RATIONAL HEURISTICS FOR ONE-SHOT GAMES*

Frederick Callaway[†] Thomas L. Griffiths[‡]
Gustav Karreskog Rehbinder [§]

May 9, 2023

Abstract

We present a theory of human behavior in one-shot interactions based on the assumption that people use heuristics that optimally trade off the expected payoff and the cognitive cost. The theory predicts that people's behavior depends on their past experience; specifically, they make choices using heuristics that performed well in previously played games. We confirm this prediction in a large, preregistered experiment. The rational heuristics model provides a strong quantitative account of participant behavior, and outperforms existing models. More broadly, our results suggest that synthesizing heuristic and optimal models is a powerful tool for understanding and predicting economic decisions.

Keywords: Bounded rationality, Experiments, Cognitive cost, Strategic thinking, Game theory, One-shot games, Heuristics

JEL classification: C72, C90, D83, D01

*We thank Drew Fudenberg, Alice Hallman, Benjamin Mandl, Erik Mohlin, Isak Trygg Kupersmidt, Jörgen Weibull, Peter Wikman, and seminar participants at SSE, Uppsala University, UCL, NHH, and Princeton, for helpful comments and insights. This work was supported by the Templeton Foundation, the NOMIS Foundation, the Jan Wallander and Tom Hedelius Foundation, and the Knut and Alice Wallenberg Research Foundation.

[†]Department of Psychology, Princeton University, Princeton, NJ 08540; fredcallaway@princeton.edu

[‡]Department of Psychology, Princeton University, Princeton, NJ 08540; tomg@princeton.edu

[§]Department of Economics, Uppsala University, Kyrkogårdsgatan 10, 751 20, Uppsala, Sweden; gustav.karreskog@nek.uu.se

1 Introduction

A key assumption underlying classical economic theory is that people behave optimally in order to maximize their subjective expected utility (Savage, 1954). However, a large body of work in behavioral economics shows that human behavior systematically deviates from this rational benchmark in many settings (Dhami, 2016). This suggests that we can improve our understanding of economic behavior by incorporating more realistic behavioral components into our models. While many of these deviations are indeed systematic and show up in multiple studies, the estimated biases vary considerably across studies and contexts. Apparent biases change or even disappear if participants have opportunities to learn or if the details of the decision task change. For example, this is the case with the endowment effect (Tunçel and Hammitt, 2014), loss aversion (Ert and Erev, 2013), numerosity underestimation (Izard and Dehaene, 2008), and present bias (Imai et al., 2020).

In order to incorporate behavioral effects into theories with broader applications—without having to run new experiments for each specific setting—we need a theory that can account for this variation. That is, we need a theory that can help us understand why—and predict when—people deviate from the rational benchmark. In this paper, we propose such a theory based on the idea that people use simple decision procedures, or *heuristics*, that are optimized to the environment to make the best possible use of their limited cognitive resources and thereby maximize utility. This allows us to predict behavior by analyzing which heuristics perform well in which environments. This paper presents an explicit instantiation of this theory tailored to one-shot games and tests it experimentally.

In situations where people play the same game multiple times against different opponents, and hence there is an opportunity to learn, both theoretical and experimental work suggests that Nash equilibria can often yield sensible long-run predictions (Fudenberg et al., 1998; Camerer, 2003). However, in experimental studies of one-shot games where players don't have experience with the particular game at hand, people seldom follow the theoretical prediction of Nash equilibrium play (see Crawford et al., 2013, for an overview). Consequently, we need an alternative theory for strategic interactions that happen only once (or infrequently).

The most common theories of behavior in one-shot games in the literature assume that players perform some kind of iterated reasoning to form beliefs about the other player's action and then select the best action in response. This includes level-k (Nagel, 1995; Stahl and Wilson, 1994, 1995), cognitive hierarchy (Camerer et al., 2004), and

noisy introspection models (Goeree and Holt, 2004). In such models, participants are characterized by different levels of reasoning. Level-0 reasoners behave naively by playing a uniformly random strategy. Level-1 reasoners best respond to level-0 behavior, while higher-level reasoners best respond to behavior based on lower-level reasoning. In meta-analyses such as Crawford et al. (2013), Wright and Leyton-Brown (2017), and Fudenberg and Liang (2019), variations of these iterated reasoning models best explain human behavior.

All iterated reasoning models assume the basic structure of belief formation and best responding to those beliefs. However, such a belief-formation and best-response process is often inconsistent with empirical evidence. For example, Costa-Gomes and Weizsäcker (2008) found that participants who were asked to state their beliefs about how the opponent would play, often failed to play a best response to those beliefs. Moreover, eye-tracking studies have revealed that the order in which participants attend to payoffs in visually presented normal-form games is inconsistent with a belief-formation and best-response process (Polonio et al., 2015; Devetag et al., 2016; Stewart et al., 2016). Furthermore, the estimated parameters of iterated reasoning models often vary considerably across different data sets (Wright and Leyton-Brown, 2017), behavior depends on aspects of the game that these models do not take into account (Bardsley et al., 2010; Heap et al., 2014), and there is evidence that games played previously have an effect on behavior, which the above static models fail to capture (Mengel and Sciubba, 2014; Peysakhovich and Rand, 2016).

In this paper, we present a theory of human behavior in one-shot games based on the rational use of heuristics (Lieder and Griffiths, 2017, 2020). That is, we assume that people use simple cognitive strategies that flexibly and selectively process payoff information to take good decisions with minimal cognitive effort. Concretely, we assume that people use heuristics that maximize expected payoff minus cognitive cost. Importantly, this optimization happens at the level of the environment; although people might not choose the best action in a given game, they will learn which heuristics generally work well (cf. *procedural rationality* in Simon, 1976).

Thus, our approach combines two perspectives on human decision-making, embracing both the notion that human behavior is adaptive in a way that can be described as optimizing and the notion that people use simple strategies that are effective for the problems they actually need to solve. The key assumption of this *resource-rational analysis* approach is that people use cognitive strategies that make optimal use of their limited computational resources (Lieder and Griffiths, 2020; Griffiths et al., 2015; cf. Howes et al., 2009; Lewis et al., 2014; Gershman et al., 2015).

It is instructive to compare resource-rational analysis with two other approaches to explaining observed deviations from perfectly rational behavior: the information-theoretic and ecological rationality approaches. Like *information-theoretic* approaches such as rational inattention (Matějka and McKay, 2015; Sims, 1998; Caplin and Dean, 2013; Hebert and Woodford, 2019; Steiner et al., 2017), the resource-rational approach assumes that the costs and benefits of information processing are optimally traded off. However, while information-theoretic approaches typically assume domain-general cost functions (e.g., based on entropy reduction), the resource-rational approach typically makes stronger assumptions about the specific computational processes and costs that are likely to be involved in a given domain. In this way, the resource-rational approach is more similar to the *ecological rationality* approach, a framework based on the idea that people use computationally frugal heuristics, which are highly effective for the kinds of problems that people actually encounter (Gigerenzer and Todd, 1999; Goldstein and Gigerenzer, 2002; Todd and Gigerenzer, 2012). For example, if the other players in an environment are using a wide variety of decision strategies, then a heuristic that ignores the other players' payoffs entirely may perform best (Spiliopoulos and Hertwig, 2020). However, while proponents of ecological rationality explicitly reject the notion of optimization under constraints (e.g., Gigerenzer and Todd, 1999, Ch. 1), optimization is at the heart of resource rationality. This makes it possible to predict when people will use one heuristic rather than another (Lieder and Griffiths, 2017) and even to discover novel heuristics (Lieder et al., 2017; Krueger et al., 2022).

One important commonality between our approach and ecological rationality is the recognition that the quality or adaptiveness of a heuristic depends on the environment in which it is used. For example, in an environment in which most interactions are characterized by competing interests (e.g., zero-sum games), a good heuristic is to look for actions with high guaranteed payoffs. On the other hand, if most interactions have common interests, a better heuristic might be to look for outcomes that would be good for everyone (cf. Spiliopoulos and Hertwig, 2020). Our theory thus predicts that people will use different heuristics in cooperative vs. competitive environments.

To test our theory's prediction that people adapt their heuristics to the environment, we conduct a large, preregistered[1] behavioral experiment. In our experiment, participants play a series of normal-form games in one of two environments characterized by different correlations in payoffs. In the *common-interests* environment, there is a positive correlation between the payoffs of the two players over the set of strategy profiles; i.e., outcomes that are good for one player tend to be good for the other as

[1] `https://osf.io/hcnzg`

well. In the *competing-interests* environment, the payoff correlation is negative; i.e., one player's loss is the other's gain, which is essentially a soft version of zero-sum games. Interspersed among these treatment-specific games, we include four *comparison games* that are the same for both treatments. If the participants are using environment-adapted heuristics to make decisions, and different heuristics are good for common-interests and competing-interests environments, the participants should behave differently in the comparison games since they are employing different heuristics. Indeed, this is what we observe.

To provide further support for the claim that participant behavior is consistent with an optimal tradeoff between the expected payoff and the cognitive cost, we define two parameterized families of heuristics and cognitive costs that can make quantitative predictions about the distribution of play in each game. However, rather than identifying the parameters that best fit human behavior (as is commonly done in model comparison), we instead identify the parameters that strike an optimal tradeoff between expected payoffs and cognitive costs, and ask how well they predict human behavior. Although we fit the cost function parameters that partially define the resource-rational heuristic, these parameters are fit jointly to data in both treatments. Strikingly, we find that this model, which has no free parameters that vary between the treatments, achieves nearly the same out-of-sample predictive accuracy as the model with all parameters fit separately to each treatment. Both the optimized and fitted versions of this model predicted the modal action with an accuracy of 88%, compared to 80% for a quantal cognitive hierarchy model.

In Section 2, we provide an overview of our theory and present a stylized example to illustrate how resource-rational heuristics depend on the structure of the environment. In Section 3, we present our experiment and model-free analyses, which demonstrate a strong causal link between previous experience and current behavior in one-shot games. In Section 4, we provide a more detailed description of the theory and introduce two different parameterized models, an interpretable, low-parameter model, and a black-box neural network model. Using these models, we demonstrate in Section 5 that the differences in behavior can be accurately predicted out-of-sample by assuming that participants use the optimal heuristics for the respective environments. In Section 5.4, we compare our models to alternative models, including quantal cognitive hierarchy and prosocial preference models, and show that our models provide better predictions of behavior than these alternatives.

2 Theory Overview and Stylized Predictions

The central tenet of our theory is that individuals use heuristics that maximize the expected payoff minus the cognitive cost in a given environment. This can be summarized in the following equation:

$$h^* = \underset{h_i \in \mathcal{H}}{\operatorname{argmax}} \, \mathbb{E}_{\mathcal{E}} \left[\pi_i \left(h_i(G), h_{-i}(G^T) \right) - c(h_i) \right]. \tag{1}$$

Here, $h_i \in \mathcal{H}$ is a heuristic, G is a game, and $h_i(G)$ is the distribution of play produced by applying heuristic h_i to game G. The optimal heuristic h^* is the one that maximizes the expected payoff $\pi_i \left(h_i(G), h_{-i}(G^T) \right)$ minus the cognitive cost $c(h_i)$, where $h_{-i}(G^T)$ gives the distribution of play by the opponent. The expectation is taken with respect to an environment $\mathcal{E}$, which defines a distribution over possible games G and opponent heuristics h_{-i}.

A key implication of this theory is that the heuristics we expect people to use depend on the types of games and opponents they encounter frequently; that is, h_i^* depends on $\mathcal{E}$. Below, we illustrate this idea with a simple example.

Consider two possible environments: one consisting entirely of coordination games (where the players want to coordinate on the same action), and one consisting entirely of constant-sum games (where the players' interests are exactly opposed). In both environments, all other players follow a heuristic where they pick the strategy with the highest average payoff (level-1 in the language of level-k reasoning). Now consider what you would do as the row player when faced with the following games from each environment.

$8, 8$	$0, 0$
$0, 0$	$9, 6$

Coordination game

$5, 4$	$2, 7$
$3, 6$	$3, 6$

Constant-sum game

In the coordination game, the column player will select column **1** because 8 is larger than 6; row **1** is thus the optimal play. In the constant-sum game, the column player will select column **2** because $7 + 6 > 4 + 6$; thus, row **2** is the optimal play. Clearly, simulating the other player as we have done here will always lead to the optimal choice. However, in each case, the optimal action could also be found by a simpler, less cognitively demanding heuristic. In the coordination game, the best action is the one that produces the outcome with the highest minimum value for each player (we will later call this the "jointmax heuristic"). In the constant-sum game, the best action is

the one that has the highest guaranteed payoff (the "maximin" heuristic).

The central claim of our theory is that people will use heuristics that identify good actions with minimal cognitive cost. Critically, a "good" action is one that achieves high payoffs on average across all the games a person encounters. Thus, if we take one person, "Lucy," and we put her in an environment where she repeatedly plays games like the one on the left, she will learn to use the jointmax heuristic because it usually selects the same action as simulation, but with less cognitive cost. If we put another person, "Rodney," in an environment where he repeatedly plays games like the one on the right, he will learn to use the maximin heuristic for the same reason. Now consider what actions each will select in a new game:

$7, 7$	$0, 9$
$9, 0$	$4, 4$

Prisoner's dilemma game

Here, the second action strictly dominates the first, and so it has to be the choice of a perfectly rational decision-maker. Rodney will play this action, as it is selected by the maximin heuristic, which has performed well in his previous experience. Importantly, he may choose this action without ever realizing that it dominates the other. By contrast, Lucy will be likely to play the first, "incorrect" action, as it is selected by the jointmax. She makes this mistake because identifying the outcome that is best for both players is easy, and it has worked well for her before. Although she might have fared better on this specific game if she had simulated the possible outcomes of each action, the cognitive cost of such an approach would not be justified by the relatively small increase in payoff across the full set of games she has played.

To summarize, the principled but costly approach of simulating the other player in order to select one's own action can sometimes be approximated by simpler heuristic strategies. When this approximation is sufficiently accurate, a resource-rational agent will use the heuristic to avoid the mental effort of simulation. But if we present the unwitting agent with a new game that lacks the structure the heuristic was taking advantage of, the agent will make predictable errors. This is the key intuition underlying our behavioral experiment.

3 Experiment

Our overarching hypothesis is that individuals choose actions in one-shot games using heuristics that optimally trade off between the expected payoff and the cognitive cost. Critically, as discussed above, this optimization occurs with respect to an environment rather than a single game. This results in a central prediction: the action a player takes in a given game will depend not only on the nature of that particular game but also on the other games she has previously played. From this central prediction, we derived four hypotheses, which we tested in a large, preregistered online experiment.

3.1 Methods

We recruited 600 participants on Amazon Mechanical Turk using the oTree platform (Chen et al., 2016). Each participant was assigned to one of 20 populations of 30 participants. They then played 50 different one-shot normal-form games, with each participant randomly matched to another player from their population after each game.[2]

Each population was assigned to one of two experimental treatments, which determined the distribution of games played. Specifically, we manipulated the correlation between the row and column players' payoffs in each cell (cf. Spiliopoulos and Hertwig, 2020). In the *common-interests* treatment, the payoffs were positively correlated, such that a cell with a high payoff for one player was likely to have a high payoff for the other player as well. By contrast, in the *competing-interests* treatment, the payoffs were negatively correlated, such that a cell with a high payoff for one player was likely to have a low payoff for the other player. Concretely, the payoffs in each cell were sampled from a bivariate Normal distribution truncated to the range $[0, 9]$ and discretized such that all payoffs were single-digit nonnegative integers.[3] Examples of each type of *treatment game* are shown in Tables 1 and 2.

$5,6$	$6,4$	$5,3$
$9,4$	$5,5$	$6,7$
$2,0$	$0,1$	$6,4$

$3,4$	$5,5$	$9,7$
$4,2$	$5,7$	$5,7$
$2,4$	$2,1$	$2,3$

$9,7$	$5,9$	$7,8$
$6,7$	$9,9$	$4,6$
$6,4$	$3,1$	$6,2$

Table 1: Three games from the common-interests treatment.

[2]To facilitate running the experiment online, we used an asynchronous scheme in which participants could play "against" an opponent who had played the game earlier. Participants were informed of this; see Figure 8 in Appendix A.

[3]The normal distribution is given by $N((5,5),\Sigma)$ with $\Sigma = 5\begin{pmatrix}1 & \rho\\ \rho & 1\end{pmatrix}$, where $\rho = 0.9$ in the common-interests treatment and $\rho = -0.9$ in the competing-interests treatment.

$5,5$	$6,2$	$5,3$
$5,3$	$1,8$	$8,4$
$3,6$	$7,4$	$4,6$

$2,4$	$4,4$	$4,6$
$1,7$	$2,6$	$9,1$
$7,1$	$4,8$	$8,6$

$4,5$	$1,5$	$7,1$
$2,7$	$8,5$	$5,7$
$2,6$	$8,3$	$3,9$

Table 2: Three games from the competing-interests treatment.

For each population, we sampled 46 treatment games, with each participant playing every game once. The remaining four games were *comparison games*, i.e., treatment-independent games that we used to compare differences in the participants' behavior between the two treatments. The comparison games were played in periods 31, 38, 42, and 49. We located these comparison games later in the experiment so that the participants would have time to adjust to the treatment environment first, while leaving intervals between them to minimize the chance that participants would notice that these games were different from the others they had played.

3.1.1 The Comparison Games

We selected four comparison games that we expected to elicit dramatically different distributions of play in the two treatments. In these games, there is a tension between choosing a row with an efficient outcome or choosing a row with a high guaranteed payoff. For two of the games, the efficient outcome was also a Nash equilibrium (NE), and for the other two games, the efficient outcome was not a NE.

$8,8$	$2,6$	$0,5$
$6,2$	$6,6$	$2,5$
$5,0$	$5,2$	$5,5$

Comparison game 1

$8,8$	$2,9$	$1,0$
$9,2$	$3,3$	$1,1$
$0,1$	$1,1$	$1,1$

Comparison game 2

$4,4$	$4,6$	$5,0$
$6,4$	$3,3$	$5,1$
$0,5$	$1,5$	$9,9$

Comparison game 3

$4,4$	$9,1$	$1,3$
$1,9$	$8,8$	$1,8$
$3,1$	$8,1$	$3,3$

Comparison game 4

Table 3: The four comparison games.

The first game is a weak-link game, where all the diagonal strategy profiles are Nash equilibria, but each has a different efficiency. The most efficient NE yields the payoffs (8,8), but it is also possible to get 0. The least efficient equilibrium yields the

payoffs (5,5), but 5 is also the guaranteed payoff. The equilibrium (6,6) is in between the aforementioned payoffs in terms of both risk and efficiency. The third row has the highest average payoff and is the best response to itself, and so any standard recursive reasoning model would predict (5,5) being the outcome.

The second comparison game is a normal prisoner's dilemma game, with an added dominated and inefficient strategy. In this game, strategy **2** dominates the other strategies. However, we still expect strategy **1** to be played more often in the common-interests treatment since, overall, it is a good heuristic to look for efficient outcomes in that environment.

The third comparison game is a game with two NE, where one is the pure NE with both players playing strategy **3**, and the other is a mixed NE involving **1** and **2**. This game is constructed so that the row averages are much higher for strategies **1** and **2** than for **3**, meaning that any level-k heuristic would result in strategy **1** or **2** being played, while the NE yielding $(9, 9)$ is much more efficient. Thus, there is a strong tension between the efficient payoff and the guaranteed payoff.

In the fourth comparison game, the risky efficient outcome $(8, 8)$ is not a NE. A standard level-k player of any level higher than 0 would play strategy **3**.

3.2 Model-free Results

We organize our results based on four preregistered hypotheses. The first two are model-free and concern behavior in the comparison games; they are presented here. The next two are model-based and concern behavior in the treatment games; these will be presented later.

Our first hypothesis is that the treatment environment has an effect on behavior in the comparison games.

Hypothesis 1. *The distribution of play in the four comparison games will differ between the two treatments.*

This hypothesis follows from the assumption that people learn to use heuristics that are adaptive within their treatment and that different heuristics are adaptive across the two treatments. Figure 1 visually confirms this prediction, and Table 4 confirms that these differences are statistically significant (χ^2-tests, as preregistered).

Inspecting Figure 1, we see that the distribution of play is not just different between the two groups; it is different in a systematic way. In particular, players in the common-interests treatment tend to coordinate on the efficient outcome, even in games 2 and 4, where the efficient outcome is not a Nash equilibrium. We expected this divergence

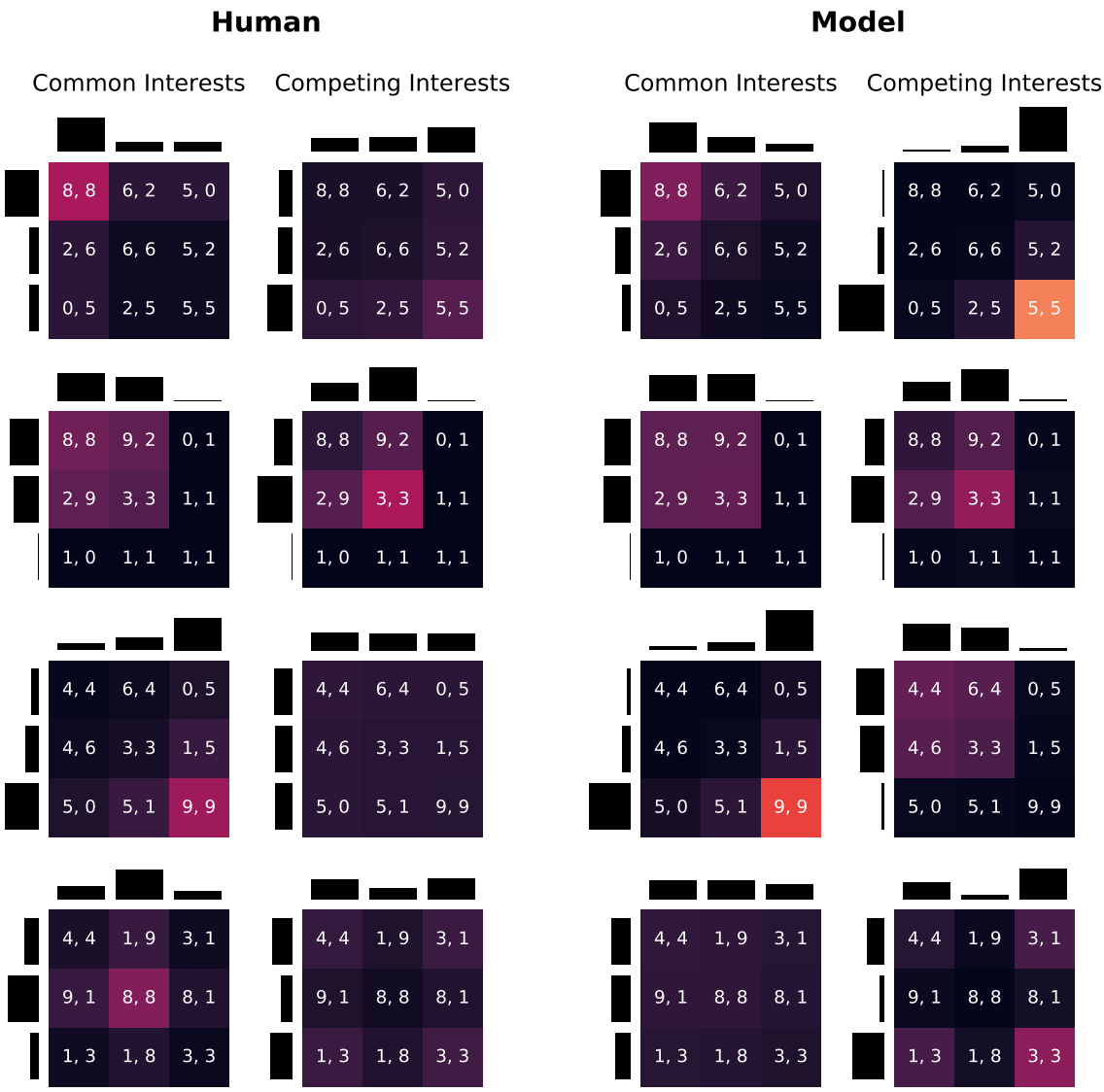


Figure 1: Distribution of play in the four comparison games. Each panel shows the joint and marginal distributions of row/column plays in a single game. The cells show the players' payoffs for the given outcome. The two columns to the left show the actual behavior in the two environments, while the two columns to the right show the predictions of the rational (optimized) metaheuristics.

	Frequencies				
	1	**2**	**3**	χ^2	p-value
Comparison Game 1				98.39	$p < .001$
Common interests	193	53	54		
Competeting interests	75	82	143		
Comparison Game 2				22.08	$p < .001$
Common interests	160	139	1		
Competeting interests	103	195	2		
Comparison Game 3				61.75	$p < .001$
Common interests	40	73	187		
Competeting interests	106	97	97		
Comparison Game 4				91.36	$p < .001$
Common interests	78	173	49		
Competeting interests	115	62	123		

Table 4: χ^2 tests for each comparison game. The results are significant at the preregistered 0.05 level.

in behavior when we constructed the comparison games, which motivates our second hypothesis.

Hypothesis 2. *The average payoff in the four comparison games will be higher in the common-interests treatment than in the competing-interests treatment.*

The comparison games were designed to create tension between efficiency and risk, with one outcome having a high payoff for both players, but requiring each player to take an action that could yield a very low payoff. We expected that the common-interests players would be more likely to coordinate on the efficient outcome, and therefore receive higher payoffs. The model makes this prediction because identifying mutually beneficial outcomes is typically an effective heuristic in common-interests games, while identifying high guaranteed payoffs is an effective heuristic in competing-interests games. Table 5 confirms this prediction. The common-interests players had a higher average payoff in all four comparison games, and the difference is statistically significant in each case (at the preregistered level of $p < .05$).

4 Theoretical Framework

We have seen that different environments lead to drastically different behavior in the comparison games. While these results are suggestive of rational adaptation, they do not directly imply that the participants are using heuristics in an optimal way. One way to strengthen this claim is to show that one can accurately predict human behavior

	Treatment average payoff			
	Common interests	Competing interests	t-value	p-value
Comparison game 1	5.09	3.64	6.851	$p < .001$
Comparison game 2	5.52	4.04	6.28	$p < .001$
Comparison game 3	5.00	4.31	2.86	$p = 0.004$
Comparison game 4	5.19	3.42	7.21	$p < .001$

Table 5: Two-sided t-tests for the difference in the average payoff between the two treatments in the comparison games.

by assuming optimal use of heuristics. In order to do this, we need to specify concrete parameterizations of the space of heuristics $\mathcal{H}$ and the cognitive costs. But first, we provide a more detailed description of the general theory.

4.1 General Model

We consider a setting where individuals in a population are repeatedly randomly matched with another individual to play a finite normal-form game. We assume they use some heuristic to decide what strategy to play.

Let $G = \langle \{1, 2\}, S_1 \times S_2, \pi \rangle$ be a two-player normal-form game with pure strategy sets $S_i = \{1, \ldots, m_i\}$ for $i \in \{1, 2\}$, where $m_i \in \mathbb{N}$. A mixed strategy for player i is denoted $\sigma_i \in \Delta(S_i)$. The *material payoff* for player i from playing pure strategy $s_i \in S_i$ when the other player $-i$ plays strategy $s_{-i} \in S_{-i}$ is denoted $\pi_i(s_i, s_{-i})$. We extend the material payoff function to the expected material payoff from playing a mixed strategy $\sigma_i \in \Delta(S_i)$ against the mixed strategy $\sigma_{-i} \in \Delta(S_{-i})$ with $\pi_i(\sigma_i, \sigma_{-i})$, in the usual way. A heuristic is a function that maps a game to a mixed strategy $h_i(G) \in \Delta(S_i)$. For simplicity, we will always consider the games from the perspective of the row player, and consider the transposed game $G^T = \langle \{2, 1\}, S_2 \times S_1, (\pi_2, \pi_1) \rangle$ when talking about the column player's behavior.

Each heuristic has an associated cognitive cost[4], $c(h) \in \mathbb{R}_+$. Simple heuristics, such as playing the uniformly random mixed strategy, have low cognitive costs, while complicated heuristics involving many precise computations have high cognitive costs. Since a heuristic returns a mixed strategy, the expected material payoff for player i using heuristic h_i when player $-i$ uses heuristic h_{-i} is

$$\pi_i \left(h_i(G), h_{-i}(G^T) \right).$$

[4]In general, the cognitive cost could depend on both the heuristic and the game. For example, it might be more costly to apply a heuristic to a 5×5 game than to a 2×2 game. But since all our games are 3×3, we can dispense with that dependency.

Since each heuristic has an associated cognitive cost, the actual expected utility derived from it is

$$u_i\left(h_i, h_{-i}, G\right) = \pi_i\left(h_i(G), h_{-i}(G^T)\right) - c(h_i).$$

A heuristic is neither good nor bad in isolation; its performance has to be evaluated with regard to some environment, in particular, with regard to the games and other-player behavior one is likely to encounter. Let $\mathcal{G}$ be the set of possible games in the environment, $\mathcal{H}$ be the set of heuristics the other player could use, and P be the joint probability distribution over $\mathcal{G}, \mathcal{H}$. In the equations below, we will assume that $\mathcal{G}$ and $\mathcal{H}$ are countable. An environment is given by $\mathcal{E} = (P, \mathcal{G}, \mathcal{H})$. Thus, an environment describes which game and other-player heuristic combinations a player is likely to face. Given an environment, we can calculate the expected performance of a heuristic as follows:

$$V(h_i, \mathcal{E}) = \mathbb{E}_{\mathcal{E}}\left[u_i\left(h_i, h_{-i}, G\right)\right] = \sum_{G \in \mathcal{G}} \sum_{h_{-i} \in \mathcal{H}} u_i\left(h_i, h_{-i}, G\right) \cdot P(G, h_{-i}). \tag{2}$$

We can also calculate the expected performance of a heuristic conditional on the specific game being played as follows:

$$V(h_i, \mathcal{E}, G) = \mathbb{E}_{\mathcal{E}|_G}\left[u_i\left(h_i, h_{-i}, G\right)\right] = \sum_{h_{-i} \in \mathcal{H}} u_i\left(h_i, h_{-i}, G\right) \cdot P(h_{-i} \mid G).$$

We can now formally define what it means for a heuristic to be rational (or optimal). A rational heuristic h^* is a heuristic that optimizes (2), i.e.,

$$h^* = \underset{h_i \in \mathcal{H}}{\operatorname{argmax}}\, V(h_i, \mathcal{E}), \tag{3}$$

or, in slightly expanded form,

$$h^* = \underset{h_i \in \mathcal{H}}{\operatorname{argmax}}\, \mathbb{E}_{\mathcal{E}}\left[\pi_i\left(h_i(G), h_{-i}(G^T)\right) - c(h_i)\right]. \tag{4}$$

That is, a rational heuristic chooses actions that yield high rewards for the games and opponents one tends to encounter, while not being costly to evaluate; more specifically, a rational heuristic achieves the best tradeoff between these two (typically, but not always, competing) desiderata. We here also see that by varying the environment, $\mathcal{E}$, we can vary which heuristics are optimal. In our experiment, we will manipulate the distribution over games, thereby varying the predictions we get by assuming rational

heuristics.

One natural critique of this approach is that the problem of selecting an optimal heuristic is actually much more complex than the problem of selecting an optimal action. Critically, however, while the optimality of an action is defined with respect to a single game, the optimality of a heuristic is defined with respect to an environment. Thus, it is possible to for a player *learn* an optimal heuristic (but not an optimal action) even if she has limited experience with the specific game being played. In Appendix E, we show that a simple learning model can reproduce the performance of the optimizing metaheuristic model.

4.2 Specific Parameterizations

We consider two parameterizations of $\mathcal{H}$ and c. Importantly, we don't claim that either parameterization perfectly matches the actual spaces of heuristics or cognitive costs faced by human beings. Instead, they are constructed to be rich enough, and close enough to actual cognitive costs and heuristics, to be able to capture the essence of behavior. As we will see, the optimal heuristics found with both these parameterizations give accurate predictions.

The first parameterization we call *metaheuristics*. It consists of three primitive heuristics that together with a selection rule create the metaheuristic. The primitive heuristics and selection rule are chosen based on existing models and descriptive evidence on choice processes. This parameterization is intuitive and interpretable, but its design involves many somewhat arbitrary researcher decisions.

The second parameterization, *deep heuristics*, makes much weaker assumptions about the space of heuristics. This parameterization is based on the neural network architecture for normal-form games proposed by Hartford et al. (2016). It captures a much larger space of possible heuristics and thus removes the researcher degrees of freedom that are a concern for the first parameterization (e.g., the choice of primitive heuristics), at the cost of losing interpretability and some control over the cognitive cost.

4.2.1 Metaheuristics

To build a formal model of heuristics for one-shot games, we begin by specifying a few general types of reasoning that such heuristics might employ: row-based reasoning, cell-based reasoning, and simulation-based reasoning. For each of these types, we specify a precise functional form with a small number of continuous parameters and an associated

cognitive cost function. The cognitive cost of a heuristic is a function of its parameters, and the cost function is itself parameterized. Finally, we consider a higher-order heuristic, which we call a *metaheuristic*, that selects among the candidate first-order heuristics based on their expected values for the current game. We emphasize that we do not claim that this specific family captures all the heuristics people might employ in a game. However, we hypothesized—and our results confirm—that this family is expressive enough to illustrate the general theory's predictions and provide a strong quantitative account of human behavior. Since this specific parameterization (metaheuristics) is not the main focus of the paper, the details can be found in Appendix B.

Row Heuristics A *row heuristic* calculates a value, $v(s_i)$, for each pure strategy, $s_i \in S_i$, based only on the player's own payoffs associated with s_i. Formally, a row heuristic is defined by the row-value function v such that

$$v(s_i) = f(\pi_i(s_i, \mathbf{1}), \ldots, \pi_i(s_i, m_i)))$$

for some function $f : \mathbb{R}^{m_{-i}} \to \mathbb{R}$. The specific parameterization of f we consider goes from min to max, passing by the mean, via a single parameter γ such that

$$v^{\gamma}(s_i) = \sum_{s_{-i} \in S_{-i}} \pi_i(s_i, s_{-i}) \cdot \frac{\exp\left[\gamma \cdot \pi_i(s_i, s_{-i})\right]}{\sum_{s \in S_{-i}} \exp\left[\gamma \cdot \pi_i(s_i, s)\right]}.$$

For example, $\gamma = 0$ implies that f is the mean function, corresponding to a level-1 strategy. $\gamma \to -\infty$ corresponds to the min function, and yields the maximin strategy. More generally, the row heuristic captures a weighted mean of each row's payoffs that might overweight either good or bad outcomes.

Once the function v is specified, we assume that the computation of v is subject to noise but that this noise can be reduced through cognitive effort, which we operationalize as a single scalar φ. In particular, following Stahl and Wilson (1994), we assume that the noise is Gumbel-distributed and thus recover a multinomial logit model with the probability that player i plays strategy s_i being

$$h^{s_i}_{\text{row}}(G) = \frac{\exp\left[\varphi \cdot v(s_i)\right]}{\sum_{k \in S_i} \exp\left[\varphi \cdot v(k)\right]}.$$

Naturally, the cost of a row heuristic is a function of cognitive effort. Specifically,

we assume that the cost is proportional to effort,

$$c(h_{\text{row}}) = \varphi \cdot C_{\text{row}},$$

where $C_{\text{row}} > 0$ is a free parameter of the cost function.

Cell Heuristics An individual might not necessarily consider all aspects connected to a strategy but find a good "cell," meaning a payoff pair $(\pi_i(s_i, s_{-i}), \pi_{-i}(s_i, s_{-i}))$. In particular, previous research has proposed that people sometimes adopt a *team view*, where each player looks for outcomes that are good for both players, and chooses actions under the (perhaps implicit) assumption that the other player will try to achieve this mutually beneficial outcome as well (Sugden, 2003; Bacharach, 2006). Alternatively, people may engage in *virtual bargaining*, where each player selects the outcome that would be agreed upon if she could negotiate with the other player (Misyak and Chater, 2014). Importantly, these approaches share the assumption that people reason directly about outcomes (rather than actions) and that there is some amount of assumed cooperation.

We refer to heuristics that reason directly about outcomes as *cell heuristics*. Based on preliminary analyses, we identified one specific form of cell heuristic that participants appear to use frequently: the *jointmax* heuristic, which identifies the outcome that is most desirable for both players. Formally, the joint desirability of a cell is given by

$$v^{\text{jointmax}}(s_i, s_{-i}) = \min\left\{\pi_i(s_i, s_{-i}), \pi_{-i}(s_i, s_{-i})\right\}$$

and the probability of playing a given strategy, with cognitive effort φ is given by

$$h^{s_i}_{\text{jointmax}}(G) = \sum_{s_{-i} \in S_{-i}} \frac{\exp\left[\varphi \cdot v^{\text{jointmax}}(s_i, s_{-i})\right]}{\sum_{(k_i, k_{-i}) \in S_i \times S_{-i}} \exp\left[\varphi \cdot v^{\text{jointmax}}(k_i, k_{-i})\right]}.$$

This can be interpreted as applying a softmax to all possible outcomes and taking the probability of each strategy to be the sum of the probabilities in the corresponding row.

Cognitive cost is again proportional to effort, and so

$$c(h_{cell}) = \varphi \cdot C_{cell},$$

where $C_{cell} > 0$ is a free parameter of the cost function.

Simulation Heuristics: Higher-level Reasoning Most previous behavioral models of initial play have a basic structure of belief formation and best response. Such models assume that people first form a belief about which strategy the other player will choose and then select the strategy with the maximal expected value given that belief. In general, effective heuristics do not necessarily have this form; indeed, for many parameter values, the row and cell heuristics described earlier might not be compatible with any beliefs. However, explicitly forming beliefs and calculating the best responses (following a *simulation heuristic*) may be a good decision-making strategy in some situations.

If a row player uses a simulation heuristic, she first considers the game from the column player's perspective, applying some heuristic (a row, cell, or simulation heuristic) that generates a distribution of likely play. She then plays a noisy best response to that distribution.

The cognitive cost of a simulation heuristic is a combination of the cognitive cost of the heuristic for the column player, a constant cost for updating the payoff matrix using that belief (C_{mul}), and a cost that is proportional to the cognitive effort parameter in the last step, as for a row heuristic,

$$c(h_{\text{sim}}) = c(h_{\text{col}}) + C_{\text{mul}} + C_{\text{row}} \cdot \varphi.$$

Notice that once the beliefs have been formed the last cost for taking a decision is based on C_{row} since this process is the same as averaging over the rows as with a row heuristic.

Selection Rule We don't expect a person to use the same heuristic in all games. Instead, they may have a set of heuristics, and choose which one to use in each situation based on an estimate of the candidate heuristics' expected values. We model this selection process as a higher-order selection rule that selects among the first-order heuristics described above. This selection rule allows the decision-maker to select from a few different primitive heuristics, and hence the term "metaheuristic."

Rather than explicitly modeling the process by which players select among the candidate heuristics, for example, by using the approach in Lieder and Griffiths (2015), we use a reduced-form model based on the rational inattention model of Matějka and McKay (2015). We make this simplifying assumption since it allows us to focus on the central parts of our theory. This functional form captures the three key properties a metaheuristic should have: (1) there is a prior weight on each primitive heuristic, (2) a

primitive heuristic will be used more on games in which it is likely to perform well, and (3) this adjustment from the prior based on expected value is incomplete and costly. See Equation 6 in the Appendix for details.

4.2.2 Deep Heuristics

A drawback of using explicitly formulated heuristics, as above, is that the results depend on somewhat arbitrary decisions made by the researchers (in particular, the set of primitive heuristics). To minimize the risk of our conclusions being driven by such decisions, we also consider a nonparametric family of heuristics implemented with neural networks. While not as interpretable as the metaheuristics, this new class includes a much larger set of possible heuristics.

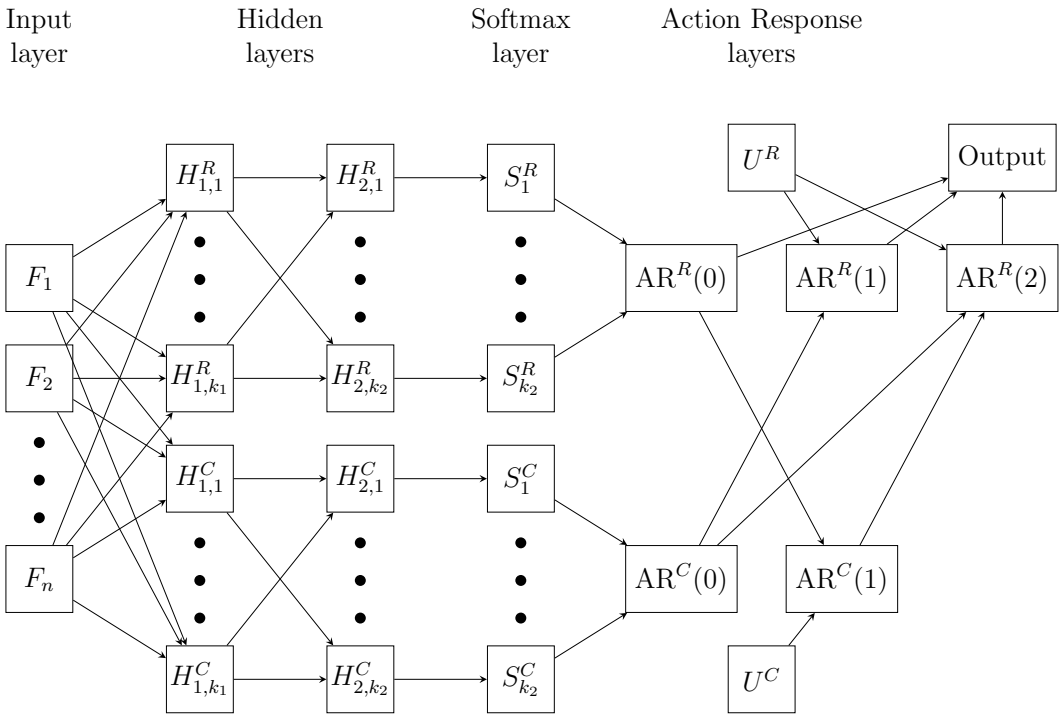


Figure 2: Architecture of the deep heuristic.

We use a neural network architecture similar to the one developed by Hartford et al. (2016), with some adjustments to allow for modeling cognitive costs. The architecture, illustrated in Figure 13, has two key properties specifically adapted to finite normal-form games. First, the connectivity structure ensures that predictions are invariant to relabeling of the strategies, thereby vastly reducing the size of the parameter space (playing a similar role to convolution in computer vision). Second, the architecture explicitly separates recursive reasoning (e.g., level-k) and direct reasoning about the

payoff matrix. This allows us to capture belief formation and best response, as well as simpler heuristics like our row and cell heuristics. Furthermore, we can assign a different cognitive cost to each type of reasoning. A detailed description of the architecture is given in Appendix C.

5 Model-based Analysis

Having formally specified our theoretical framework, we can now take the models to our experimental data. Specifically, we ask whether the behavioral differences found in our experiment are consistent with the rational use of heuristics. To do this, we first compare the predictive accuracy and payoffs achieved by models that are either payoff-optimized or fit directly with the data; in particular, we confirm two preregistered hypotheses generated by the general theory. Next, we compare our model to previously proposed models, and demonstrate that our model yields more accurate predictions.

5.1 Model Estimation

We take an out-of-sample prediction approach to model comparison. Each data set is divided into a training set on which model parameters are estimated and a test set on which predictive performance is evaluated. We used the first 30 treatment games from each population as the training set and the remaining 16 treatment games as the test set. We chose this split so that we could test the predictions on the later games when people would be most likely to be using a consistent decision strategy. We consider each game to consist of two observations: the empirical distribution of play for each player role (row and column). The games are sampled separately for each population but are the same within a population, and we have 10 populations for each treatment. For each treatment, we thus have 600 observations in the training set and 320 observations in the test set. This separation was preregistered and can thus be considered a "true" out-of-sample prediction.

We define separate environments for the two treatments using the actual games and empirical distributions of play in all populations of the corresponding treatment. We thus define the common-interests environment, $\mathcal{E}^+$, by letting $\mathcal{G}^+$ be all the treatment games played in the common-interests treatment, and letting the opponent's behavior, $h^+(G)$, be the actual distribution of play in G. Lastly, P is a uniform distribution over all games in $\mathcal{G}^+$ and always returns h^+ as the heuristic for the opponent. We define the competing-interests environment $\mathcal{E}^-$ correspondingly. Lastly, we divide the games into

the training games, i.e., $\mathcal{G}^+_{\text{train}}$, and test games $\mathcal{G}^+_{\text{test}}$.

The measure of the fit we use is the average negative log-likelihood (or, equivalently, the cross-entropy), where a lower value means a better fit. If p is the observed distribution of play for the row or column player role in some game, and q is the predicted distribution of play from some model, the negative log-likelihood (NLL) is defined as

$$\text{NLL}(q, p) = -\sum_s p_s \cdot \log(q_s).$$

We define the total NLL of a metaheuristic m with cognitive costs C evaluated on the common-interests training set $\mathcal{E}^+_{\text{train}}$ as

$$\text{NLL}(m, \mathcal{E}^+_{\text{train}}, C) = \sum_{G \in \mathcal{G}^+_{\text{train}}} \text{NLL}(m(G, h^+, C), h^+(G)),$$

and analogously for the three remaining training and test sets. We write $m(G, h^+, C)$ since the actual prediction of the metaheuristic m in a given game depends on the performance of the different primitive heuristics, which in turn depend on the opponent's behavior, h^+, and the cognitive costs, C, as given by Equation (6).

The behavior of the metaheuristic model depends on three factors: the consideration set of possible primitive heuristics, the cognitive cost of those heuristics, and the prior distribution for the selection rule. We assume that the consideration set includes one of each type of primitive heuristic: a cell heuristic, a row heuristic, and a simulation heuristic. The model thus has twelve free parameters: six that specify the behavior of the primitive heuristics, four for the cognitive costs, and two for the selection rule's prior.

The cognitive cost parameters are fixed from the decision-maker's perspective, reflecting constraints imposed by the decision-maker's cognitive abilities. We thus fit the cost parameters to data. By contrast, the parameters of the heuristics and the selection rule prior are under the decision-maker's control. We consider two methods for estimating the parameters of the heuristics: fitting them to the data, or optimizing them such that they maximize expected utility. The latter method instantiates our theory that people use heuristics in a resource-rational way. For a given set of cognitive cost parameters $C = (C_{\text{row}}, C_{\text{cell}}, C_{\text{mul}}, \lambda)$, the *fitted* common-interests metaheuristic is given by

$$m_{\text{fit}}(\mathcal{E}^+_{\text{train}}, C) = \operatorname*{argmin}_{m \in \mathcal{M}} \text{NLL}(m, \mathcal{E}^+_{\text{train}}, C),$$

where $\mathcal{M}$ is the space of metaheuristics we restrict our analysis to. The fitted parameters

thus capture the heuristics that empirically best explain human behavior.

The *optimal* common-interests metaheuristic, for cognitive cost C, is instead given by

$$m_{\text{opt}}(\mathcal{E}^+_{\text{train}}, C) = \underset{m \in \mathcal{M}}{\text{argmax}}\, V(m, \mathcal{E}^+_{\text{train}}, C) = \underset{m \in \mathcal{M}}{\text{argmax}} \sum_{G \in \mathcal{G}^+_{\text{train}}} u(m, h^+, G, C)$$

where $u(m, h^+, G, C)$ is the expected utility from employing metaheuristic m against behavior h^+ in game G with cognitive cost parameters C. The optimized parameters thus identify the heuristics that objectively achieve the best cost-benefit tradeoff, given the fitted cost parameters. The fitted and optimal metaheuristics for the competing-interests environment are defined analogously.

Having defined the fitted and optimal heuristics with cognitive costs C, we now turn to the question of how to estimate the cognitive costs. Since the participants are drawn from the same distribution and are randomly assigned to the two treatments, we assume that the cognitive costs are always the same for both treatments.

To estimate the costs, we find the costs that minimize the average NLL of the optimized or fitted heuristics on the training data. Therefore

$$C_{\text{fit}} = \underset{C \in \mathbb{R}^4_+}{\text{argmin}}\ \text{NLL}(m_{\text{fit}}(\mathcal{E}^+_{\text{train}}, C), \mathcal{E}^+_{\text{train}}, C) + \text{NLL}(m_{\text{fit}}(\mathcal{E}^-_{\text{train}}, C), \mathcal{E}^-_{\text{train}}, C),$$

and

$$C_{\text{opt}} = \underset{C \in \mathbb{R}^4_+}{\text{argmin}}\ \text{NLL}(m_{\text{opt}}(\mathcal{E}^+_{\text{train}}, C), \mathcal{E}^+_{\text{train}}, C) + \text{NLL}(m_{\text{opt}}(\mathcal{E}^-_{\text{train}}, C), \mathcal{E}^-_{\text{train}}, C).$$

Notice the crucial difference between the fitted and optimized metaheuristics. For the fitted metaheuristics, we fit both the cognitive cost parameters and the heuristic parameters to match actual behavior in the two training sets. For the optimized metaheuristics, we fit only the cognitive cost parameters; the heuristic parameters are set to maximize the payoff minus the cognitive cost. As a result, any difference between the optimal common-interests metaheuristic and the optimal competing-interests metaheuristic is entirely driven by differences in performance between the different heuristics in the two environments.

5.2 Results for Metaheuristics

Next, we consider our two model-based hypotheses regarding the metaheuristic model's ability to capture the difference in the participants' behavior between the two treatments.

Hypothesis 3. *Participants' behavior will differ between the two treatments in a way that the model can capture. Specifically, their behavior in the common-interests test games should be better predicted by the common-interests metaheuristic than by the competing-interests metaheuristics. Conversely, their behavior in the competing-interests test games should be better predicted by the competing-interests metaheuristics than by the common-interests metaheuristic. This should hold both for the fitted metaheuristics and for the optimized metaheuristics.*

Concretely, this hypothesis states that the following four inequalities should hold:

$$\begin{aligned}
\mathrm{NLL}(m_{\mathrm{fit}}(\mathcal{E}^-_{\mathrm{train}}), \mathcal{E}^-_{\mathrm{test}}) &< \mathrm{NLL}(m_{\mathrm{fit}}(\mathcal{E}^+_{\mathrm{train}}), \mathcal{E}^-_{\mathrm{test}}) \\
\mathrm{NLL}(m_{\mathrm{opt}}(\mathcal{E}^-_{\mathrm{train}}), \mathcal{E}^-_{\mathrm{test}}) &< \mathrm{NLL}(m_{\mathrm{opt}}(\mathcal{E}^+_{\mathrm{train}}), \mathcal{E}^-_{\mathrm{test}}) \\
\mathrm{NLL}(m_{\mathrm{fit}}(\mathcal{E}^-_{\mathrm{train}}), \mathcal{E}^+_{\mathrm{test}}) &> \mathrm{NLL}(m_{\mathrm{fit}}(\mathcal{E}^+_{\mathrm{train}}), \mathcal{E}^+_{\mathrm{test}}) \\
\mathrm{NLL}(m_{\mathrm{opt}}(\mathcal{E}^-_{\mathrm{train}}), \mathcal{E}^+_{\mathrm{test}}) &> \mathrm{NLL}(m_{\mathrm{opt}}(\mathcal{E}^+_{\mathrm{train}}), \mathcal{E}^+_{\mathrm{test}}),
\end{aligned}$$

where the notation for C_{fit} and C_{opt} is omitted for brevity.

In order to facilitate comparisons between treatments and between games, we use "relative prediction loss", that is, the difference in NLL between the model's predictions and the theoretical minimum NLL. Let y be the observed empirical distribution of play in some game G. Then the lowest possible NLL in that game is $NLL(y, y)$. The relative prediction loss for model m in game G is thus given by[5]

$$NLL(m, G, C) - NLL(y, y).$$

We compute confidence intervals of the relative prediction loss over all the games in the test set. Since we consider each game separately for the two different player roles, there are 320 observations per test set.

Figure 3 shows the relative prediction loss on the test data in each treatment achieved by each possible method of fitting the model. We clearly see that the models that were trained on data from the same treatment as the test set outperform models trained on the other treatment. This confirms Hypothesis 3.

An even more striking result is that the optimized metaheuristics achieve nearly the same predictive performance as the fitted metaheuristics. That is, a model that uses the

[5]The resulting measure of performance is related to the completeness measure of Fudenberg et al. (2022). However, since we have only fifteen participants per game and role, and there is randomness in behavior, even the perfect model would not be able to get the exact distribution of play right. Therefore, the theoretical minimum is truly theoretical.

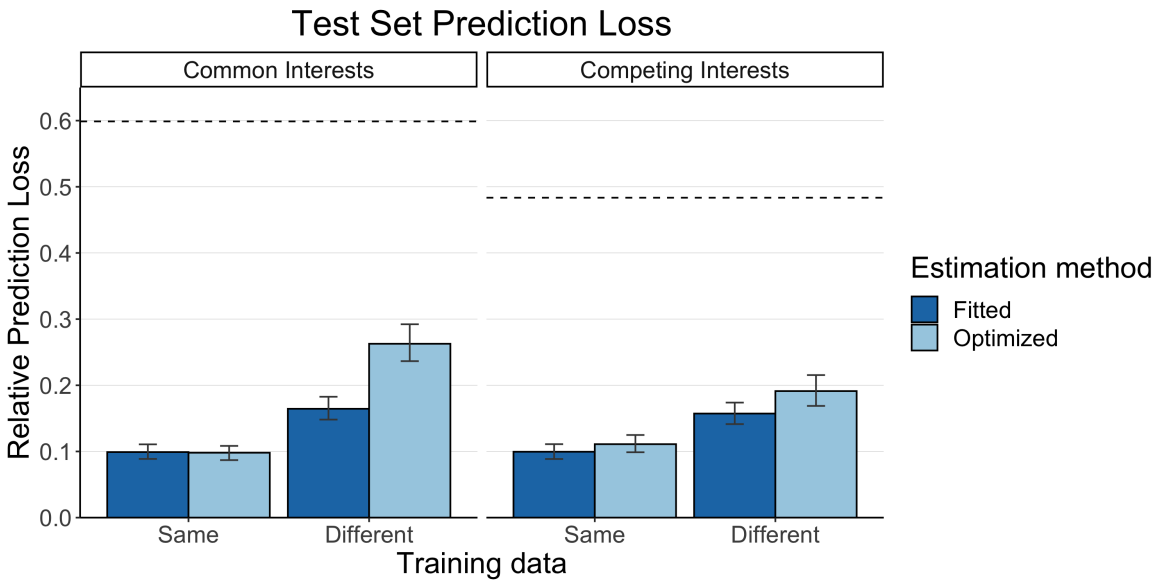


Figure 3: Predictive performance of the metaheuristics. Each panel shows the relative prediction loss (average negative log-likelihood minus lowest possible value) of the test data for one treatment (competing interests or common interests). Models are fitted or optimized to either the competing-interests training games or the common-interests training games. The error bars show 95% confidence intervals. The dashed line corresponds to uniform random play, which assigns the same probability to each action in each game.

same set of cognitive cost parameters in both treatments (with the heuristic parameters set to optimize the resultant expected payoff-cognitive cost tradeoff) explains participant data almost as well as the fully parameterized model, in which the heuristic parameters are separately fitted to each treatment.

Not only do we confirm our hypothesis and show that the rational heuristic is a strong predictor, but we also see that we capture most of the distance between the uniform random play and the theoretical minimum NLL. Table 6 in the Appendix shows the accuracy and average NLL for all models we consider in the paper. There, we see that the average accuracy of the optimal metaheuristic is 88%, meaning that in 88% of the games, the modal action is assigned the highest probability. It should also be noted that in the games where the optimal metaheuristic makes an incorrect prediction, the modal action is on average only played by 54% of the participants, while the modal action was played by 75% of the participants in all of the test games. Therefore, in the games where the proposed model fails to assign the highest probability to the modal action, play is quite even and hence difficult to predict.

Our final model-based hypothesis provides an additional test that the metaheuristics

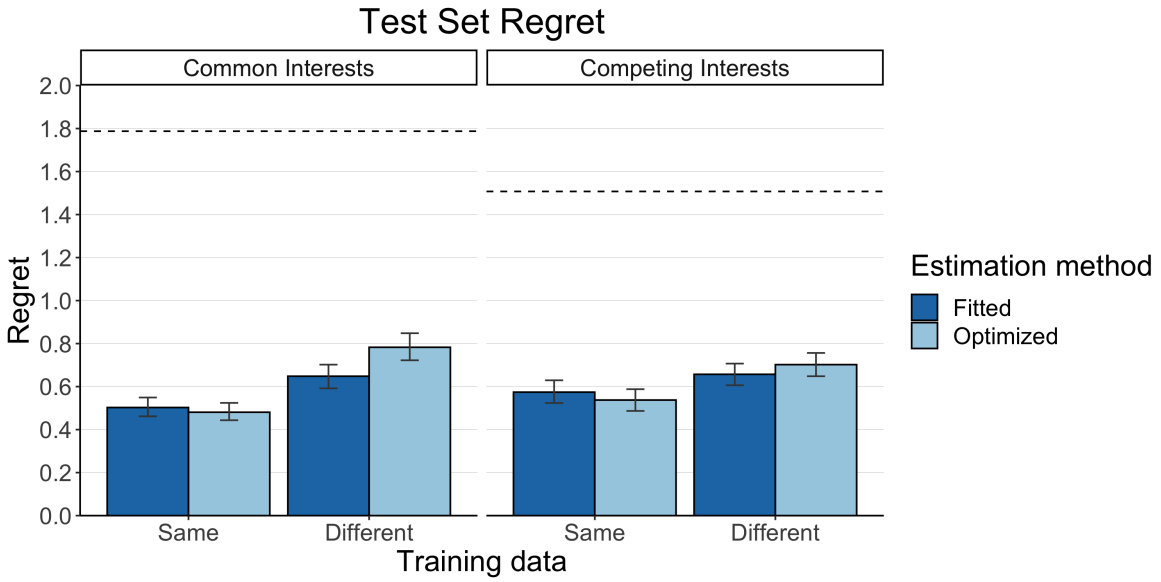


Figure 4: Payoff performance of the metaheuristics. Each panel shows the regret (best possible expected payoff minus true expected payoff) attained by models that are trained and tested in different combinations of common-interests and competing-interests environments. The dashed line shows the performance of uniform random play.

that participants use are adapted to their treatment environment:

Hypothesis 4. *The fitted heuristics estimated for a given treatment should achieve higher expected payoffs on the test games for that treatment, as compared to the heuristics estimated for the other treatment.*

The logic of this hypothesis is that even if we do not assume that participants use optimal heuristics, we should still see that the heuristics that best describe participants' behavior in each treatment achieve higher payoffs in that treatment. As with prediction loss, we use a relative performance measure that accounts for differences in maximal payoff in the two treatments. Specifically, we quantify performance in terms of regret, the difference between the expected payoff given the predicted behavior and the maximum expected payoff in each game.

As illustrated in Figure 4, the results confirm our hypothesis. When testing on games from either treatment, the models fitted to human behavior in the same treatment achieved lower regret than those fitted to the other treatment, although the difference is larger for the common-interests games.

In Appendix D.2 we present results from pairwise tests of both Hypotheses 3 and 4. We see there that all the differences in both relative prediction loss and regret are

significant at the 0.01 level.[6]

5.3 Results for Deep Heuristics

By applying the same estimation method to the deep heuristics as we did to the metaheuristics, we can test whether Hypotheses 3 and 4 also hold for a completely different specification of the space of heuristics and cognitive costs. In Figure 5, we see that Hypothesis 3 holds for this specification as well: the models make more accurate predictions for the treatment on which they were trained or optimized. We also see that the predictive performance of the optimal heuristic is close to the fitted heuristic, given optimized cognitive costs.

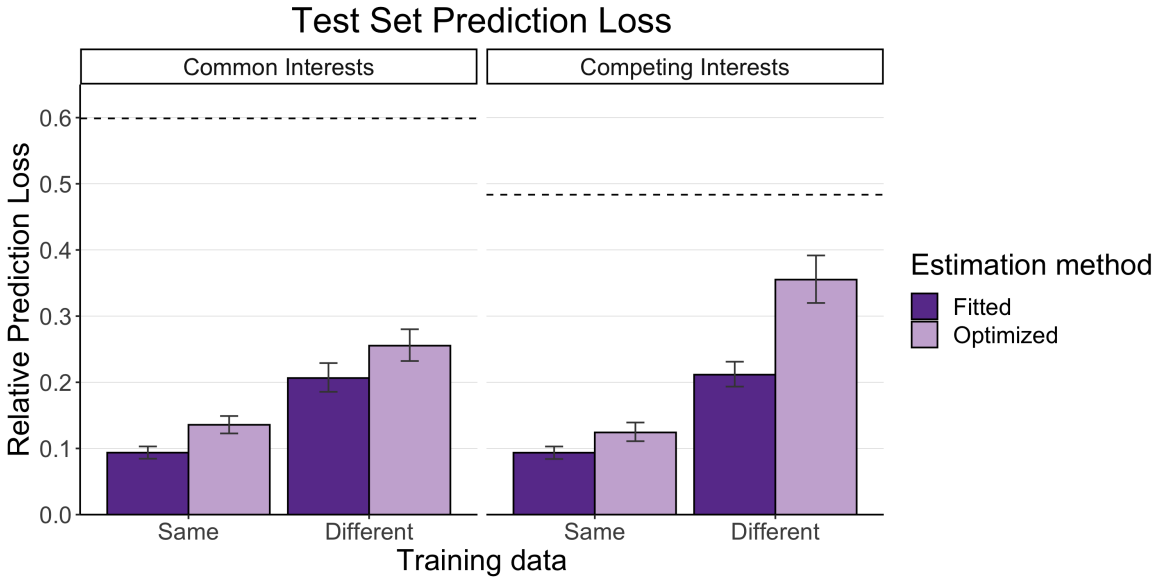


Figure 5: Predictive performance of the deep heuristics.

We can also test Hypothesis 4 in the same way by looking at the expected payoff from the two different deep heuristics fitted to the behavior of the participants in the two different treatments. As before, we see that the fitted models achieved lower regret in the treatment on which they were trained, again suggesting that the heuristics people use are well adapted to their environment.

[6]In the preregistration, we did not specify a formal testing procedure for these differences and did not originally include such a test in the paper. However, after discussions and presentations, it became clear that such tests are sought after by readers, and we have therefore added them.

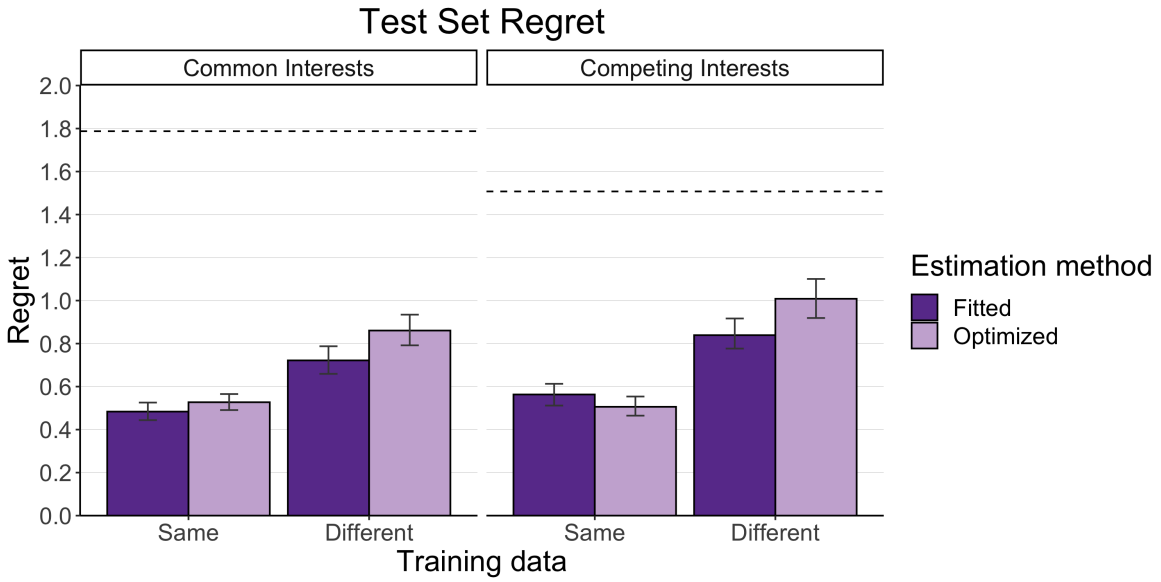


Figure 6: Payoff performance of the deep heuristics.

5.4 Model Comparison

In the previous sections, we have shown that the rational use of heuristics can explain and predict people's behavior in one-shot games, in particular how their behavior depends on the previous games they have played. Next, we ask whether our proposed theory provides a more accurate account of that behavior when compared to previously proposed models. Specifically, we consider three alternative models: quantal cognitive hierarchy (QCH), QCH with prosocial preferences, and noisy best-response to the true distribution of play with prosocial preferences.

Quantal Cognitive Hierarchy. In previous comparisons between behavioral models of one-shot games, variations of cognitive hierarchy models are usually the best performing (Camerer et al., 2004; Wright and Leyton-Brown, 2017). In such a model, we consider agents of different cognitive levels. In the quantal cognitive hierarchy (QCH) model we consider here, a level-0 agent plays the uniformly random strategy, playing each action with an equal probability. A level-1 player (logit) best responds to a level-0. Finally, a level-2 player best responds to a combination of level-0 and level-1 players.[7] The model has 4 parameters: the share of level-0 and level-1 players (which together determine the share of level-2 players), the sensitivity λ_1 of level-1 players, and the sensitivity λ_2 of level-2 players.

Prosocial Preferences

We have attributed the difference in the participants' behavior between the two

[7]We found that adding higher levels of play did not improve predictive performance.

treatments to their learning different heuristics. However, this pattern of behavior could be explained by a change not in their decision-making strategy but in their underlying preferences. In particular, participants in the common-interests environment may develop a sense of camaraderie that makes them care about the other players' payoffs, while participants in the competing-interests environment may become jaded or even spiteful, leading them to disregard the others' payoffs.

To test this alternative explanation, we augmented the QCH model with a prosocial utility function (Fehr and Schmidt, 1999; Bruhin et al., 2019), i.e.,

$$u_i(s_i, s_{-i}) = (1 - \alpha s - \beta r) \times \pi_i(s_i, s_{-i}) + (\alpha s + \beta r) \times \pi_{-i}(s_i, s_{-i}), \tag{5}$$

where s indicates whether $\pi_i(s_i, s_{-i}) < \pi_{-i}(s_i, s_{-i})$ and r indicates whether $\pi_i(s_i, s_{-i}) > \pi_{-i}(s_i, s_{-i})$. In other words, α determines how much player i values the payoff of player $-i$ when i gains less than $-i$, and β how much player i values the payoff of player $-i$ when i gains more than $-i$. This augmentation thus adds two parameters to the QCH model, α and β. In this model, beliefs are formed using a standard QCH model, but the payoffs are changed according to the prosocial preferences model (Equation 5) before the last quantal best-response step.[8] This model can account for differences in behavior between the two treatments both by assuming different levels of prosociality and by assuming different levels of reasoning or sensitivity in the QCH step.

Differing Beliefs

A second possible source of differing behavior across treatments is differing beliefs. Since people behave differently in the two treatments, participants may form different beliefs about what they expect the other player to do. In particular, participants in the common-interests treatment may expect the other player to cooperate by selecting an action with a jointly beneficial outcome, while participants in the competing-interests treatment may expect the other player to select the safest action for themselves.

To test this account, we replace the recursively formed beliefs of QCH with the correct (empirical) belief. This model thus plays a noisy best response to the actual distribution of participants' play. In this model, we additionally allow for prosocial preferences, resulting in a three-parameter model.

Results. In Figure 7, we compare the out-of-sample predictive performance of these two alternative models and our two suggested specifications for the space of heuristics. While the alternative models are estimated by fitting the parameters to match the

[8] We also considered another model combining QCH and prosocial preferences, in which the player also has some beliefs about the other player's prosociality that informs the beliefs formed during the QCH steps. This didn't make a meaningful difference in fit.

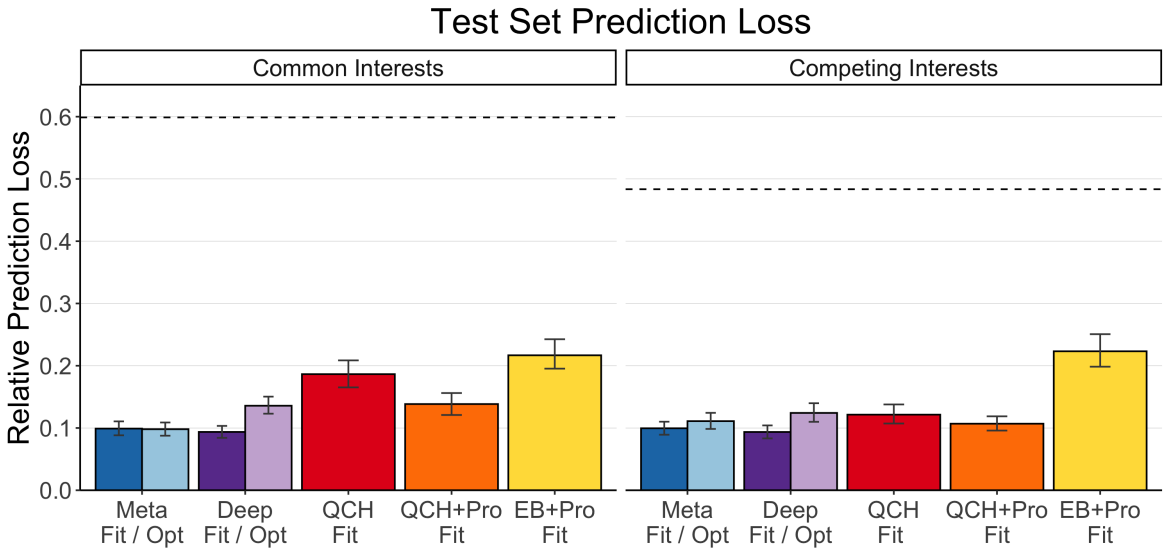


Figure 7: Out-of-sample relative prediction loss for alternative models of behavior. All the models are estimated on the training games of the same environment as the test games. The error bars show a 95% confidence interval. Legend: QCH = quantal cognitive hierarchy, Pro = prosocial preferences, EB = empirical beliefs.

participants' behavior, we also include the optimized versions of our two specifications.

For the common-interests games, it is clear that both the fitted and optimized versions of our models outperform both the quantal cognitive hierarchy model and the noisy best response with prosocial preferences (prosociality) model. The model with both prosocial preferences and recursive reasoning (Pro+QCH) performs better, but is still outperformed by three of our models (excluding the optimized deep heuristics). For the competing-interests games, the prosociality model is still clearly performing worst, but the other models all perform similarly. This suggests that the QCH model predicts participants' behavior better in the competing-interests environment than in the common-interests environment. Taken together, our proposed models are better at predicting behavior than alternative models, including the current best-performing model in the literature (QCH).

We also see clearly in Figure 7 that the predictive performance of the metaheuristics and fitted deep heuristics are very close, even though the deep heuristics encompasses a much larger space of heuristics. This suggests that we have managed to capture the relevant space of heuristic strategies with our parameterization of the metaheuristics. That is, the metaheuristic model is nearly "complete" in the sense of Fudenberg et al. (2022).

6 Discussion

In the theory presented, we combine two perspectives. On the one hand we assume that people use simple cognitive strategies to choose actions that are often inconsistent with rational behavior in any given game. On the other hand, we don't assume that the specific heuristics used are predetermined or insensitive to incentives. On the contrary, we assume that the heuristics people use are chosen resource-rationally, such that they strike an optimal balance between expected payoffs and cognitive costs. We have seen that by combining these two perspectives, we can predict behavior more accurately and better understand the influence of the larger environment on behavior in a given game.

In particular, the proposed approach can help us predict when we should expect behavior to coincide with rational behavior and when we might see systematic deviations from a rational benchmark. Behavior will coincide with rational behavior if two conditions are satisfied. Firstly, there has to exist a simple heuristic that leads to the optimal action. Secondly, that heuristic has to perform well in the larger environment so that the decision-maker can learn to use it. When there doesn't exist a simple and high-performing heuristic, or when the heuristic that normally works well leads to the wrong decision, we will observe consistent deviations. This latter case is nicely illustrated in our comparison games.

The optimal heuristic will focus on the features of the games that are often of importance, but miss opportunities that are rare. Specifically, a person used to common-interests games might miss an opportunity for personal gain at the other player's expense while a person used to competing-interests games might fail to notice an outcome that is actually best for everyone.

Our findings relate to those of Peysakhovich and Rand (2016), who showed that varying the sustainability of cooperation in an initial session of the repeated prisoner's dilemma affects how much prosocial behavior and trust is shown in later games, including the one-shot prisoner's dilemma. Our results provide a qualitative replication of this idea. In particular, we found that putting people in an environment in which prosocial heuristics (such as jointmax) perform well leads them to choose prosocial actions in the comparison games and in some cases, even to select dominated options. By contrast, putting people in an environment where prosocial actions often result in low payoffs prevents people from achieving efficient outcomes, even when they are Nash equilibria. Consistent with our theory, Peysakhovich and Rand interpreted their findings as the result of heuristic decision-making. We build on this intuitively appealing notion by specifying formal models of heuristics in one-shot games that make quantitative

predictions. We also emphasize the influence of cognitive costs (in addition to payoffs) on the heuristics people use.

Finally, we would like to emphasize an important difference between our theory and previously proposed models of learning in games. Previous learning models have been posed at the level of *action*; people learn which action to take in a specific (repeatedly played) game (e.g. Jehiel, 2005; Grimm and Mengel, 2012). In contrast, in our theory, learning happens at the level of *reasoning*; people learn how to decide what to do in a new game. We believe that this more abstract form of learning is more broadly applicable in the real world, as it is rare that we ever encounter the exact same situation twice (a feature that is captured by the randomly generated games in our experiment).

7 Conclusion

We have proposed a theory of human behavior in one-shot normal-form games based on the resource-rational use of heuristics. According to this theory, people select their actions using simple cognitive heuristics that flexibly and selectively process payoff information; the heuristics people choose to use are ones that strike a good tradeoff between the expected payoffs and the cognitive cost.

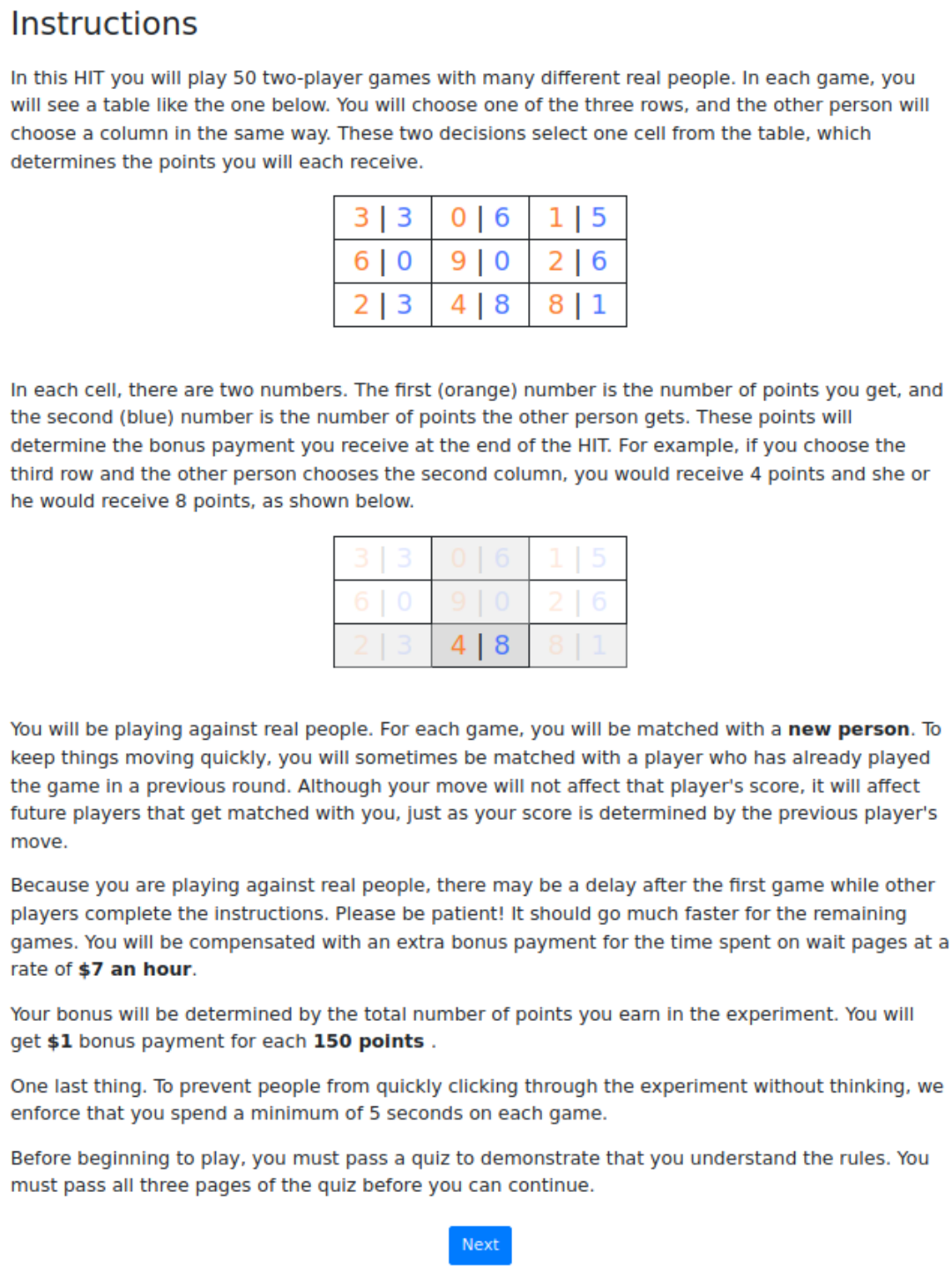
In a large preregistered experiment, we confirmed one of the primary qualitative predictions of the theory: people learn which heuristics are resource-rational in a given environment, and thus their recent experience affects the choices they make. In particular, we found that placing participants in environments with common (vs. competing) interests leads them to select the most efficient (or least efficient) equilibrium in a weak-link game and to cooperate (or defect) in a prisoner's dilemma.

Furthermore, we found that our theory provides a strong quantitative account of our participants' behavior, making more accurate out-of-sample predictions than both the quantal cognitive hierarchy model and a model with prosocial preferences and a noisy best response. Strikingly, we found that a resource-rational model, in which behavior in both common-interests and competing-interests treatments is predicted using a single set of fitted cost parameters (with the heuristic parameters set to optimize the resultant expected payoff-cognitive cost tradeoff), achieved nearly the same accuracy as the fully parameterized model in which the heuristic parameters are estimated separately to match the behavior in each treatment. Coupled with the overall high predictive accuracy of the model, this provides strong evidence in support of the theory that people use heuristics that optimally trade off between the expected payoff and the cognitive cost. We also found similar results using an entirely different neural network-based family

of heuristics, indicating that these findings are robust to the parameterization of the heuristics.

From a broader perspective, our theory speaks to a decades-long debate on the rationality of human decision-making. In contrast to classical models based on optimization and utility maximization, which fail to capture systematic patterns in human choice behavior, recent models instead emphasize our systematic biases, suggesting that we rely on simple and error-prone heuristics to make decisions. In this paper, we hope to have offered a synthesis of these two perspectives, by treating heuristics as things that can themselves be optimized in a utility-maximization framework. We hope this approach will prove to be a valuable step forward toward a more unified understanding of economic decision-making.

A Instructions for the experiment



Instructions

In this HIT you will play 50 two-player games with many different real people. In each game, you will see a table like the one below. You will choose one of the three rows, and the other person will choose a column in the same way. These two decisions select one cell from the table, which determines the points you will each receive.

3 | 3	0 | 6	1 | 5
6 | 0	9 | 0	2 | 6
2 | 3	4 | 8	8 | 1

In each cell, there are two numbers. The first (orange) number is the number of points you get, and the second (blue) number is the number of points the other person gets. These points will determine the bonus payment you receive at the end of the HIT. For example, if you choose the third row and the other person chooses the second column, you would receive 4 points and she or he would receive 8 points, as shown below.

3 | 3	0 | 6	1 | 5
6 | 0	9 | 0	2 | 6
2 | 3	4 | 8	8 | 1

You will be playing against real people. For each game, you will be matched with a **new person**. To keep things moving quickly, you will sometimes be matched with a player who has already played the game in a previous round. Although your move will not affect that player's score, it will affect future players that get matched with you, just as your score is determined by the previous player's move.

Because you are playing against real people, there may be a delay after the first game while other players complete the instructions. Please be patient! It should go much faster for the remaining games. You will be compensated with an extra bonus payment for the time spent on wait pages at a rate of **$7 an hour**.

Your bonus will be determined by the total number of points you earn in the experiment. You will get **$1** bonus payment for each **150 points** .

One last thing. To prevent people from quickly clicking through the experiment without thinking, we enforce that you spend a minimum of 5 seconds on each game.

Before beginning to play, you must pass a quiz to demonstrate that you understand the rules. You must pass all three pages of the quiz before you can continue.

Next

Figure 8: The instructions one the first page when a participant joins the experiment.

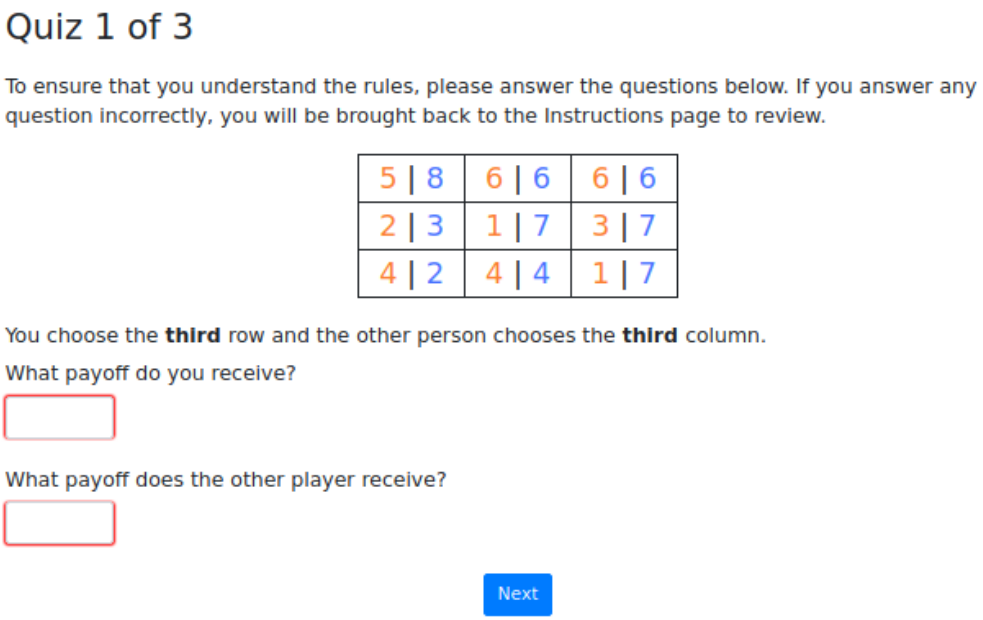


Figure 9: The participants have to complete three questions like this in a row in order to be allowed to participate in the experiment.

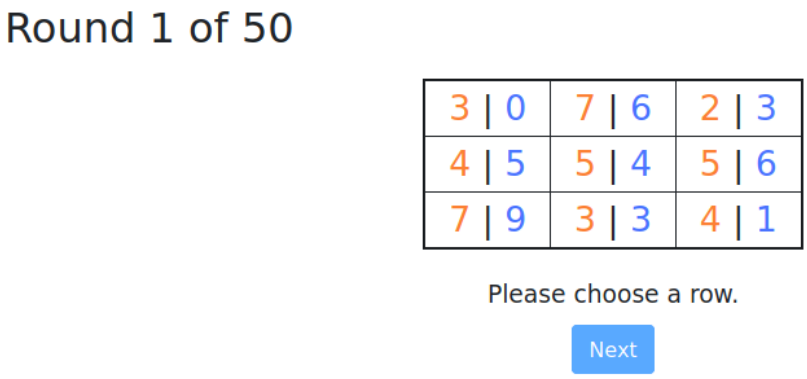


Figure 10: In each round, the participant chose a row by clicking on it. Once it is clicked it is highlighted and they have to click the next button to proceed.

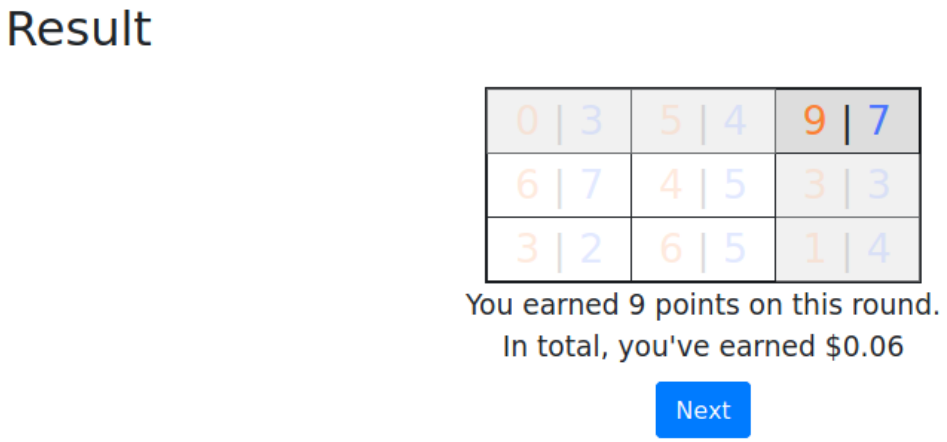


Figure 11: Once the matched participant chooses a column, either by making a decision or by sampling from previous decisions in the game from the same population, the result is shown.

B Complete Description of Metaheuristics

To exemplify the different heuristics, we will apply them to the following example game.

	1	**2**	**3**
1	$0, 1$	$0, 2$	$8, 8$
2	$5, 6$	$5, 5$	$2, 2$
3	$6, 5$	$6, 6$	$1, 1$

Figure 12: Example normal-form game represented as a bi-matrix. The row player chooses a row and the column player chooses a column. The first number in each cell is the payoff of the row player and the second number is the payoff of the column player.

B.1 Row Heuristics

A *row heuristic* calculates a value, $v(s_i)$, for each pure strategy, $s_i \in S_i$, based only on the player's own payoffs associated with s_i. That is, it evaluates a strategy based only on the first entry in each cell of the corresponding row of the payoff matrix (see Figure 12). Formally, a row heuristic is defined by the row-value function v such that

$$v(s_i) = f(\pi_i(s_i, \mathbf{1}), \ldots, \pi_i(s_i, m_i)))$$

for some function $f : \mathbb{R}^{m_{-i}} \to \mathbb{R}$. For example, if f is the mean function, then we have

$$v^{\text{mean}}(s_i) = \frac{1}{m_{-i}} \sum_{s_{-i} \in S_{-i}} \pi_i(s_i, s_{-i}),$$

which evaluates each strategy by the average payoff in the corresponding row of the payoff matrix. Deterministically selecting $\arg\max_{s_i} v^{\text{mean}}(s_i)$ gives exactly the behavior of a level-1 player in the classical level-k model.

If, instead, we let f be min, we recover the *maximin* heuristic, which calculates the minimum value associated with each strategy and tries to choose the row with the highest minimum value,

$$v^{\min}(s_i) = \min_{s_{-i} \in S_{-i}} \pi_i(s_i, s_{-i}),$$

and, similarly, if we let f be max, we recover the *maximax* heuristic,

$$v^{\max}(s_i) = \max_{s_{-i} \in S_{-i}} \pi_i(s_i, s_{-i}).$$

While one can imagine a very large space of possible functions f, we consider a one-dimensional family that interpolates smoothly between min and max. We construct such a family with the following expression:

$$v^{\gamma}(s_i) = \sum_{s_{-i} \in S_{-i}} \pi_i(s_i, s_{-i}) \cdot \frac{\exp\left[\gamma \cdot \pi_i(s_i, s_{-i})\right]}{\sum_{s \in S_{-i}} \exp\left[\gamma \cdot \pi_i(s_i, s)\right]},$$

which approaches $v^{\min}(s_i)$ as $\gamma \to -\infty$, $v^{\max}(s_i)$ as $\gamma \to \infty$, and $v^{\text{mean}}(s_i)$ when $\gamma = 0$. Intuitively, we can understand this expression as computing an expectation of the payoff for s_i under different degrees of optimism about the other player's choice of s_{-i}. In the above example game (Figure 12), the heuristic will assign the highest value to **1** (the top row) when γ is large and positive, to **2** when γ is large and negative, and to **3** when $\gamma = 0$. Notice that if $\gamma \neq 0$, the values associated with the different strategies do not necessarily correspond to a consistent belief about the other player's action. For example, if γ is positive, the highest payoff in each row will be overweighted, but this might correspond to a different column in each row; in the example game (Figure 12), column 3 is overweighted when evaluating row 1 but downweighted when evaluating rows 2 and 3. Although this internally inconsistent weighting may appear irrational, it provides an extra degree of freedom that can increase the expected payoff in a given environment without additional cognitive effort.

We assume that the computation of v is subject to noise, but that this noise can

be reduced through cognitive effort, which we operationalize as a single scalar φ. In particular, following Stahl and Wilson (1994), we assume that the noise is Gumbel-distributed and thus recover a multinomial logit model with the probability that player i plays strategy s_i being

$$h^{s_i}_{\text{row}}(G) = \frac{\exp\left[\varphi \cdot v(s_i)\right]}{\sum_{k \in S_i} \exp\left[\varphi \cdot v(k)\right]}.$$

We assume that the cost is proportional to the effort, i.e.,

$$c(h_{\text{row}}) = \varphi \cdot C_{\text{row}},$$

where $C_{\text{row}} > 0$ is a free parameter of the cost function.

B.2 Cell Heuristics

We refer to heuristics that reason directly about outcomes as *cell heuristics*. Based on preliminary analyses, we identified one specific form of cell heuristic that participants appear to use frequently: The *jointmax* heuristic, which identifies the outcome that is most desirable for both players. Formally, the joint desirability of a cell is given by

$$v^{\text{jointmax}}(s_i, s_{-i}) = \min\left\{\pi_i(s_i, s_{-i}), \pi_{-i}(s_i, s_{-i})\right\}$$

and the probability of playing a given strategy with cognitive effort φ is given by

$$h^{s_i}_{\text{jointmax}}(G) = \sum_{s_{-i} \in S_{-i}} \frac{\exp\left[\varphi \cdot v^{\text{jointmax}}(s_i, s_{-i})\right]}{\sum_{(k_i, k_{-i}) \in S_i \times S_{-i}} \exp\left[\varphi \cdot v^{\text{jointmax}}(k_i, k_{-i})\right]}.$$

This can be interpreted as applying a softmax to all possible outcomes and taking the probability of each strategy to be the sum of the probabilities in the corresponding row. In the example game (Figure 12), the jointmax heuristic would assign the highest probability to row **1** because the cell $(\mathbf{1}, \mathbf{3})$ with payoffs $(8, 8)$ has the highest minimum payoff.

The cognitive cost is again proportional to effort, and so

$$c(h_{cell}) = \varphi \cdot C_{cell},$$

where $C_{cell} > 0$ is a free parameter of the cost function.

B.3 Simulation Heuristics: Higher-Level Reasoning

If a row player uses a simulation heuristic, she first considers the game from the column player's perspective, applying some heuristic that generates a distribution of likely play. She then plays a noisy best response to that distribution. Let G^T denote the transposed game, i.e., the game from the column player's perspective. Let h_{col} be the heuristic the row player uses to estimate the column player's behavior; then, $h_{\text{sim}}(G)$ is given by

$$h_{\text{row}}^{s_i} = \frac{\exp\left[\varphi \cdot \left(\sum_{s_{-i} \in S_{-i}} \pi_i(s_i, s_{-i}) \cdot h_{\text{col}}^{s_{-i}}(G^T)\right)\right]}{\sum_{s_i \in S_i} \exp\left[\varphi \cdot \left(\sum_{s_{-i} \in S_{-i}} \pi_i(s_i, s_{-i}) \cdot h_{\text{col}}^{s_{-i}}(G^T)\right)\right]},$$

where φ is the cognitive effort parameter. A simulation heuristic is thus defined by a combination of a heuristic and an effort parameter $(h_{\text{col}}, \varphi)$.

The cognitive cost for a simulation heuristic is calculated by first calculating the cognitive cost associated with the heuristic used for the column player's behavior, then a constant cost for updating the payoff matrix using that belief (C_{mul}), and one additional cost that is proportional to the cognitive effort parameter in the last step, as if it was a row heuristic,

$$c(h_{\text{sim}}) = c(h_{\text{col}}) + C_{\text{mul}} + C_{\text{row}} \cdot \varphi.$$

Notice that once the beliefs have been formed and the beliefs have been incorporated, the last cost for taking a decision is based on C_{row} since this process is the same as averaging over the rows as for a row-heuristic.

B.4 Selection Rule

We model the selection of primitive heuristics using the rational inattention model of Matějka and McKay (2015). While we don't think about the underlying selection process as inherently one of rational inattention, the rational inattention model captures the key properties we expect from the selection rule: (1) there is a prior weight on each heuristic, (2) a heuristic will be used more on games in which it is likely to perform well, and (3) the adjustment from the prior based on expected value is incomplete and costly.

Assume that an individual is choosing between n heuristics $H = \{h^1, h^2, \ldots, h^N\}$.

Then the probability of using heuristic h^n when playing game G is given by

$$\begin{aligned}\mathbb{P}\left[\{\text{use } h^n \text{ in } G\}\right] &= \frac{\exp\left[(a_n + V(h^n, \mathcal{E}, G))/\lambda\right]}{\sum_{j=1}^{N} \exp\left[(a_j + V(h^j, \mathcal{E}, G))/\lambda\right]} \\ &= \frac{p_n \exp\left[V(h^n, \mathcal{E}, G)/\lambda\right]}{\sum_{j=1}^{N} p_j \exp\left[V(h^j, \mathcal{E}, G)/\lambda\right]} \end{aligned} \tag{6}$$

where λ_i is an adjustment cost parameter and a_n are weights that give the prior probability of using the different heuristics, $p_n = \frac{\exp(a_n/\lambda_i)}{\sum_{j=1}^{N} \exp(a_j/\lambda_i)}$.

A metaheuristic is defined by a tuple $m = \langle H, P \rangle$ where $H_i = \{h^1, h^2, \ldots, h^N\}$ is a finite consideration set of heuristics, and $P = \{p^1, p^2, \ldots, p^N\}$ a prior over those heuristics. We can express the performance of a metaheuristic in an environment $\mathcal{E}$, analogously to (2) for heuristics, as

$$V^{meta}(m, \mathcal{E}) = \sum_{G \in \mathcal{G}} \sum_{h \in H} V(h^n, \mathcal{E}, G) \cdot \frac{p_n \exp\left[V(h^n, \mathcal{E}, G)/\lambda\right]}{\sum_{j=1}^{N} p_j \exp\left[(V(h^j, \mathcal{E}, G))/\lambda\right]} \cdot P(G). \tag{7}$$

The optimization problem faced by the individual, subject to the adjustment cost λ, is then to maximize (7), i.e., to choose the optimal consideration set and corresponding priors,

$$m^* = \underset{H \in \mathcal{P}_{fin}(\mathcal{H})}{\operatorname{argmax}} \underset{P \in \Delta(H)}{\operatorname{argmax}} V^{meta}\left(\langle H, P \rangle, \mathcal{E}\right),$$

where $\mathcal{P}_{fin}(\mathcal{H})$ is the set of all finite subsets of all possible heuristics. In practice, this is not a solvable problem when the consideration set of possible heuristics, $\mathcal{H}$, is large. Therefore, we will assume a small set of heuristics and jointly find optimal parameters of those heuristics and priors P.

C Deep Heuristics

Our neural network architecture is based on that developed by Hartford et al. (2016). The idea is to let every element of the input and hidden layers be a matrix of the same size as the game, instead of a single value as is typical. Each cell in those matrices is then treated in the same way. This ensures that the deep heuristic is invariant to relabeling of strategies, as should be expected from any decision rule for normal-form games.

Higher-level reasoning is incorporated by first having two separated neural networks, representing a "level-0" heuristic for the row player and the column player separately, and then possibly taking into account the thus formed beliefs about the column player's

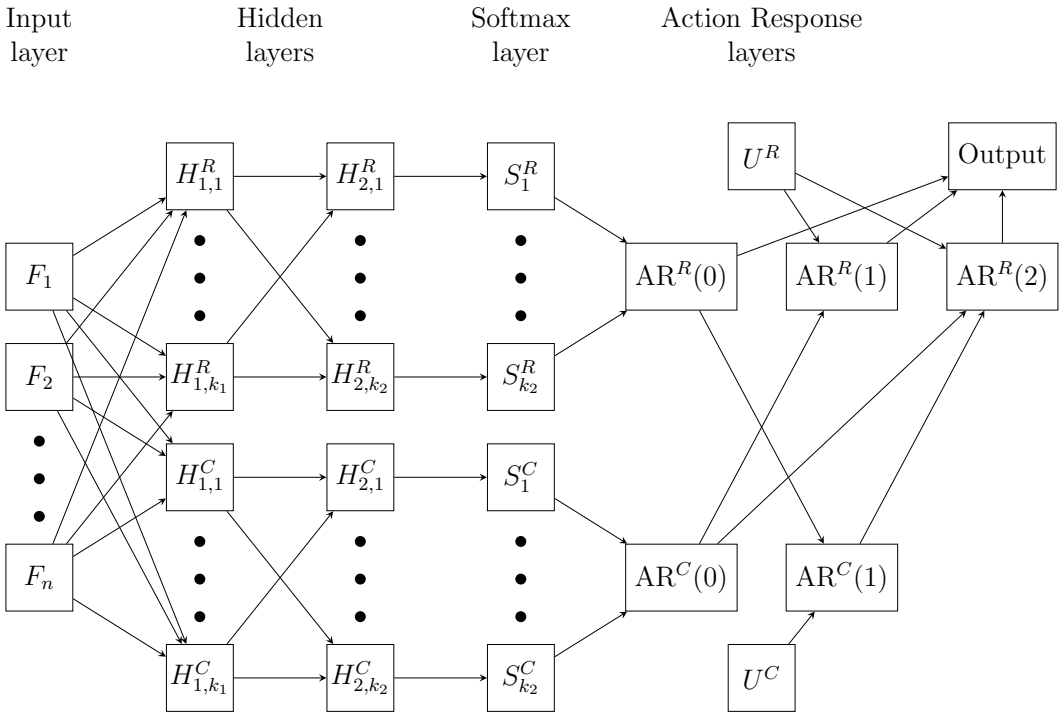


Figure 13: Architecture of the deep heuristic.

behavior in separate "action response" layers. The different action response layers are then combined into a response distribution. A heuristic that did not explicitly form beliefs about the other player's behavior would let $AR^R(0)$ be the output, a person who applies a heuristic to estimate the opponent's behavior and then best responds to it would only use $AR^R(1)$, etc. The neural network architecture is illustrated in Figure 13.

C.1 Feature Layers

The hidden layers are updated according to

$$H^R_{l,k} = \phi_l \left(\sum_j w^R_{l,k,j} H^R_{l-1,j} + b^R_{l,k} \right) \quad H^R_{l,k} \in \mathbb{R}^{m_R \times m_C}$$

and similarly for H^C. For the first hidden layer $H^R_{0,i} = H^C_{0,i} = F_i$, and so the two disjoint parts have the same feature matrices, but different weights.

The feature matrices consist of matrices where each cell contains information associated with the row or column of one payoff matrix. The payoff matrices for the row and column players are denoted by U^R and U^C, respectively. More specifically, we

$$\begin{pmatrix} \min_{R,C}\{U^R_{1,1}, U^C_{1,1}\} & \min_{R,C}\{U^R_{1,2}, U^C_{1,2}\} & \min_{R,C}\{U^R_{1,3}, U^C_{1,3}\} \\ \min_{R,C}\{U^R_{2,1}, U^C_{2,1}\} & \min_{R,C}\{U^R_{2,2}, U^C_{2,2}\} & \min_{R,C}\{U^R_{2,3}, U^C_{2,3}\} \\ \min_{R,C}\{U^R_{3,1}, U^C_{3,1}\} & \min_{R,C}\{U^R_{3,2}, U^C_{3,2}\} & \min_{R,C}\{U^R_{3,3}, U^C_{3,3}\} \end{pmatrix}$$

Figure 14: Examples of input feature matrices.

calculate the maximum, minimum, and mean of each row and column for both payoff matrices. Furthermore, F_1 and F_2 are the payoff matrices as they are, and lastly, we have a feature matrix where each value is the minimum payoff that either one of the players receives from the strategy profile. Below are three examples of such feature matrices.

$$\begin{pmatrix} \max_i U^R_{i,1} & \max_i U^R_{i,2} & \max_i U^R_{i,3} \\ \max_i U^R_{i,1} & \max_i U^R_{i,2} & \max_i U^R_{i,3} \\ \max_i U^R_{i,1} & \max_i U^R_{i,2} & \max_i U^R_{i,3} \end{pmatrix}, \quad \begin{pmatrix} \max_j U^R_{1,j} & \max_j U^R_{1,j} & \max_j U^R_{1,j} \\ \max_j U^R_{2,j} & \max_j U^R_{2,j} & \max_j U^R_{2,j} \\ \max_j U^R_{3,j} & \max_j U^R_{3,j} & \max_j U^R_{3,j} \end{pmatrix}$$

C.2 Softmax and Action Response Layers

After the last feature layer, a play distribution is calculated from each feature matrix in the last layer. This is done by first summing over the rows (columns) and then taking a softmax over the sums. The first action response layer is then given by a weighted average of those different distributions. For example, the distribution $S^R_1 \in \Delta^{m_R}$ is given by

$$S^R_1 = \text{softmax}\left(\sum_i (H^R_{2,1})_{1,i}, \sum_i (H^R_{2,1})_{2,i}, \ldots, \sum_i (H^R_{2,1})_{m_R,i}\right)$$

while the sums for the column player taken over the columns are given by

$$S^C_1 = \text{softmax}\left(\sum_j (H^C_{2,1})_{j,1}, \sum_j (H^C_{2,1})_{j,2}, \ldots, \sum_j (H^C_{2,1})_{j,m_C}\right).$$

The first action response distribution is then $\text{AR}^R(0) = \sum_l^{k_2} w^R_l S^R_l$ for $w^R \in \Delta^{k_2}$, and similarly for the column player.

The $\text{AR}^R(0)$ corresponds to a level-0 heuristic, i.e., a heuristic where the column player's behavior isn't explicitly modeled and taken into account. To do this, we move to Action Response layer 1, and use $\text{AR}^C(0)$ as a prediction for the behavior of the opposing player. Once the beliefs of the column player are formed, the $\text{AR}^R(1)$

calculates the expected value from each action, conditioned on that expected play, and takes a softmax over those payoffs:

$$\mathrm{AR}^R(1) = \mathrm{softmax}\left(\lambda \sum_j U^R_{1,j} \cdot \mathrm{AR}^C(0)_j, \ldots, \lambda \sum_j U^R_{m_R,j} \cdot \mathrm{AR}^C(0)_j\right)$$

As in the cognitive hierarchy model, the second Action Response layer, $\mathrm{AR}^R(2)$, forms a belief about the other player by taking a weighted average of the $\mathrm{AR}^R(1)$ and $\mathrm{AR}^R(0)$ layers and computing a noisy best response to it:

$$\mathrm{AR}^R(2) = \mathrm{softmax}\left(\lambda \sum_j U^R_{1,j} \cdot \left(\gamma \mathrm{AR}^C(0)_j + (1-\gamma)\mathrm{AR}^C(1)_j\right), \ldots\right)$$

C.3 Output Layer

The output layer takes a weighted average of the row player's action response layers. This is the final predicted distribution of play for the row player.

C.4 Cognitive Costs

When the deep heuristic is optimized with respect to the received payoff, the cognitive cost comes from two features of the network. Firstly, there is an assumed fixed cost associated with simulating, which is then proportional to the weight given to $\mathrm{AR}^R(1)$. Secondly, it is assumed that more exact predictions are cognitively more costly. The second cognitive cost is thus proportional to the reciprocal of the entropy of the resulting prediction.

D Detailed Results

D.1 Accuracy and Prediction Loss

In Table 6 we see the accuracy (how often the modal action is assigned the highest probability) and the average NLL of the different models.

D.2 Pairwise Tests

For Hypotheses 3 and 4 we can test significance with pairwise tests. For each of the games in the test set, we compare the difference in either the prediction loss or the payoff between the relevant models. For each game, we get two observations, one for

		Common		Competing		Total	
Model	Estimation	Accu	NLL	Accu	NLL	Accu	NLL
Deep heuristics	Fitted	89.4%	0.593	85.3%	0.709	87.3%	0.651
Metaheuristics	Fitted	88.4%	0.599	86.6%	0.715	87.5%	0.657
Metaheuristics	Optimized	89.1%	0.598	86.6%	0.726	87.8%	0.662
QCH+Pro	Fitted	85.3%	0.638	85.6%	0.722	85.5%	0.68
Deep heuristics	Optimized	85.3%	0.636	85.0%	0.739	85.2%	0.687
QCH	Fitted	82.2%	0.686	84.1%	0.737	83.1%	0.711
EB+Pro	Fitted	80.9%	0.717	71.2%	0.838	76.1%	0.777

Table 6: Average accuracy and negative log-likelihood for different models. Here we only report the models when estimated and evaluated on the same environments.

each role. For each of these comparisons, we perform both a t-test and a nonparametric, Wilcoxon rank test. As can be seen in the tables below, all of these tests are significant.

Model	Test set	Estimation	Difference	t-test	Wilcoxon
Metaheuristics	Common	Fitted	-0.065	$p < .001$	$p < .001$
Metaheuristics	Common	Optimized	-0.165	$p < .001$	$p < .001$
Metaheuristics	Competing	Fitted	-0.058	$p < .001$	$p < .001$
Metaheuristics	Competing	Optimized	-0.080	$p < .001$	$p < .001$
Deep heuristics	Common	Fitted	-0.113	$p < .001$	$p < .001$
Deep heuristics	Common	Optimized	-0.120	$p < .001$	$p < .001$
Deep heuristics	Competing	Fitted	-0.118	$p < .001$	$p < .001$
Deep heuristics	Competing	Optimized	-0.231	$p < .001$	$p < .001$

Table 7: Pairwise tests for differences in prediction loss in the test sets between the models estimated on training data from the same vs. the different environment. The prediction loss is lower for the model estimated on training data from the same environment for all pairs.

In Table 9 we see a pairwise test for the difference in the predictive ability between the optimized metaheuristic and the alternative models. Prosocial EB is a model with prosocial preferences and correct beliefs. We see that the optimized metaheuristic model is significantly better than the alternative models QCH, prosociality, and prosocial QCH.

Considering pairwise comparisons of models for each treatment in isolation, we see that the optimized metaheuristic makes better predictions than alternative models in the common-interests treatment. For the competing-interests treatment, the difference is not significant for either the QCH model with prosocial preferences or the standard QCH model.

Model	Test set	Estimation	Difference	t-test	Wilcoxon
Metaheuristics	Common	Fitted	-0.145	$p < .001$	$p < .001$
Metaheuristics	Common	Optimized	-0.302	$p < .001$	$p < .001$
Metaheuristics	Competing	Fitted	-0.083	$p < .001$	$p < .001$
Metaheuristics	Competing	Optimized	-0.165	$p < .001$	$p < .001$
Deep heuristics	Common	Fitted	-0.238	$p < .001$	$p < .001$
Deep heuristics	Common	Optimized	-0.333	$p < .001$	$p < .001$
Deep heuristics	Competing	Fitted	-0.276	$p < .001$	$p < .001$
Deep heuristics	Competing	Optimized	-0.502	$p < .001$	$p < .001$

Table 8: Pairwise tests for differences in regret in the test sets between the models estimated on training data from the same vs. the different environment. Regret is lower for the model estimated on training data from the same environment for all pairs.

Model	Estimation	Difference	t-test	Wilcoxon
Deep heuristics	Fitted	-0.011	$p = .003$	$p = .001$
Metaheuristics	Fitted	-0.005	$p = .079$	$p = .052$
QCH+Pro	Fitted	0.018	$p < .001$	$p = .001$
Deep heuristics	Optimized	0.025	$p < .001$	$p < .001$
QCH	Fitted	0.049	$p < .001$	$p < .001$
EB+Pro	Fitted	0.115	$p < .001$	$p < .001$

Table 9: Pairwise tests for differences in prediction loss between the optimized metaheuristic model and the alternative models across both treatments.

Model	Estimation	Difference	t-test	Wilcoxon
Deep heuristics	Fitted	-0.004	$p = .384$	$p = .194$
Metaheuristics	Fitted	0.001	$p = .801$	$p = .373$
Deep heuristics	Optimized	0.038	$p < .001$	$p < .001$
QCH+Pro	Fitted	0.040	$p < .001$	$p < .001$
QCH	Fitted	0.088	$p < .001$	$p < .001$
EB+Pro	Fitted	0.119	$p < .001$	$p < .001$

Table 10: Pairwise tests for differences in prediction loss between the optimized metaheuristic model and the alternative models for the common-interests games.

E Explaining Adaptation via Learning

In the main text, we assume that the participants manage to find rational heuristics without going into the details about how that is done. Here, we show that a learning model could explain this adaptation to rational metaheuristics.

We assume that all individuals arrive at the experiment with the same initial metaheuristic $m(\cdot \mid \theta(0))$, where θ are the parameters of the metaheuristic, including the parameters of both the primitive heuristics and the priors.

Model	Estimation	Difference	t-test	Wilcoxon
Deep heuristics	Fitted	-0.017	$p < .001$	$p = .001$
Metaheuristics	Fitted	-0.011	$p = .011$	$p = .059$
QCH+Pro	Fitted	-0.004	$p = .483$	$p = .619$
QCH	Fitted	0.010	$p = .146$	$p = .062$
Deep heuristics	Optimized	0.013	$p = .090$	$p = .151$
EB+Pro	Fitted	0.112	$p < .001$	$p < .001$

Table 11: Pairwise tests for differences in prediction loss between the optimized metaheuristic model and the alternative models for the competing-interests games.

For each experimental population ξ, the players play a sequence of $(G_{\xi,t})_{t=1}^{50}$, each time with a single realized action of the other player. Given the observed behavior of player $-i$, the utility in round t for player i is given by

$$u(m(\cdot \mid \theta), G_{\xi,t}, s_{-i}, c) = \pi_{G_{\xi,t}}\left(m(G_{\xi,t} \mid \theta), s_{-i}\right) - c(m(\cdot \mid \theta)),$$

where $m(\cdot \mid \theta)$ is the metaheuristic with parameters θ, $G_{\xi,t}$ is the game played in round t by population ξ, c is the cognitive cost function, and s_{-i} is the action taken by the other player.

After observing the action s_{-i} taken by the other player, player i can calculate the gradient with respect to the parameters to see how the metaheuristic used could have been improved, i.e.,

$$\nabla_\theta u(m(\cdot \mid \theta), G_{\xi,t}, s_{-i}, c).$$

A simple learning model is one where each individual changes the metaheuristic used in the direction of the gradient after each round of the experiment, with some step-size κ. We can write this as

$$\theta_{\xi,i}(t+1) = \theta_{\xi,i}(t) + \kappa \nabla_\theta u(m(\cdot \mid \theta_{\xi,i}(t)), G_{\xi,t}, s_{-i}, c).$$

In other words, after each game, the metaheuristic is moved in the direction that would have yielded a higher utility in that game.

For simplicity, we consider a population-level model, rather than modeling the behavior of each individual player separately. The behavior in round t is given by

$$\theta_\xi(t+1) = \theta_\xi(t) + \kappa \mathbb{E}_{s_{-i} \sim P_\xi(\cdot \mid G_{\xi,t})}\left[\nabla_\theta u(m(\cdot \mid \theta_{\xi,i}(t)), G_{\xi,t}, s_{-i}, c)\right],$$

where $P_\xi\left(s_{-i} \mid G_{\xi,t}\right)$ is the empirical probability that s_{-1} is used in game $G_{\xi,t}$. Thus, after each game, the population parameters for the next round move in the average

direction of improvement defined by the empirical behavior in that game.

In our estimation of the learning model, we use the costs estimated for the optimal metaheuristics. To estimate this model we thus need a baseline heuristic, $\theta(0)$, and a learning parameter, κ. To make the performance of this model comparable to that of the other models, we estimate the common starting parameters $\theta(0)$ and the common learning rate κ in order to minimize loss on the first 30 games of each population in both treatments. We then predict the remaining 16 treatment games of each population in both treatments.

Model	Estimation	Common	Competing	Both
Deep heuristics	Optimize	0.636	0.739	0.687
Deep heuristics	Fit	0.593	0.709	0.651
Metaheuristics	Optimize	0.598	0.726	0.662
Metaheuristics	Fit	0.599	0.715	0.657
Learning		0.605	0.724	0.664

Table 12: Out-of-sample NLL prediction loss.

In Table 12 we see that the performance of the learning model is comparable to but slightly lower than the performances of the fitted models. In Table 13 the expected payoffs in the test set games are shown for the learning model, the optimized metaheuristics, the optimized deep heuristics, and relevant benchmarks. It is clear that the expected payoffs from this learning model are similar to both the actual payoffs and those of the optimization-based models.

Model	Estimation	Common	Competing
Metaheuristic	Optimize	6.69	5.43
Deep heuristic	Optimize	6.65	5.45
Learning		6.68	5.38
Random		5.38	4.45
Human behavior		6.74	5.43
Maximum		7.17	5.95

Table 13: Out-of-sample expected payoffs.

In conclusion, this simple learning model appears to be a possible explanation for how the participants come to use these near-optimal heuristics in our experiment with simple adjustments of the heuristics used after each game.

References

Bacharach, M. (2006). *Beyond individual choice: teams and frames in game theory*. Princeton University Press.

Bardsley, N., J. Mehta, C. Starmer, and R. Sugden (2010). Explaining focal points: Cognitive hierarchy theory versus team reasoning. *Economic Journal 120*(543), 40–79.

Bruhin, A., E. Fehr, and D. Schunk (2019). The many faces of human sociality: Uncovering the distribution and stability of social preferences. *Journal of the European Economic Association 17*(4), 1025–1069.

Camerer, C. F. (2003). *Behavioral Game Theory: Experiments in Strategic Interaction*. Princeton University Press.

Camerer, C. F., T.-H. Ho, and J.-K. Chong (2004, aug). A Cognitive Hierarchy Model of Games. *The Quarterly Journal of Economics 119*(3), 861–898.

Caplin, A. and M. Dean (2013). Behavioral implications of rational inattention with shannon entropy. Technical report, National Bureau of Economic Research.

Chen, D. L., M. Schonger, and C. Wickens (2016, March). oTree—An open-source platform for laboratory, online, and field experiments. *Journal of Behavioral and Experimental Finance 9*, 88–97.

Costa-Gomes, M. A. and G. Weizsäcker (2008, jul). Stated Beliefs and Play in Normal-Form Games. *Review of Economic Studies 75*(3), 729–762.

Crawford, V. P., M. A. Costa-Gomes, and N. Iriberri (2013, mar). Structural Models of Nonequilibrium Strategic Thinking: Theory, Evidence, and Applications. *Journal of Economic Literature 51*(1), 5–62.

Devetag, G., S. Di Guida, and L. Polonio (2016, mar). An eye-tracking study of feature-based choice in one-shot games. *Experimental Economics 19*(1), 177–201.

Dhami, S. (2016). *The foundations of behavioral economic analysis*. Oxford University Press.

Ert, E. and I. Erev (2013). On the descriptive value of loss aversion in decisions under risk: Six clarifications. *Judgment and Decision Making 8*(3), 214–235.

Fehr, E. and K. M. Schmidt (1999). A theory of fairness, competition, and cooperation. *The Quarterly Journal of Economics 114*(3), 817–868.

Fudenberg, D., F. Drew, D. K. Levine, and D. K. Levine (1998). *The theory of learning in games*, Volume 2. MIT press.

Fudenberg, D., J. Kleinberg, A. Liang, and S. Mullainathan (2022). Measuring the completeness of economic models. *Journal of Political Economy 130*(4), 956–990.

Fudenberg, D. and A. Liang (2019, dec). Predicting and Understanding Initial Play. *American Economic Review 109*(12), 4112–4141.

Gershman, S. J., E. J. Horvitz, and J. B. Tenenbaum (2015). Computational rationality: A converging paradigm for intelligence in brains, minds, and machines. *Science 349*(6245).

Gigerenzer, G. and P. M. Todd (1999). *Simple Heuristics That Make Us Smart*. Oxford University Press, USA.

Goeree, J. K. and C. A. Holt (2004). A model of noisy introspection. *Games and Economic Behavior 46*(2), 365–382.

Goldstein, D. G. and G. Gigerenzer (2002). Models of ecological rationality: The recognition heuristic. *Psychological Review 109*(1), 75–90.

Griffiths, T. L., F. Lieder, and N. D. Goodman (2015). Rational use of cognitive resources: Levels of analysis between the computational and the algorithmic. *Topics in Cognitive Science 7*(2), 217–229.

Grimm, V. and F. Mengel (2012). An experiment on learning in a multiple games environment. *Journal of Economic Theory 147*(6), 2220–2259.

Hartford, J. S., J. R. Wright, and K. Leyton-Brown (2016). Deep Learning for Predicting Human Strategic Behavior. In *Advances in Neural Information Processing Systems*, Volume 29. Curran Associates, Inc.

Heap, S. H., D. R. Arjona, and R. 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.

Hebert, B. and M. Woodford (2019). Rational inattention when decisions take time. *Journal of Chemical Information and Modeling 53*(9), 1689–1699.

Howes, A., R. L. Lewis, and A. Vera (2009). Rational Adaptation Under Task and Processing Constraints: Implications for Testing Theories of Cognition and Action. *Psychological Review 116*(4), 717–751.

Imai, T., T. A. Rutter, and C. F. Camerer (2020, sep). Meta-Analysis of Present-Bias Estimation Using Convex Time Budgets. *The Economic Journal 186*(2), 227–236.

Izard, V. and S. Dehaene (2008). Calibrating the mental number line. *Cognition 106*(3), 1221–1247.

Jehiel, P. (2005, aug). Analogy-based expectation equilibrium. *Journal of Economic Theory 123*(2), 81–104.

Krueger, P., F. Callaway, S. Gul, T. Griffiths, and F. Lieder (2022, January). Discovering Rational Heuristics for Risky Choice.

Lewis, R. L., A. Howes, and S. Singh (2014). Computational rationality: Linking mechanism and behavior through bounded utility maximization. *Topics in Cognitive Science 6*(2), 279–311.

Lieder, F. and T. L. Griffiths (2015). When to use which heuristic: A rational solution to the strategy selection problem. *Proceedings of the 37th annual conference of the cognitive science society 1*(3), 1–6.

Lieder, F. and T. L. Griffiths (2017, November). Strategy selection as rational metareasoning. *Psychological Review 124*(6), 762–794.

Lieder, F. and T. L. Griffiths (2020). Resource-rational analysis: Understanding human cognition as the optimal use of limited computational resources. *Behavioral and Brain Sciences 43*, e1.

Lieder, F., P. M. Krueger, and T. Griffiths (2017). An automatic method for discovering rational heuristics for risky choice. In *Proceedings of the 39th Annual Meeting of the Cognitive Science Society*.

Matějka, F. and A. McKay (2015, jan). Rational Inattention to Discrete Choices: A New Foundation for the Multinomial Logit Model. *American Economic Review 105*(1), 272–298.

Mengel, F. and E. Sciubba (2014). Extrapolation and structural similarity in games. *Economics Letters 125*(3), 381–385.

Misyak, J. B. and N. Chater (2014). Virtual bargaining: A theory of social decision-making. *Philosophical Transactions of the Royal Society B: Biological Sciences 369*(1655).

Nagel, R. (1995). Unraveling the Guessing Game. *American Economic Review 85*(5), 1313–1326.

Peysakhovich, A. and D. G. Rand (2016). Habits of virtue: Creating norms of cooperation and defection in the laboratory. *Management Science 62*(3), 631–647.

Polonio, L., S. Di Guida, and G. Coricelli (2015, nov). Strategic sophistication and attention in games: An eye-tracking study. *Games and Economic Behavior 94*, 80–96.

Savage, L. J. (1954). *The Foundations of Statistics*. John Wiley & Sons.

Simon, H. A. (1976). From substantive to procedural rationality. In *25 years of economic theory*, pp. 65–86. Springer.

Sims, C. A. (1998). Stickiness. *Carnegie-Rochester Conference Series on Public Policy 49*, 317–356.

Spiliopoulos, L. and R. Hertwig (2020, March). A map of ecologically rational heuristics for uncertain strategic worlds. *Psychological Review 127*(2), 245–280.

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

Stahl, D. O. and P. W. Wilson (1995). On players' models of other players: Theory and experimental evidence.

Steiner, J., C. Stewart, and F. Matějka (2017). Rational Inattention Dynamics: Inertia and Delay in Decision-Making. *Econometrica 85*(2), 521–553.

Stewart, N., S. Gächter, T. Noguchi, and T. L. Mullett (2016). Eye Movements in Strategic Choice. *156*(October 2015), 137–156.

Sugden, R. (2003). The logic of team reasoning. *Philosophical explorations 6*(3), 165–181.

Todd, P. M. and G. E. Gigerenzer (2012). *Ecological Rationality: Intelligence in the World.* Oxford University Press.

Tunçel, T. and J. K. Hammitt (2014). A new meta-analysis on the WTP/WTA disparity. *Journal of Environmental Economics and Management 68*(1), 175–187.

Wright, J. R. and K. Leyton-Brown (2017, nov). Predicting human behavior in unrepeated, simultaneous-move games. *Games and Economic Behavior 106*(2), 16–37.