize if their choice does not coincide with the randomly-selected box. Finally, I asked 96 participants to rate the options' optic salience (SalienceRating) and 102 participants to rate how well each of the boxes within a frame represented all four boxes within that frame (RepresentRating).[12] In both tasks, participants saw the boxes in the same horizontal line-up as in the other tasks. Below each box, they would have a slider (empty at the outset) to indicate the level of optical salience (between "extremely conspicuous", top, and "extremely nondescript", bottom) or representativeness (between "totally representative", top, and "not representative at all", bottom).

3 The amended models and a simple heuristic

This section presents the models I use in the paper. As I argued in the introduction, there are no 'standard' alternatives available that would come close to explaining the data. Starting from the observation that the BettingTask choices are surprisingly similar to the data from the different games, I construct two 'sensible' level-k alternatives and an equilibrium with payoff-perturbations based on the BettingTask choices. Then, I compare the three amended models to the standard Nash equilibrium, a standard level-k model with uniformly mixing level-0, the salience-based level-k of Crawford and Iriberri (2007) (with an empirically-defined level-0), and the simple two-step heuristic that ended up explaining the data surprisingly well. I start out with the equilibrium models, followed by the level-k models and the heuristic, at the end.

NashEqm. The unique symmetric mixed-strategy equilibrium that has both players randomise uniformly over all locations.

[12] The rating tasks were included in BettingTask and HiderBettingTask sessions. One could argue that this procedure could bias the rating-task data. However, in particular the SalienceRating data has so little variance and corresponds so well with intuition, that I see little value in repeating the task in separate sessions. The RepresentRating data has more variance, but I do not use it to inform any of the models in Section 3.

LuckyNoEqm. A Nash-equilibrium variant in which participants derive extra utility from choosing certain locations (cf. Crawford and Iriberri, 2007). For this model, I interpret the BettingTask data as a measure of participants' inherent preferences for the different locations.[13] I compute utility values from the BettingTask data and re-define the game in terms of these utility values: A multinomial-logit utility model estimated by maximum likelihood yields utility values that I transform in an affine-linear way (to obtain positive utility values). Then, I calculate the mixed-strategy equilibria for the games that result when the non-zero entries in the standard game matrix are replaced by the transformed utility values. Finally, I use another layer of maximum-likelihood estimation to obtain the transformation of utility values and probability of trembles by players that yield the best-possible fit to the data. Note that the transformation of utility values does not affect BettingTask choice under the multinomial-logit model, but it does affect the calculated mixed-strategy equilibria.

Take the example of a discoordination game played on abaa: (Absolute) choice frequencies in the BettingTask were 18, 46, 58, and 18. If these frequencies are the result of a multinomial-logit choice process, the maximum-likelihood estimates for utilities are -0.52, 0.42, 0.65, and -0.52. I recalibrate those utilities to 0.65, 1.27, 1.43, and 0.65 (which are still in accordance with the BettingTask choice frequencies) and use the recalibrated values as the corresponding entries in the normal form game: when a participant chooses one of the end-as and her opponent chooses another location, the participant's utility will be 0.65. Likewise, when she successfully discoordinates by choosing b, her utility will be 1.27. Using the resulting normal-form game, the unique symmetric equilibrium (mixed) strategy would be (0, 0.47, 0.53, 0). As I point out above, I allow for errors and allow the maximum-likelihood procedure to optimise over another layer of utility-recalibration for the model comparison in part 4.1.

Standard Lk. The predictions of the standard level-k model with a uniformly mixing level-0 coincide with those of the standard Nash equilibrium for the games examined here. Therefore, the model will be subsumed under "NashEqm" for the remainder of the paper.

Salience-Lk. Crawford and Iriberri's (2007) level-k model in which level-0 follows salience, and level-k players with $k > 0$ play a best-response to level-$(k-1)$ players. Rather than making assumptions about what is salient, I use data from the SalienceRating task as the level-0 to base the model on.[14] For

[13] Of course, this assumes that people are homogeneous in what utilities they derive from the different locations. This is a strong assumption, but it is the best approximation that I have.

[14] I use the distribution of locations that participants ranked as most salient, to obtain a metric that is comparable to the data from the BettingTask. Using the average salience rating for each

our discoordination-game example on ABAA, the first A is held to be the most salient location by 2% of all SALIENCERATING participants, B by 91%, and the two other As by 4% each. Therefore, we would expect level-0 to choose with probabilities $(0.02, 0.91, 0.04, 0.04)$, uneven levels to choose the first A, and even levels to randomise between the other three locations.

BETTING-Lk. This level-k model uses as level-0 the data from the BETTINGTASK. In level-k theories, level-0 is supposed to be people's intuitive reaction to the game, which may well coincide with the choice they make in a lottery. In the ABAA-discoordination-game example, betting proportions—and hence, level-0 choices—are 13%, 33%, 41%, and 13%, uneven levels randomise between the end As and even levels between the two locations in the middle.

BOUNDED Lk. This model differs from standard level-k with a uniformly randomising level-0 only in terms of level-1. It incorporates that level-1 players may respond to uniform randomisation by non-uniform randomisation (or by not randomising at all). The BETTINGTASK and HIDERBETTINGTASK elicit what participants do when facing uniform randomisation. Level-1 seekers will act like participants in the BETTINGTASK, whereas level-1 discoordinators and level-1 hiders will act like participants in the HIDERBETTINGTASK. For ABAA, the HIDERBETTINGTASK choice frequencies are 9%, 53%, 21%, and 17%. Therefore, in our discoordination-game example, level-0 would randomise uniformly, level-1 would choose with probabilities $(0.09, 0.53, 0.21, 0.17)$, even levels would choose the first A, and uneven levels of level-3 or higher would randomise uniformly among all locations but the first A.

LUCKYORANYTHING. The simple two-step heuristic meant to provide a benchmark to assess what drives the potentially better fit of the amended models—basing them on 'lucky numbers' or including strategic reactions. An agent following the heuristic proceeds in the following two steps:

1. choose your 'lucky number'. This is the choice you would pick in a lottery. With a certain probability $(1-p)$, end here. With the complementary probability, check whether the chosen option is a best-response if others pick 'lucky numbers', too (assuming that 'lucky numbers' are perfectly correlated between players). If so, stick to your choice and end here. If not,

2. make a uniformly-random choice amongst the remaining options.[15]

location does not change the results in any significant way.

[15] Alternatively, this heuristic could be rationalised as being an abbreviation of the following

In the first step, agents are assumed to think about their choice again with probability p. In principle, the probability p should be a constant (across games) that is related to the personal characteristics of the individual, most notably, with their tendency to re-assess first intuitions as measured, for example, by the cognitive-reflection test (Frederick, 2005). For the purpose of this paper, however, we will assume p to be the same for all participants.

According to the heuristic, seekers simply go by their lucky numbers: both are happy to stick to their choice when thinking that their opponent is likely to make the same choice. Discoordinators proceed to step 2 with probability p, because discoordinators do not want to choose the same items and 'lucky numbers' will be correlated. Hence, they will choose 'lucky numbers' with probability $(1-p)$, and uniformly-random amongst the 'non-lucky numbers' with probability p. The same prediction applies to hiders and to players in the 'to your right' game. To test the heuristic against data from the Colonel-Blotto game, we have to adapt its lucky-number step slightly as meaning that people deploy their resources according to the 'lucky-numberedness' of the options.[16] Again, a fraction p of Blotto players will conclude that playing 'lucky numbers' is not a best-response to others playing lucky numbers, and proceed to uniform randomisation amongst the 'non-lucky numbers'. These predictions mean that the LuckyOrAnything heuristic has two general implications:

invariance the qualitative distribution of choices for a given frame follow the qualitative distribution of 'lucky numbers', irrespective of the game; here, the "qualitative distribution" refers to which items are chosen the most often, the second-most often, *etc.*; and

L-differential the prevalence of 'lucky numbers' is highest for seekers in the hide-and-seek game (and coordinators; both play 'non-lucky numbers' only when

train of thought that players approaching a new situation might be following: "is there any obvious best option for me (checking for obvious dominance)? If not, is it clear what the other player will do (checking for obvious dominance for the other player)? If so, react correspondingly, if not, do we both want the same? If so, what do we have to do to make the best out of it (use team-reasoning)? If not, that is, if I still don't know what to do, I'll just pick what sounds best to me from among the options I have (step 1). Ah, wait, maybe I shouldn't do that if the other player does the same, should I (step 1b)? In that case, I'll just choose anything else (step 2)." Note that dominance does not have a bite in the games I consider, and team-reasoning would only help in the coordination games.

[16] It is unclear whether the 'lucky-numberedness' of an option can be measured by the fraction of people choosing it as their most lucky number. In principle, it seems more sensible to elicit participants' individual lucky-number orderings over all locations (*e.g.*, in a conditional betting task, in which participants have to specify what they bet on if their most-preferred option is not available) and then define the lucky-numberedness of the options by the resulting distribution of lucky-number orderings. Given we do not have this data, the best we can do is to assume both orderings will be correlated and to use the data we have.

making an error). Hiders, discoordinators, and 'to-your-righters' rely on 'lucky numbers' to the same (lower) degree.

Because 'lucky numbers' will not be the same for everybody, sampling participants into the experiment will introduce randomness in the aggregate data. This may change the qualitative choice pattern when the 'lucky-number' pattern is weak, but it should rarely do so when the pattern is strong. Hence, we obtain a third implication:

predict-differential the LuckyOrAnything heuristic accounts for the qualitative data pattern the better, the stronger the 'lucky-number' pattern is.

4 Results

4.1 Accounting for behaviour

model	fitted on	LogL	MSE	modes predicted	parameters
Betting-Lk	discoordination	-2980	0.0056	5 out of 11	3
NashEqm/Standard Lk		-2975	0.0045	2.75† out of 11	—
Salience-Lk		-2972	0.0039	6 out of 11	3
Bounded Lk		-2967	0.0034	7 out of 11	3
LuckyNoEqm		-2960	0.0038	6 out of 11	3
LuckyOrAnything		-2959	0.0039	6 out of 11	1
NashEqm/Standard Lk	hide & seek	-2412	0.0160	(2,2)† out of (8,8)	—
LuckyNoEqm		-2361	0.0116	(0,8) out of (8,8)	3
Salience-Lk		-2356	0.0126	(2,6) out of (8,8)	5
Bounded Lk		-2340	0.0089	(5,6) out of (8,8)	5
LuckyOrAnything		-2319	0.0081	(8,8) out of (8,8)	1
Betting-Lk		-2299	0.0066	(8,6) out of (8,8)	5
Salience-Lk	hide & seek +	-5595	0.0388	12 out of 27	5
NashEqm/Standard Lk	discoordination	-5392	0.0113	9† out of 27	—
LuckyNoEqm		-5334	0.0095	14 out of 27	3
Bounded Lk		-5319	0.0072	18 out of 27	5
LuckyOrAnything		-5280	0.0066	22 out of 27	1
Betting-Lk		-5279	0.0065	21 out of 27	5

†Expected number of correctly-predicted modes under uniform randomisation.

Table 2: Performance of the models in terms of data fitting, ordered by log-likelihood.

Observation. The amended models generally fit the data better than the standard game-theoretic models.

Looking at Table 2, the amended models (BETTING-Lk, BOUNDED Lk, and LUCKYNOEQM) generally outperform their corresponding standard model (STANDARD Lk and NASHEQM) both in terms of exhibiting a larger log-likelihood and a smaller mean squared error (MSE). Only in the discoordination game, BETTING-Lk performs worse than STANDARD Lk, albeit predicting more of the prediction modes.

Result 1. None of the amended models fits the data consistently better than the simplistic LUCKYORANYTHING heuristic, which on top has fewer parameters than the amended models.

Looking at the upper part of Table 2 again, the LUCKYORANYTHING heuristic exhibits the largest log-likelihood of the models when fitted on discoordination-game data.[17] It is outperformed in terms of the mean squared error (MSE) by SALIENCE-Lk and LUCKYNOEQM, and in terms of both the MSE and the number of choice-distribution modes correctly fitted by BOUNDED Lk. However, each of these four models performs clearly worse than LUCKYORANYTHING when fitted on hide-and-seek data, as the middle part of Table 2 shows. Here, BETTING-Lk—which LUCKYORANYTHING outperformed clearly in the upper half of Table 2—takes on the role of the main contender, with higher log-likelihood, lower MSE but less correctly-fitted modes. When fitted on all three player roles simultaneously (lower part of Table 2), BETTING-Lk and LUCKYORANYTHING go head-to-head on all three criteria.

So, while LUCKYORANYTHING does not dominate any of the other models, it always performs best on at least one criterion, and it does so using only one free parameter as opposed to three (for the discoordination data) or five (for the hide-and-seek data) as in the amended level-k models. As further suggestive evidence, the fitted BOUNDED Lk has virtually only levels 0 (uniform randomisation) and 1 (BETTINGTASK/HIDERBETTINGTASK), no matter which game the model is fitted on (a combined 100% if fitted on discoordination, 88% if fitted on hide and seek).

4.2 The *invariance* implication

The LUCKYORANYTHING heuristic predicts that the qualitative data pattern in all games will be the same as that of the corresponding BETTINGTASK. Looking at the modes as a first, crude measure, this prediction seems to hold for hiders and seekers (16 out of 16 modes correctly predicted), and to a lesser degree also for to-your-right players (8 out of 11), discoordinators and Blotto players (both times

[17] To fit the models, I calculate the predicted marginal probabilities as a function of the model parameters (for level-k models, the level distribution, for LUCKYORANYTHING, the 're-thinking probability' p). Using those marginal probabilities, I calculate the likelihood of the observed samples. The maximum-likelihood algorithm then optimises over the parameter values.

	Spearman coefficient	$p-$value
hiders	0.63	0.000
seekers	0.60	0.001
discoordinators	0.24	0.122
to-your-right players	0.50	0.001
Blotto players	0.41	0.007

Table 3: Correlations of ranks: game data and BettingTask data.

6 out of 11).[18] While the modes are interesting, the *invariance* implication speaks about the complete distribution. Table 3 presents the Spearman correlations of ranks between the game data and the BettingTask data for each of the player roles.

Result 2. *Invariance* tends to hold: the correlation of ranks between the game data and the BettingTask data is strong and significant for hiders, seekers, to-your-right players, and Blotto players, and still sizable for discoordinators.

As Table 3 shows, the correlation of ranks between the game data and the BettingTask data is 0.63 for hiders, 0.60 for seekers, 0.50 for to-your-right players, and 0.41 for Blotto players (all $p < 0.007$). The rank correlation for discoordinators is 0.24 ($p = 0.122$). At the same time, all predictions are clearly better than the predictions made by random choice. To see that, I calculate the mean squared difference in ranks between the game data and the BettingTask data, and compare this difference to the difference to be expected under random choice. To compare the mean squared rank differences to the BettingTask with that to the random-choice benchmark, I draw for each frame 100'000 sets of 110 draws from a uniform distribution over four items and convert the sets to rankings.[19] I then compare the mean squared rank difference between the game data and the BettingTask to the distribution of mean squared rank differences from the simulation, by means of a Kolmogorov-Smirnov test (bootstrapped to correct for the discreteness of the rankings). The corresponding p-values are all $p \leq 0.042$.[20]

[18]There is nothing systematic to be learnt from the deviations: the frames on which the different games deviate from the prediction overlap only partially, and when they do, the modes coincide in only 2 out of 5 cases.

[19]I chose 110 draws to match the median number of observations in my data set.

[20]Another way of looking at the question is to locate the median mean squared rank difference of a game with the BettingTask data within the simulated distribution of differences to random play. Here, we note that the median mean squared rank difference over all frames is always between 0.7 and 1.3 of a standard deviation (of the simulated random-play distribution) lower than the median of the simulated distribution.

	fitted prob('lucky numbers') in %
seekers	77
hiders	45
discoordinators	30
to-your-right players	15
Blotto players	9

Table 4: Fitted probability of choosing by 'lucky numbers' for each player role. NOTE: in the absence of errors, seekers should play 'lucky numbers' in 100% of the cases, while the other player roles should play 'lucky numbers' with the same probability (which is different from the heuristic's p: for ease of interpretation of the above exercise, we calculate the optimal weight in a mix with uniform randomisation over all options rather than with 'non-lucky numbers' only).

4.3 L-differential

The LuckyOrAnything heuristic predicts that seekers should be relying on 'lucky numbers' the most, because those who reconsider their choice would stick to it. The heuristic does not distinguish between the degrees to which hiders, discoordinators, and players of the to-your-right game should rely on 'lucky numbers'. To test the *L-differential* implication, I calculate the mix of 'lucky-number' choices and uniform randomisation that best explains the data separately on hiders, seekers, discoordinators, and to-your-right players, and report the fitted proportion of 'lucky-number' choices for each of them in Table 4.

Result 3. *L-differential* holds partially: while seekers are clearly the most likely to rely on 'lucky numbers', there also is a difference among the other player roles.

Table 4 clearly shows that seekers are the most likely to play according to 'their lucky numbers'. Whether there is a difference between the other player roles is unclear at first sight: they may all have a propensity to choose by 'lucky numbers' of around 30%, but there also seems to be a clear difference between hiders and to-your-right/Blotto players that the heuristic does not account for. Likelihood-ratio tests reveal that all of the differences reported in Table 4 are significant (with $p \leq 0.021$), except for the difference between to-your-right and Blotto players ($p = 0.485$).[21]

[21] To obtain these p-values, I run maximum likelihood estimates for each role with the probability of uniform mixing being constrained to each of the other estimates. The Likelihood-ratio test then compares the maximum likelihood of the constrained and the unconstrained estimations.

4.4 Predict-differential

In the preceding two sections, I presented evidence that the two implications *invariance* and *L-differential* offer considerable guidance in understanding behaviour. Yet, none of the two implications holds perfectly. As also implied by the heuristic under random sampling of participants, we can use information from the BettingTask data to identify the frames in which the heuristic fits better.

Result 4. There is a positive correlation between the predictive accuracy of the LuckyOrAnything heuristic and the strength of the 'lucky-number' pattern.

I measure the strength of the 'lucky-number' pattern by the difference between the relative frequencies of the most popular and the least popular choices in the BettingTask. Then, I relate this difference to the mean squared rank difference between game data and LuckyOrAnything prediction, for each player role. I find that all five correlations are negative (for discoordinators, hiders, seekers, to-your-right and Blotto players). The probability of all five correlations showing as negative if there was no true relationship is $p = (\frac{1}{2})^5 = 0.03125$.

4.5 Individual-level data

Some readers of an earlier version of this paper commented that it would be extremely useful to obtain individual-level data, in the sense of having the same people go through the BettingTask and playing (some of) the games. While I was (and still am) somewhat sceptical about this procedure—for reasons I will discuss below—I heeded the commenters' advice and ran three additional sessions (with a total of 82 new participants). In the sessions, participants would play hide-and-seek (in one of the roles), discoordination, and to-your-right games on all eleven frames before doing the BettingTask on all frames.

In particular, a participant would have played a discoordination game on one frame, a hide-and-seek game on another frame, a to-your-right game on the next frame, and then the next discoordination game on a fourth frame, and so on, until all games had been played on all frames. The order of the eleven frames was randomised individually (and then repeated four times). I also randomised the starting games and a participant's opponents for each of the three randomly-selected payoff-relevant decisions. Thus, participants *de facto* faced a protocol with random-rematching after each round and game, and without any feedback in between.

The left column of Table 5 shows the predictive success of choices in the BettingTask for choices in the different games under the corresponding frames. What is obvious is that the average probability of choosing the same option as

	fraction of betting-like choices (in %)	fraction of players (in %)
discoordinators	30	63 ($p = 0.020$)
to-your-right players	28	60 ($p = 0.097$)
seekers	32	63 ($p = 0.117$)
hiders	33	59 ($p = 0.349$)

Table 5: Relative frequency of choices for each game that where identical to the chosen option under the same frame in the BettingTask (left); relative frequency of participants who chose the same option as in the BettingTask more often than at chance level (right; binomial-test p-values in parentheses).

	To-your-right	Hiding	Seeking
Discoordination	0.265 (p = 0.016)	0.792 (p < 0.001)	0.004 (p = 0.981)
To-your-right		0.462 (p = 0.002)	0.362 (p = 0.020)

Table 6: Pearson coefficients correlating the frequency of betting-like choices in pairs of games/roles.

in the corresponding BettingTask is always higher than chance (25%). Similarly, the fraction of players choosing the same option as in the corresponding BettingTask more often than 25% of the time (displayed in the right column of Table 5) is always higher than 50%. In addition, the figures for seekers are always among the highest of the four values. However, they are far from being as high as we should have expected if the heuristic was an accurate model of the choice process.

Could it be that only a subset of participants can be described well be the heuristic? In that case, participants' frequency of betting-like choices in one game should be correlated with the same participants' frequency of betting-like choices in another game. And indeed, this is the case for most combinations we can examine, as depicted in Table 6 where 5 out of 6 correlations are substantial and significant. At the same time, histogrammes of the frequencies of betting-like choices (not depicted here) do not show any clear trace of a bimodal distribution. So, while there seems to be a personality-related proneness to using one's 'lucky number' also outside of lotteries, the individual-level dataset shows no clear evidence of participants following the LuckyOrAnything heuristic.

Having said this, the individual-level data has to be treated with some caution when it is used for assessing the type of models referred to in this paper. In particular, level-k is meant to explain the choices people make when they see a game for the first time, and the heuristic cannot sensibly be expected to explain anything else, either. However, people may think about a certain frame (like ABAA) very differently the first time they face this frame compared to if they have faced the same frame in a different game altogether. Although participants did not receive any feedback during the experiment, it thus may be a different thing to play hide-and-seek on an ABAA frame if you have already played a discoordination game on the same frame, compared to if hide-and-seek is the first (and only) thing you play on that frame. The fact that the data patterns resulting in the individual-level experiment tend to be rather different from the data patterns we know from the stand-alone treatments might give some credence to the argument presented here.[22]

4.6 Relating 'lucky numbers' and other concepts

As a final point, I relate 'lucky numbers' to characteristics of the labels within their frame. As characteristics, I use subjective and relatively objective criteria. As objective criteria, I include relative position (0.5 for the middle, 1 for the rightmost locations) and valence (positive, negative, or neutral). Subjective criteria are salience and representativeness, which I objectivise by measuring students' assessment of them (in SalienceRating and RepresentRating; representativeness is important for choice among evidently equivalent items, cf. Bar-Hillel, 2015). Table 7 reports the corresponding regression.

Observation. Several characteristics interact to make an item a 'lucky number', among them salience, position, and valence.

As we can see from the Table, there seem to be four characteristics that increase the relative frequency of an item being picked in the BettingTask: (i) being rated as more salient, (ii) being positioned in the middle (to see this, combine relative position and its square); (iii) having a positive connotation when the item is salient (in the context of our frames, this essentially means that the positively connoted item is presented along with three negatively connoted items); and potentially, (iv) being rated as being representative. While this analysis has to be taken with caution because the set of frames is rather peculiar, it may serve as a first indication of what may determine 'lucky numbers' in general.

5 Discussion

In this paper, I have looked at games in which players' strategies are indistinguishable once we remove the strategies' labels. Many of these games are at

[22]The data from the individual-level experiment is shown in Appendix D.

	coefficient	s.e.	p-value
(Intercept)	-0.21	(0.14)	0.1461
negative	0.01	(0.12)	0.9425
positive	-0.11	(0.08)	0.1959
SalienceRating	0.43	(0.13)	0.0020
RepresentRating	0.25	(0.13)	0.0693
relative position	0.62	(0.11)	$1 \cdot 10^{-6}$
(relative position)^2	-0.54	(0.10)	$5 \cdot 10^{-6}$
negative·SalienceRating	-0.26	(0.22)	0.2574
positive·SalienceRating	0.33	(0.15)	0.0364
R^2	0.72		
Adj. R^2	0.66		
Num. obs.	44		

Table 7: Regression of relative choice frequencies in the BettingTask on label characteristics.

the heart of game-theoretic reasoning, at the same time as capturing important elements of every-day life. This paper focuses on versions of the games that incorporate a key aspect of reality, namely that options carry *descriptive* labels.

Using data from both earlier studies and new experiments, I identify four characteristics of the data that none of the popular models accounts for: (i) the qualitative choice distribution tends to be the same, irrespective of the game being played;[23] (ii) the data pattern is strongly correlated with 'lucky numbers', the options people bet on in a lottery; (iii) seekers rely on 'lucky numbers' the most; and (iv) the stronger the 'lucky-number' pattern deviates from uniform randomisation, the more likely we will observe it also in the game data.

I adapt two of the most popular models, Nash equilibrium and level-k, in straightforward ways to accommodate some of the above characteristics. The amended models clearly do better in accounting for the data than the standard models. At the same time, none of the amended models beats a simple benchmark in which participants play their 'lucky numbers' and (only) sometimes reconsider, in which case they under certain conditions mix uniformly among the 'non-lucky numbers'. Moreover, the simple benchmark model accounts for all four characteristics, while the amended models do not.

[23]Models that predict uniform randomisation like the standard Nash-equilibrium do predict *ex-ante* invariance to the game. However, when sampling from a uniform distribution for two different games, we would not expect invariance in the rank distributions, which is what we observe in the data.

Is this benchmark model an adequate representation of people's choice process? Probably not. The reason for why it performs so well is that it captures well two important aspects of the choice process. First, people seem to be attracted by 'lucky numbers' in all games. And second, their strategic reaction to others favouring 'lucky numbers' seems to be weaker and less directed than the 'best-response' concept suggests.

People's strategic reaction seems to be systematic only in the sense that hide-and-seek-game seekers rely on 'lucky numbers' the most—which makes sense if players sometimes check whether their action would be a good idea if others also play 'lucky numbers'. However, it is not as strategic as our standard models like Nash equilibrium or level-k would prescribe: there is no evidence of several layers of best-responding behaviour or even mutual best-responding.

In summary, the common models of strategic behaviour probably err on the side of ascribing too much strategic reasoning to the average participant in our experiments. The heuristic I have presented here is likely to err on the other side. And yet, it seems to capture important elements of decision-making. The heuristic thus should be seen as a thought-provoking impulse to help us finally get to grips with the conundrum we face since the papers by Rubinstein, Heller, and Tversky.

References

Arad, A. and Rubinstein, A. (2012a). The 11–20 money request game: A level-k reasoning study. *American Economic Review*, 102(7):3561–3573.

Arad, A. and Rubinstein, A. (2012b). Multi-dimensional iterative reasoning in action: The case of the colonel blotto game. *Journal of Economic Behavior & Organization*, 84(2):571–585.

Bar-Hillel, M. (2015). Position effects in choice from simultaneous displays: A conundrum solved. *Perspectives on Psychological Science*, 10(4):419–433.

Bardsley, N., Mehta, J., Starmer, C., and Sugden, R. (2010). Explaining focal points: Cognitive hierarchy theory *Versus* team reasoning. *Economic Journal*, 120:40–79.

Bauer, D. and Wolff, I. (2018). Biases in beliefs: Experiment evidence. TWI Research Paper Series, No. 109.

Borel, E. (1921). La théorie du jeu les équations intégrales à noyau symétrique. *Comptes Rendus de l'Académie*, 173:1304–1308. English translation by Savage, L.: The theory of play and integral equations with skew symmetric kernels. Econometrica 21:97–100 (1953).

Camerer, C. F., Ho, T.-H., and Chong, J.-K. (2004). A cognitive hierarchy model of games. *Quarterly Journal of Economics*, 119(3):861–898.

Capra, C. M., Goeree, J. K., Gomez, R., and Holt, C. A. (1999). Anomalous behavior in a traveler's dilemma? *American Economic Review*, 89(3):678–90.

Chowdhury, S. M., Kovenock, D., and Sheremeta, R. M. (2013). An experimental investigation of colonel blotto games. *Economic Theory*, 52(3):833–861.

Costa-Gomes, M. A., Crawford, V. P., and Broseta, B. (2001). Cognition and behavior in normal-form games: An experimental study. *Econometrica*, 69(5):1193–1235.

Crawford, V. (2003). Lying for strategic advantage: Rational and boundedly rational misrepresentation of intentions. *American Economic Review*, 93(1):133–49.

Crawford, V. P., Costa-Gomes, M. A., and Iriberri, N. (2013). Structural models of nonequilibrium strategic thinking: Theory, evidence, and applications. *Journal of Economic Literature*, 51(1):5–62.

Crawford, V. P. and Iriberri, N. (2007). Fatal attraction: Salience, naïveté, and sophistication in experimental 'hide-and-seek' games. *American Economic Review*, 97(5):1731–1750.

Eliaz, K. and Rubinstein, A. (2011). Edgar allan poe's riddle: Framing effects in repeated matching pennies games. *Games and Economic Behavior*, 71(1):88–99.

Fischbacher, U. (2007). z-tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics*, 10(2):171–178.

Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives*, 19(4):25–42.

Greiner, B. (2015). An online recruitment system for economic experiments. *Journal of the Economic Science Association*, 1(1):114–125.

Groseclose, T. and Snyder, Jr., J. M. (1996). Buying supermajorities. *American Political Science Review*, 90(2):303–315.

Hargreaves Heap, S., Rojo Arjona, D., and Sugden, R. (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, T. and Wolff, I. (2012). Strategic reasoning in hide-and-seek games: A note. Research Paper 74, Thurgau Institute of Economics.

McKelvey, R. D. and Palfrey, T. R. (1995). Quantal response equilibria for normal form games. *Games and Economic Behavior*, 10(1):6 – 38.

Mehta, J., Starmer, C., and Sugden, R. (1994). The nature of salience: An experimental investigation of pure coordination games. *American Economic Review*, 84(3):658.

Nagel, R. (1995). Unraveling in guessing games: An experimental study. *American Economic Review*, 85(5):1313–1326.

O'Neill, B. (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*, 84:2106–2109.

Penczynski, S. P. (2016). Strategic thinking: The influence of the game. *Journal of Economic Behavior & Organization*, 128:72–84.

Rubinstein, A. (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, A. and Tversky, A. (1993). Naive strategies in zero-sum games. Working paper 17-93, The Sackler Institute of Economic Studies.

Rubinstein, A., Tversky, A., and Heller, D. (1997). Naive strategies in competitive games. In Albers, W., Güth, W., Hammerstein, P., Moldovanu, B., and van Damme, E., editors, *Understanding Strategic Interaction–Essays in Honor of Reinhard Selten*, pages 394–402. Springer-Verlag.

Scharlemann, J., Eckel, C., Kacelnik, A., and Wilson, R. (2001). The value of a smile: Game theory with a human face. *Journal of Economic Psychology*, 22(5):617–40.

Schelling, T. C. (1960). *The Strategy of Conflict.* Harvard University Press, Cambridge, Massachusetts.

Stahl, D. O. and Wilson, P. W. (1994). Experimental evidence on players' models of other players. *Journal of Economic Behavior & Organization*, 25(3):309–327.

Wolff, I. (2015). Foundations of strategic thinking and strategic behaviour. Unpublished.

Wolff, I. (2016). Elicited salience and salience-based level-k. *Economics Letters*, 141:134–137.

Wolff, I. and Bauer, D. (2018). Elusive beliefs: Why uncertainty leads to stochastic choice and errors. TWI Research Paper Series, No. 111.

Appendix A Full data

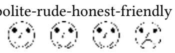
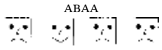
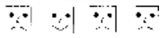
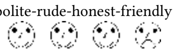
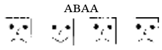
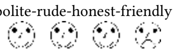
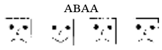
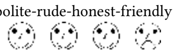
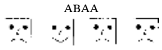
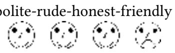
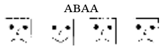
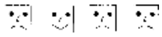
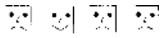
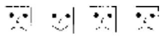
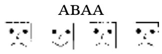
Player role	frame	location 1	location 2	location 3	location 4
coordinators		86	0	10	4
	polite-rude-honest-friendly	6	54	12	28
		6	6	14	74
	ABAA	14	72	13	1
		6	88	6	0
	hate-detest-love-dislike	2	6	88	4
	1-2-3-4	38	17	29	15
	AABA	5	27	54	14
discoordinators		39	14	18	29
	polite-rude-honest-friendly	28	20	32	20
		17	27	23	33
	ABAA	18	21	38	24
		17	40	29	15
	hate-detest-love-dislike	16	29	26	29
	1-2-3-4	21	32	30	17
	AABA	26	24	32	18
	Ace-2-3-Joker	31	17	21	31
	BAAA	34	23	19	23
	AAAB	31	22	18	29
hiders		23	23	43	11
	polite-rude-honest-friendly	15	26	51	8
		21	26	34	19
	ABAA	15	29	33	23
		15	40	34	11
	hate-detest-love-dislike	11	23	38	28
	1-2-3-4	25	22	36	18
	AABA	22	35	19	25
seekers		29	24	42	5
	polite-rude-honest-friendly	8	40	40	11
		7	25	34	34
	ABAA	9	21	53	17
		16	55	21	8
	hate-detest-love-dislike	20	21	55	14
	1-2-3-4	20	18	48	14
	AABA	13	51	21	15
to-your-right players		15	30	32	24
	polite-rude-honest-friendly	22	22	33	24
		18	22	33	27
	ABAA	15	19	34	32
		16	20	33	31
	hate-detest-love-dislike	23	17	30	30
	1-2-3-4	17	21	39	23
	AABA	22	23	29	26
	Ace-2-3-Joker	25	23	31	21
	BAAA	15	26	34	25
	AAAB	24	23	28	25
Colonel-Blotto players		30	26	26	18
	polite-rude-honest-friendly	26	25	26	22
		24	26	28	22
	ABAA	23	29	24	24
		25	26	26	23
	hate-detest-love-dislike	25	23	30	21
	1-2-3-4	24	25	26	25
	AABA	24	25	28	23
	Ace-2-3-Joker	26	26	24	24
	BAAA	29	25	26	20
	AAAB	24	26	25	24

Table A.1: Full data of the games (relative choice frequencies; for Colonel Blotto: average proportion of troops).

Task	frame	location 1	location 2	location 3	location 4
BettingTask		25	29	36	11
	polite-rude-honest-friendly	12	8	53	27
		16	34	35	16
	ABAA	13	33	41	13
		6	58	26	11
	hate-detest-love-dislike	7	12	69	12
	1-2-3-4	18	21	36	25
	AABA	13	36	36	15
	Ace-2-3-Joker	33	14	19	34
	BAAA	27	25	24	24
	AAAB	18	21	34	26
HiderBettingTask		40	19	19	22
	polite-rude-honest-friendly	12	17	50	21
		24	17	28	31
	ABAA	9	53	21	17
		14	59	16	12
	hate-detest-love-dislike	12	14	53	21
	1-2-3-4	19	28	31	22
	AABA	26	19	40	16
	Ace-2-3-Joker	28	14	29	29
	BAAA	31	24	19	26
	AAAB	16	22	22	40
RepresentRating†		10	32	27	31
	polite-rude-honest-friendly	34	7	26	32
		38	32	27	3
	ABAA	38	5	30	28
		29	19	30	21
	hate-detest-love-dislike	34	15	20	30
	1-2-3-4	33	23	16	28
	AABA	34	23	8	35
	Ace-2-3-Joker	39	15	17	29
	BAAA	11	35	25	29
	AAAB	43	25	24	8
SalienceRating†		94	2	4	0
	polite-rude-honest-friendly	14	57	21	8
		5	6	8	81
	ABAA	2	91	4	4
		13	72	11	4
	hate-detest-love-dislike	16	19	62	3
	1-2-3-4	38	21	25	16
	AABA	3	4	93	1
	Ace-2-3-Joker	28	3	3	66
	BAAA	92	3	5	0
	AAAB	5	6	3	86

†In case a participant rated several items as most representative/most salient, her count would be evenly distributed on all corresponding locations.

Table A.2: Full data from the complementary tasks (relative choice frequencies; for the Rating tasks: relative frequencies of location ranked the highest).

Appendix B Analysis of the coordination-game data

The coordination-game data is explained best by participants choosing what they see as the most salient option. The top part of Table 2 shows that this is in accordance with the Salience-Lk model but not with Betting-Lk, Bounded-Lk, or the heuristic (I omit the Nash-equilibrium and the LuckyNoEqm models here as they do not make a unique prediction). We know from earlier studies that in pure coordination games, team reasoning often is important (Bardsley et al., 2010). We did not include team reasoning here, as (i) in our coordination games, it would make the same prediction as Salience-Lk, and (ii) it does not help predicting behaviour in any of the other games.

model	fitted on	LogL	MSE	modes predicted	parameters
Standard Lk	coordination	-1027	0.0782	2† out of 8	—
Betting-Lk		-951	0.0616	3 out of 8	2
LuckyOrAnything		-951	0.0600	3 out of 8	—
Bounded Lk		-945	0.0635	3 out of 8	3
Salience-Lk		-885	0.0154	8 out of 8	2

†Expected number of correctly-predicted modes under uniform randomisation.

Table B.1: Performance of the models in terms of data fitting, ordered by log-likelihood.

Appendix C "Neutral non-neutral" frames only

model	fitted on	LogL	MSE	modes predicted	parameters
Betting-Lk	coordination	-627	0.0376	1 out of 3	2
Bounded Lk		-626	0.0389	1 out of 3	3
Salience-Lk		-684	0.0234	3 out of 3	2
LuckyOrAnything†		-684	0.0225	3 out of 3	0
Betting-Lk	discoordination	-2635	0.0033	3 out of 6	3
NashEqm		-2637	0.0036	1.5‡ out of 6	—
Salience-Lk		-2635	0.0032	3 out of 6	3
Bounded Lk		-2632	0.0032	5 out of 6	3
LuckyNoEqm		-2623	0.0027	2 out of 6	3
LuckyOrAnything		-2621	0.0026	4 out of 6	1
NashEqm	hide & seek	-1615	0.0135	(.75,.75)‡ out of (3,3)	—
LuckyNoEqm		-1572	0.0094	(0,1) out of (3,3)	3
Salience-Lk		-1604	0.0225	(0,3) out of (3,3)	5
Bounded Lk		-1579	0.0091	(1,3) out of (3,3)	5
LuckyOrAnything		-1558	0.0071	(3,3) out of (3,3)	1
Betting-Lk		-1540	0.0042	(3,3) out of (3,3)	5

†Model includes a tremble with 1% probability to take care of zero-probability events. The fit improves further when allowing for more randomisation (*e.g.*, 20% randomisation, LogL = -603, mse = 0.0106). ‡Expected number of correctly-predicted modes under uniform randomisation.

Table C.1: Data-fitting performance of the models, order as in the main text (by LogL of the original estimate).

	Spearman coefficient	$p-$value	No. of frames
hiders	0.47	0.118	3
seekers	0.67	0.025	3
discoordinators	0.28	0.185	6
to-your-right players	0.37	0.078	6
Blotto players	0.30	0.149	6

Table C.2: Correlations of ranks: game data and BettingTask data.

	fitted prob(uniform mixing) in %
seekers	0
hiders	59
discoordinators	65
to-your-right players	79
Blotto players	92

Table C.3: Fitted probability of uniform mixing for each player role.

Appendix D Individual-level data

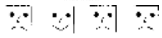
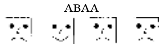
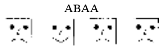
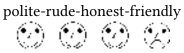
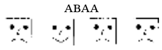
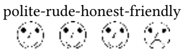
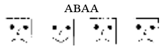
Player role (No. participants)	frame	location 1	location 2	location 3	location 4
discoordinators (82)		34	29	16	21
	polite-rude-honest-friendly	32	26	27	16
		33	27	20	21
	ABAA	24	27	27	22
		24	21	28	27
	hate-detest-love-dislike	22	23	27	28
	1-2-3-4	30	20	17	33
	AABA	24	17	24	34
	Ace-2-3-Joker	29	22	27	22
	BAAA	30	17	26	27
	AAAB	24	20	24	32
hiders (41)		37	22	17	24
	polite-rude-honest-friendly	27	17	41	15
		39	22	22	17
	ABAA	24	24	20	32
		32	20	20	29
	hate-detest-love-dislike	17	32	15	37
	1-2-3-4	24	15	29	32
	AABA	24	27	20	29
	Ace-2-3-Joker	34	20	12	34
	BAAA	32	27	7	34
	AAAB	24	34	20	22
seekers (41)		49	15	12	24
	polite-rude-honest-friendly	15	29	29	27
		29	22	22	27
	ABAA	7	59	10	24
		10	41	34	15
	hate-detest-love-dislike	15	12	59	15
	1-2-3-4	22	17	37	24
	AABA	10	22	46	22
	Ace-2-3-Joker	24	12	22	41
	BAAA	41	20	17	22
	AAAB	15	17	27	41
to-your-right players (82)		22	28	29	21
	polite-rude-honest-friendly	27	26	22	26
		18	26	28	28
	ABAA	10	30	26	34
		9	30	37	24
	hate-detest-love-dislike	20	17	38	26
	1-2-3-4	17	26	28	29
	AABA	18	26	37	20
	Ace-2-3-Joker	27	23	23	27
	BAAA	26	27	28	20
	AAAB	26	22	30	22
BettingTask (82)		23	23	33	21
	polite-rude-honest-friendly	18	24	34	23
		16	18	30	35
	ABAA	16	27	34	23
		17	29	29	24
	hate-detest-love-dislike	18	23	34	24
	1-2-3-4	24	22	30	23
	AABA	22	24	24	29
	Ace-2-3-Joker	24	22	22	32
	BAAA	27	18	30	24
	AAAB	20	23	28	29

Table D.1: Relative choice frequencies (in %), individual-level experiments.

THURGAU INSTITUTE
OF ECONOMICS
at the University of Konstanz

Hauptstrasse 90
Postfach
CH-8280 Kreuzlingen 2

T +41 (0)71 677 05 10
F +41 (0)71 677 05 11

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