Rational Heuristics for One-Shot Games*

Frederick Callaway[1], Thomas L. Griffiths[2], and Gustav Karreskog[†3]

[1,2]Department of Psychology, Princeton
[3]Department of Economics, Stockholm School of Economics

March 29, 2021

For the most recent version, click here

Abstract

Insights from behavioral economics suggest that perfect rationality is an insufficient model of human decision-making. However, the empirically observed deviations from perfect rationality or biases vary substantially among environments. There is, therefore, a need for theories that inform us when and how we should expect deviations from rational behavior. We suggest that such a theory can be found by assuming optimal use of limited cognitive resources. In this paper, we present a theory of human behavior in one-shot interactions based on the rational use of heuristics. We test our theory by defining a broad family of heuristics for one-shot games and associated cognitive cost functions. In a large, preregistered experiment, we find that behavior is well predicted by our theory, which yields better predictions than existing models. We find that the participants' actions depend on their environment and previous experiences, in the way the rational use of heuristics suggest.

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, SUDSWEC, UCL, NHH, and Princeton, for helpful comments and insights. This work was supported by a grant to TLG by the Templeton foundation, Tom Hedelius Foundation, and Knut and Alice Wallenberg Research Foundation,

†gustav.karreskog@phdstudent.hhs.se

1 Introduction

A key assumption underlying classical economic theory is that people behave optimally in order to maximize their 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 (Kahneman, 2011). This suggests we can improve our understanding by incorporating more realistic behavioral components into our models of economic behavior. While many of these deviations are indeed systematic and show up in multiple studies, the estimated biases vary considerably between studies and contexts. Apparent biases change or even disappear if participants have opportunities for learning or if the details of the decision task change. For example, this is the case for the endowment effect (Tunçel and Hammitt, 2014), loss aversion (Ert and Erev, 2013), numerosity underestimation (Izard and Dehaene, 2008), and present bias (Imai, Rutter and Camerer, 2020).

In order to incorporate behavioral effects into theories with broader applications, without having to run new experiments for every specific setting, we need a theory that can account for these variations. That is, we need a theory that can help us understand why and predict when we should expect deviations from the rational benchmark, and when we can safely assume behavior is close to rational. 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. In this paper, we present an explicit version of this theory tailored to one-shot games and test it experimentally.

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

The most common theories for 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. Examples of such models are so-called level-k models, introduced by Nagel (1995); Stahl and Wilson (1994, 1995), and closely related Cognitive Hierarchy (CH) models, introduced by Camerer, Ho and Chong (2004), or models of noisy introspection (Goeree and Holt, 2004). In such models, participants are characterized by different levels of reasoning. Level-0 reasoners behave naively, often assumed to play a uniformly random strategy. Level-1 reasoners best respond to level-0 behavior, and even higher levels best respond to

behavior based on lower level reasoning. In meta-analyses such as Crawford, Costa-Gomes and Iriberri (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, empirical evidence on information acquisition and elicited beliefs is often inconsistent with such a belief-response process. When participants are asked to state their beliefs about how the opponent will play, they often fail to play a best response to those beliefs (Costa-Gomes and Weizsäcker, 2008). 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, Di Guida and Coricelli, 2015; Devetag, Di Guida and Polonio, 2016; Stewart et al., 2016). Furthermore, the estimated parameters of these models often vary considerably between different data sets, behavior seems to depend on the underlying game in a way not captured by the models (Bardsley et al., 2010; Heap, Arjona and Sugden, 2014), and there is evidence of earlier games played having an effect on behavior not captured by existing models (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, 2019). That is, we assume that people use simple cognitive strategies that flexibly and selectively process payoff information to construct a decision with minimal cognitive effort. These heuristics do not necessarily involve any explicit construction of beliefs to which the players best respond. However, we assume that people adapt the heuristics in order to maximize utility. Although they might not choose the best action in a given game, they will learn which heuristics generally work well in an environment.[1]

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 optimization and the notion that people use simple strategies that are effective for the problems they actually need to solve. The key assumption in this approach, *resource-rational analysis*, is that people use cognitive strategies that make optimal use of their limited computational resources (Lieder and Griffiths, 2019; Griffiths, Lieder and Goodman, 2015 c.f. Lewis, Howes and Singh, 2014; Gershman, Horvitz and Tenenbaum, 2015).

In comparison with traditional rational models, resource-rational analysis is distinctive in that it explicitly accounts for the cost of allocating limited computational resources to a given decision. It specifies an objective function that includes both the utility of a decision's outcome as well as the cost of the cognitive process that produced the decision. In comparison with theories of bounded or ecological rationality (Gigerenzer and Todd, 1999; Goldstein and Gigerenzer, 2002; Smith, 2003; Todd and Gigerenzer, 2012), resource-rational analysis is distinctive in its assumption that people optimize this objective function. This makes it possible to predict when people will use one

[1]This idea is related to that of procedural rationality in Simon (1976).

heuristic versus another (Lieder and Griffiths, 2017) and even to automatically discover novel heuristics (Lieder, Krueger and Griffiths, 2017).

Finally, our approach is perhaps most compatible with information-theoretic approaches such as rational inattention (Matějka and McKay, 2015; Sims, 1998; Caplin and Dean, 2013; Hebert and Woodford, 2019; Steiner, Stewart and Matějka, 2017), in which the costs and benefits of information processing are optimally traded off. Resource-rational analysis is distinct, however, in making stronger assumptions about the specific computational processes and costs that are likely to be involved in a given domain.

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 to be used. For example, in an environment in which most interactions are characterized by competing interests (e.g., zero-sum games), a good heuristic is one that looks for actions with high guaranteed payoffs. On the other hand, if most interactions are common interest, focusing on the guaranteed payoff will lead to many missed opportunities for mutually beneficial outcomes, so it might be better to look for the common interest. This is the key insight that allows us to test our theories.

To examine whether people adapt their heuristics to the environment, as our theory predicts, we conduct a large, preregistered[2] 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 interest* environment, there is a positive correlation between the payoffs of the row and column player over the set of strategy profiles. In the *competing interest* environment, the correlation is negative. As a result, the games in the common interest environment are often such that there is a jointly beneficial outcome for the players to coordinate on. In contrast, the games in the competing interest environment are similar to zero-sum games where one player's loss is the other's gain. Interspersed among these treatment-specific games, we include four *comparison games*, which are the same for both conditions (and all sessions). If the participants are using environment-adapted heuristics to make decisions, and different heuristics are good for common interest and competing interest environments, the participants should behave differently in the comparison games since they are employing different heuristics. Indeed, this is what we observe.

To take our analysis further, we define a parameterized family of heuristics and cognitive costs in order to test the critical resource-rational hypothesis that our participant's behavior is consistent with an optimal tradeoff between payoff and cognitive cost. Rather than identifying the parameters that best fit human behavior we identify the parameters that strike this optimal tradeoff, and ask how well they predict the effect of the environment on human behavior. Although we fit the cost function parameters that partially define the resource-rational heuristic—critically—these parameters are fit jointly to data in both treatments. Thus, any difference in predicted behavior is an *a priori* prediction. Strikingly, we find that this model, which has no free parameters

[2]The preregistration is embargoed at the open science foundations preregistration platform. Email the author Gustav Karreskog at gustav.karreskog@phdstudent.hhs.se if you need access to it.

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.

We will start by introducing the general model in Section 2, capturing the connection between heuristics, their associated cognitive costs, the environment, and resource-rationally optimal heuristics. In Section 3, we introduce our main specification of the available heuristics and their cognitive costs, metaheuristics. We then introduce the experiment in Section 4, followed by the model-free results based on the comparison games. We there confirm that the two different environments indeed lead to large and predictable differences in behavior. After that, we test the two model-based hypotheses using the metaheuristics. Based on these out-of-sample predictions, we show that the differences in behavior between the different treatments can be accurately predicted by assuming that the participants use the optimal metaheuristics in the respective environments. In Section 5, we can confirm the model-based hypothesis also by considering a completely different representation of the possible heuristics using a constrained neural network design, which we call *deep heuristics*. Lastly, in Section 6, we compare our model to a quantal cognitive hierarchy model and a model with noisy-best reply and pro-social preferences and show that our model predicts behavior better than both these alternatives.

2 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, $c(h) \in \mathbb{R}_+$.[3] 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

[3] In general, we can imagine that the cognitive cost depends on both the heuristic and the game, for example, it might be more costly to apply it to a 5×5 game than a 2×2 game. But since all our games will be 3, we drop that dependency.

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

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

Since each heuristic has an associated cognitive cost, the actual expected utility derived 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 available heuristics, 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 opponent heuristic combinations a player is likely to face. Given an environment, we can calculate the expected performance of a heuristic as

$$V(h_i, \mathcal{E}) = \mathbb{E}_{\mathcal{E}} \left[u_i \left(h_i, h_{-i}, G \right) \right] = \sum_{G \in \mathcal{G}} \sum_{h_{-i} \in \mathcal{H}} u_i \left(h_i, h_{-i}, G \right) \cdot P(G, h_{-i}). \tag{1}$$

We can also evaluate the performance of a heuristic conditional on the specific game being played

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

We can now define formally what we mean with a rational, or optimal, heuristic. A rational heuristic h^* is a heuristic that optimizes (1), i.e.,

$$h^* = \operatorname*{argmax}_{h \in \mathcal{H}} V(h, \mathcal{E}).$$

We here also see that by varying the environment, we can vary which heuristics are optimal. In the experiment, we will vary $\mathcal{P}$, thereby varying the predictions we get by assuming rational heuristics.

3 Metaheuristics

To build a formal model of heuristics for one-shot games, we begin by specifying a small set of candidate forms that such a heuristic might take: row-based reasoning, cell-based reasoning, and simulation-based reasoning. We specify a precise functional form for

each class, each employing a small number of continuous parameters and a cognitive cost function. The cognitive cost of a heuristic is a function of its parameters, and the form of 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. However, we hypothesized, and our results show that it is expressive enough to illustrate the general theory's predictions and provide a strong quantitative explanation of human behavior.

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 1: Example normal form game represented as a bi-matrix. The row player chooses a row and 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.

3.1 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 first entries in each cell of the corresponding row of the payoff matrix (see Figure 1). 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 $f : \mathbb{R}^m \to \mathbb{R}$. 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 results in level-1 like behavior. Indeed, 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 chose the row with highest minimum value,

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

and similarly the maximax heuristic when f is max,

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

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, with mean in the center. We construct such a family with 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 example game above (Figure 1), the heuristic will assign highest value to s_1 (the top row) when γ is large and positive, s_2 when γ is large and negative, and s_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 over-weighted, but these might correspond to different columns in each row; in the example game (Figure 1), column 3 would be over-weighted when evaluating row 1 but down-weighted when evaluating rows 2 and 3. Although this internally inconsistent weighting may appear irrational, this extra degree of freedom can increase the expected payoff in a given environment without necessarily being more cognitively taxing.

Given a row-value function v, the most obvious way to select an action would be to select $\arg\max_{s_i} v$. However, exactly maximizing even a simple function may be challenging for an analog computer such as the human brain. Thus, 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

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

Naturally, the cost of a row heuristic is a function of the cognitive effort. Specifically, we assume that the cost is proportional to effort,

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

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

3.2 Cell Heuristics

An individual might not necessarily consider all aspects connected to a strategy, but find a good "cell", meaning payoff pair $(\pi_1(s_1, s_2), \pi_2(s_1, s_2))$. In particular, previous research has proposed that people sometimes adopt a *team view*, looking for outcomes

that are good for both players, and choosing 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*, selecting the outcome that would be agreed upon if they 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, thereby ignoring the dependency of the other player's behavior, as *cell heuristics*. Based on preliminary analyses, we identified one specific form of cell heuristic that participants appear to use frequently: This *jointmax* heuristic identifies the outcome that is most desirable for both players, formally

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

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

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

In the example game (Figure 1), 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 the accuracy, so

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

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

3.3 Simulation Heuristics - Higher level reasoning

The row and cell heuristics don't construct explicit beliefs about how the other player will behave.[4] Belief formation and best response has formed the basis of many previous models of initial play, and might very well be a sensible thing to do. We consider such a decision-making process as one possible heuristic people might use.

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 use to estimate the column players behavior, then $h_{\text{sim}}(G)$ is given by

[4] They might do so implicitly, however. For example, a row heuristic that assigns a higher weight to high payoffs works well only if the other player is more likely to play those columns. Ignoring the low payoffs might correspond to an implicit belief that the other player will not play those columns.

$$h_{\text{row}}^{s_r} = \frac{\exp\left[\varphi \cdot \left(\sum_{s_c \in S_c} \pi_r(s_r, s_c) \cdot h_{col}^{s_c}(G^T)\right)\right]}{\sum_{s_r \in S_r} \exp\left[\varphi \cdot \left(\sum_{s_c \in S_c} \pi_r(s_r, s_c) \cdot h_{col}^{s_c}(G^T)\right)\right]}$$

where φ is the cognitive effort parameter. A simulation heuristic is thus defined by a combination of a heuristic and a 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 players 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,

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

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.

3.4 Metaheuristic

We don't expect a person to use the same heuristic in all games. Instead, they may have a set of heuristics, choosing which one to use in each situation based on an estimate of the candidate heuristics' expected value. We model such a process as a higher-order heuristic that selects among the first-order heuristics described above. We call this heuristic-selecting heuristic a metaheuristic.

Rather than explicitly modeling the process by which players select among the candidate heuristics, for example, 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 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

$$\begin{aligned}\mathbb{P}\left[\{\text{use } h^n \text{ in } G\}\right] &= \frac{\exp\left[(a_n + V(h^n, \mathcal{E}, G))/\lambda\right]}{\sum_{j=1}^{N} \exp\left[(a_j + V(h^j, \mathcal{E}, G))/\lambda\right]} \\ &= \frac{p_n \exp\left[V(h^n, \mathcal{E}, G)/\lambda\right]}{\sum_{j=1}^{N} p_j \exp\left[V(h^j, \mathcal{E}, G)/\lambda\right]} \end{aligned} \tag{2}$$

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

3.4.1 The individual's optimization problem

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 set of consideration heuristics, and $P = \{p^1, p^2, \ldots, p^N\}$ a prior over those heuristics. We can write down the performance of a metaheuristic in an environment $\mathcal{E}$, analogously to (1) for heuristics, as

$$V^{meta}(m, \mathcal{E}) = \sum_{G \in \mathcal{G}} \sum_{h \in H} V(h^n, \mathcal{E}, G) \cdot \frac{p_n \exp\left[V(h^n, \mathcal{E}, G)/\lambda\right]}{\sum_{j=1}^{N} p_j \exp\left[(V(h^j, \mathcal{E}, G))/\lambda\right]} \cdot P(G) \qquad (3)$$

The optimization problem faced by the individual, subject to the adjustment cost λ, is then to maximize (3), 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 set of possible heuristics, $\mathcal{H}$, is infinite. Even with a finite set of heuristics, the size of the power set will grow very quickly. Therefore, we will assume a small set of heuristics and jointly find optimal parameters of those heuristics and priors P.

4 Experiment

Our overarching hypothesis is that individuals choose actions in one-shot games using heuristics that optimally trade off between expected payoff and cognitive cost. It is unlikely, however, that people compute these expected payoffs each time they need to make a decision. Instead, we hypothesize that people *learn* to use heuristics that are generally adaptive in their environment. This results in a critical 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. We test this prediction in a large, online experiment in which participants play one-shot normal form games.

4.1 Methods

We recruited 600 participants on Amazon Mechanical Turk using the oTree platform (Chen, Schonger and Wickens, 2016). Each participant was assigned to one of 20 populations of 30 participants each. They then played 50 different one-shot normal form games, in each period randomly matched to a new participant in their population.

Each population was assigned to one of two experimental treatments, which determined the distribution of games that were played. Specifically, we manipulated the correlation

between the row and column players' payoffs in each cell. In the *common interest* 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. In contrast, in the *competing interest* 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. 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 non-negative integers.[5] Examples of each type of *treatment game* are shown in Tables 1 and 2.

For each population, we sampled 46 treatment games, each participant playing every game once. The remaining four games were *comparison games*, treatment-independent games that we used to compare behavior in the two treatments when playing the same game. The comparison games were played in rounds 31, 38, 42, and 49. We placed them all later in the experiment so that the participants would have time to adjust to the treatment environment first, leaving gaps to minimize the chance that participants would notice that these games were different from the others they had played.

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

Common interest example 1

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

Common interest example 2

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

Common interest example 3

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

Common interest example 4

Table 1: Four games from the common interest treatment.

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

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

Competing interest example 1

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

Competing interest example 2

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

Competing interest example 3

$8,0$	$4,1$	$3,8$
$4,7$	$2,7$	$2,7$
$3,5$	$3,9$	$7,5$

Competing interest example 4

Table 2: Four games from the competing interest treatment.

4.1.1 The Comparison Games

We selected 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 a row that gives a high guaranteed pay off. 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.

First Comparison Game

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

Comparison game 1

The first game is a weak-link game, where all the diagonal strategy profiles are Nash Equilibria, but all are not as efficient. The most efficient NE gives payoffs (8,8), but it is also possible to get 0. The least efficient equilibrium yields a payoff of (5,5), but that is also the guaranteed payoff. The equilibrium (6,6) is in between the two in terms of both risk and efficiency. The third row has the highest average payoff and is the best response to itself, so any standard level-k model would predict (5,5) being played.

Second Comparison Game

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

Comparison game 2

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 the common interest treatment to play strategy 1 more often since it is usually a good heuristic for them to look for the common interest.

Third Comparison Game

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

Comparison game 3

The third game is a game with two NE, where one is the pure NE $(3, 3)$, and the other is a mixed NE involving **1** and **2**. This game is constructed so that the row averages are much higher for strategy **1** and **2** compared to **3**, meaning that any level-k heuristic ends up there, while the NE yielding $(9, 9)$ is much more efficient. So, there is a strong tension between efficiency and guaranteed payoff.

Fourth Comparison Game

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

Comparison game 4

In this 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**.

4.2 Model estimation and evaluation

We take an out-of-sample prediction approach to model comparison. Each data set is divided into a training set and a test set. The models are estimated on the training data and evaluated on the test data. The training data consisted of the first 30 games of each session, and the other 16 treatment games are the test data. We consider each game as two observations, one for empirical distribution of play for each player role. The games are sampled separately for each session, but are the same within a session, and we have 10 sessions for each treatment. For each treatment, we thus have 600 observations in the training games and 320 observations of in the test games. This separation was preregistered, and can thus be considered a "true" out of sample prediction.

We define the two different environments with the actual games and empirical distributions of play in the corresponding sessions. We thus define the common interest environment, $\mathcal{E}^+$, by letting $\mathcal{G}^+$ be all the treatment games played in the common interest treatment, and let the opponents behavior always be given by the actual distribution of play, so $h^+(G)$ returns the actual distribution of play in G. Lastly, P is uniform distribution over all games in $\mathcal{G}^+$, and always returns h^+ as the heuristic for the opponent. We define the competing interest environment $\mathcal{E}^-$ in the corresponding way. Lastly, we can divide the games in to the training games, e.g., $\mathcal{G}^+_{\text{train}}$, 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), so a lower value means a better fit. If y is the observed distribution of play for for some role in some game, and x is the predicted distribution of play from some model, the negative log-likelihood (NLL) is defined

$$\text{NLL}(x, y) = -\sum_s y_s \cdot \log(x_s).$$

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

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

and analogously for the other possible environments. 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 opponents behavior, h^+, and the cognitive costs, C, via Equation (2).

The metaheuristics described previously have several free parameters that control their behavior, the parameters of the primitive heuristics and the priors for the different primitive heuristics. We consider two methods for estimating these parameters and

the cognitive costs. Fitting the parameters to the data, or optimizing the parameters such that they maximize expected utility.

For a given set of cognitive cost parameters $C = (C_{\text{row}}, C_{\text{cell}}, C_{\text{mul}}, \lambda)$, the *fitted* common interest 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 metaheuristics we consider consists of three primitive heuristics, a jointmax cell heuristic, a row heuristic, and a simulation heuristic, where a row heuristic models the other player's behavior.

The *optimal* common interest metaheuristic, for cognitive costs C, is instead given by

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

The fitted and optimal metaheuristics for the competing interest environment are defined in the analogous way.

Having defined the fitted and optimal heuristics for given 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 the two treatments.

To estimate the costs, we find the costs the minimizes the average NLL of the optimize, or fitted, heuristics on the training data. So

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

and

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

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

4.3 Results

We organize our results based on our four pre-registered hypotheses. The first two are model-free and concern the behavior in the comparison games. The latter two are model-based and concern the behavior in the treatment games.

4.3.1 Model-free analysis of comparison games

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

Hypothesis 1. *The distribution of play in the four comparison games will be different in the two treatment populations.*

This hypothesis follows from the assumption that people learn to use heuristics that are adaptive in their treatment and that different heuristics are adaptive in the two treatments. Figure 2 visually confirms this prediction and Table 3 confirms that these differences are statistically significant (χ^2-tests, as preregistered).

	Frequencies				
	1	**2**	**3**	χ^2	p-value
Comparison Game 1				98.39	$2.2 \cdot 10^{-16}$
Common interest	193	53	54		
Competeting interest	75	82	143		
Comparison Game 2				22.08	$1.6 \cdot 10^{-5}$
Common interest	160	139	1		
Competeting interest	103	195	2		
Comparison Game 3				61.75	$3.9 \cdot 10^{-14}$
Common interest	40	73	187		
Competeting interest	106	97	97		
Comparison Game 4				91.36	$2.2 \cdot 10^{-16}$
Common interest	78	173	49		
Competeting interest	115	62	123		

Table 3: χ^2 tests for each comparison games. All of the them significant at the preregistered 0.05 level.

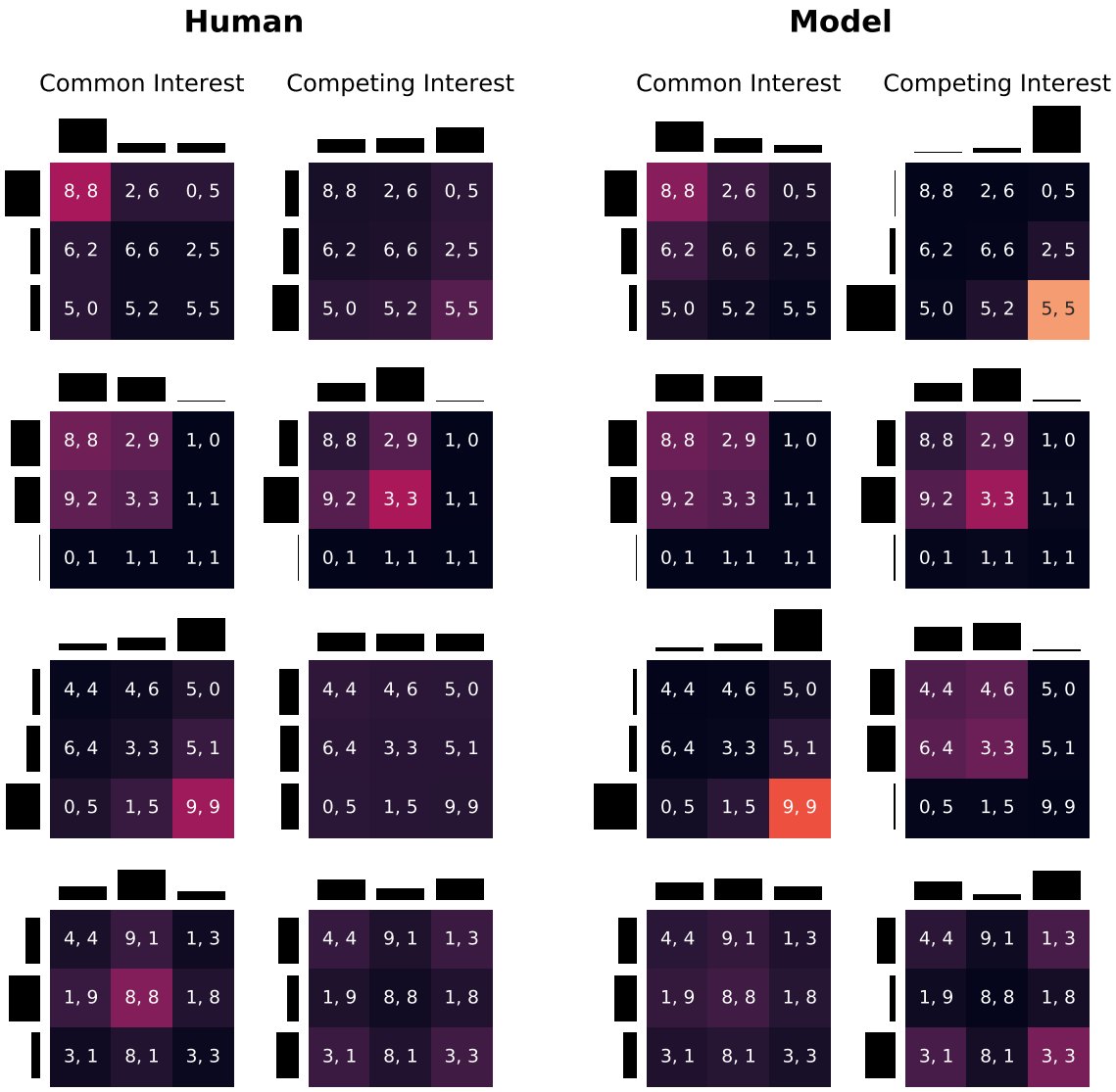


Figure 2: Distribution of plays in the four comparison games. Each panel shows the joint and marginal distributions of row/column plays in a single game. The cells are annotated with each player's payoffs for the given outcome. The two columns to the left show the actual behavior in the two environments, while the two columns to the right show the predictions of the rational (optimized) metaheuristics.

Inspecting Figure 2, we see that the distribution of play is not just different in the two groups; it is different in a systematic way. In particular, players in the common interest 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 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 interest treatment than in the competing interest treatment.*

Since the comparison games were chosen to exhibit a tension between the efficient outcome and a high guaranteed payoff, we expected that the common interest population would better coordinate on the efficient outcome. In our metaheuristic model, this prediction results from the fact that the jointmax heuristic generally performs quite well in the common interest games. Thus, the cognitive cost of checking for each game, whether another heuristic performs better, generally outweighs the potential gains. Coordinating on the efficient outcomes then leads to a higher average payoff. Table 4 confirms this prediction. The common interest population had a higher average payoff in all four comparison games, and the difference is significant in each case (at the pre-registered level of $p < .05$).

	Treatment average payoff			
	Common interest	Competing interest	t-value	p-value
Comparison game 1	5.09	3.64	6.851	$2 \cdot 10^{-11}$
Comparison game 2	5.52	4.04	6.28	$6.7 \cdot 10^{-10}$
Comparison game 3	5.00	4.31	2.86	0.0044
Comparison game 4	5.19	3.42	7.21	$1.9 \cdot 10^{-12}$

Table 4: Two-sided t-tests for the difference in average payoff between the two treatments in the comparison games.

4.3.2 Model-based analysis of treatment games

Next, we consider our two model-based hypotheses regarding the metaheuristic model's ability to capture the difference in strategies used in the two treatments. The first hypothesis is that the behavior will be different and that the model will capture some aspect of that difference.

Hypothesis 3. *Behavior in the two treatments is consistently different. In other words, the common interests metaheuristics should predict behavior in the common interest test games better than the competing interests metaheuristics. Similarly, the competing interest heuristics should predict behavior in competing interest test games better than the common interests heuristics. This should hold for both the fitted and the optimized heuristics.*

This hypothesis can also be formulated as that the following four inequalities should hold:

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

where we have suppressed the notation for C_{fit} and C_{opt} for clarity.

In order to facilitate comparisons between treatments and between games, we consider the relative prediction loss with respect to the theoretical minimum. Let y be the observed empirical distribution of play in some game G. Then the lowest possible NLL in that game is $NLL(y, y)$.[6] We therefore transform the prediction loss so that the relative prediction loss for model m on game G is thus given by

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

The confidence intervals of the relative prediction loss are then computed over all the games in the test set. Since we consider each game separately for the two different roles, this is 320 observations per test set.

Figure 4 shows the relative prediction loss of the held out test data in each treatment according to heuristic models fit to each treatment. We clearly see that each model predict the test games of the treatment corresponding to treatment it was trained on better than the other model, thus confirming our hypothesis.

An even more striking result is that the optimized metaheuristics achieve nearly the same predictive performance as the fitted metaheuristics. That is, a model with one set of cognitive cost parameters that applies for both treatments (with the heuristic parameters set to optimize the resultant payoff-cost tradeoff) explains participant data almost as well as the fully-parameterized model, in which the heuristic parameters are fit directly and separately to behavior in each treatment.

[6]Note that since only have 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. So the theoretical minimum is truly theoretical.

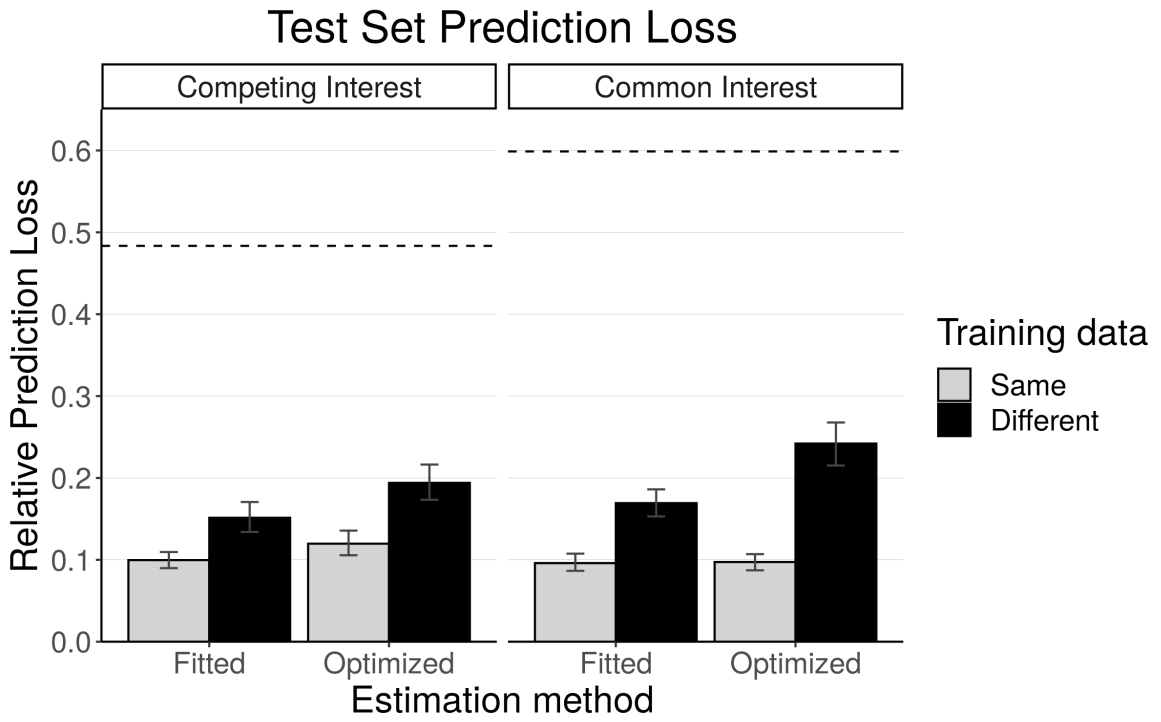


Figure 3: Model performance. Each panel shows the relative prediction loss of the held-out test data for one treatment (competing interest vs common interest). Models are fitted or optimized to either the competing interest training games or the common interest training games. The error bars show 95% confidence interval. The dotted line corresponds to uniformed random guess, which assigns each action the same probability in each game.

Not only do we confirm our hypothesis and show that the rational heuristic is a strong predictor. We also see that we capture most of the distance between the uniformed guess and the theoretical minimum.

Our final model-based hypothesis provides an additional test that the metaheuristics 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 than should the heuristics estimated for the other treatment.*

The logic for 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 participant behavior in each treatment achieve higher payoffs in that treatment. Similarly to the testing for hypothesis 3, we consider the regret, or the relative payoff loss. The regret is the difference between the maximum expected payoff in each game, and the expected payoff given the predicted behavior.

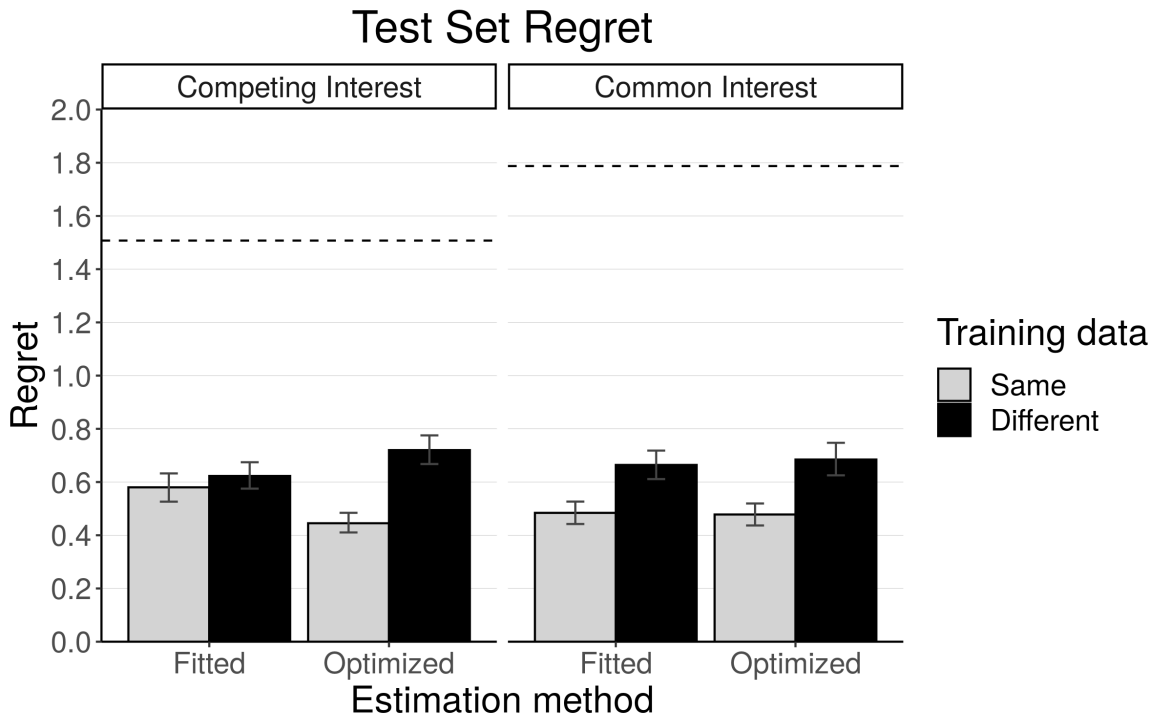


Figure 4: Testing of hypothesis 4 using the metaheuristics. Each panel shows the relative prediction loss of the held-out test data for one treatment (competing vs common interest). Models are fitted or optimized to either the competing interest training games or the common interest training games. The error bars show a 95% confidence interval.

In appendix B we present results from pairwise tests of both hypothesis 3 and 4. We see there that all the differences in both relative prediction loss and regret are significant at at least the 0.01 level. [7]

4.3.3 Estimated Metaheuristics

Inspecting the heuristics' parameters, we find that the model captures the difference in what behavior is adaptive in each treatment in an intuitive manner. Focusing first on the fitted metaheuristics (Table 5), we highlight three interesting differences between the treatments. First, jointmax is used more than twice as often in the common interest condition; this aligns with the intuition that looking for an outcome that is good for both players is a better strategy when interests are usually aligned. Second, the optimism parameter of the row heuristic, γ, is lower in the competing interest treatment; this makes sense because the better outcomes for you are less likely to be

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

realized when your partner has opposing interests. Third, simulation is used more than twice as often in the competing interest treatment; this suggests that careful reasoning about other agents is more important in competitive environments—or conversely, simple heuristics like jointmax are more successful in cooperative environments.

Fitted Metaheuristics

		Jointmax	Row Heuristic	Sim
		φ	γ, φ	$(\gamma, \varphi), \varphi$
Common interests	Params	2.47	-0.26, 1.79	(-0.26, 0.05), 0.97
	Share	35 %	48 %	17 %
Competing interests	Params	1.47	-1.34, 2.66	(1.45, 0.51), 0.96
	Share	15 %	47 %	37 %

Table 5: The estimated meta heuristics for the two treatments.

Optimal Metaheuristics

		Jointmax	Row Heuristic	Sim
		φ	γ, φ	$(\gamma, \varphi), \varphi$
Common interests	Params	1.23	1.2, 1.43	(0.64, 0.53), 1.26
	Share	45 %	53 %	2 %
Competing interests	Params	0.98	-1.52, 1.17	(-1.11, 0.67), 1.13
	Share	0 %	70 %	30 %

Table 6: The optimal meta heuristics for the two treatments.

Table 5 shows the parameters for the optimized heuristics. All three of the patterns we highlight above are replicated, although they are all stronger. This perhaps suggests that participants' adaptation to the experimental environment was only partial, which is not surprising given the relative weight of 30 games in an experiment versus years of experience interacting with people in the real world.

5 Deep Heuristics

A drawback of using the explicitly formulated heuristics, that we use to build up the metaheuristics, is that it is dependent on the decisions made by the researchers. The space of possible heuristics is extremely large, and the ultimately somewhat arbitrary decisions the researcher makes about which heuristics to include might affect the ultimate results. To minimize the risk of our conclusions being driven by such decisions, we also take a very different route to modeling heuristics. While not lending itself to as clear interpretability as the metaheuristics, it includes a much larger set of possible heuristics.

To get a general model of possible heuristics, we use a neural network architecture close to the one developed in Hartford, Wright and Leyton-Brown (2016), with some adjustments to improve interpretability and easier modeling of cognitive costs. The idea is to let every element of the input and hidden layers be a matrix, instead of a single value, of the same size as the game. 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 take into account the thus formed beliefs about the column player's behavior in the action response layers. The different action response layers are then mixed into a response. A heuristic that does not explicitly form beliefs about the other player's behavior would let $AR^R(0)$ be the output, a person who only applies a heuristic two estimate the opponent's behavior and then best responds to it would only use $AR^R(1)$, etc. A figure of the neural network architecture can bee sen in Figure 5.

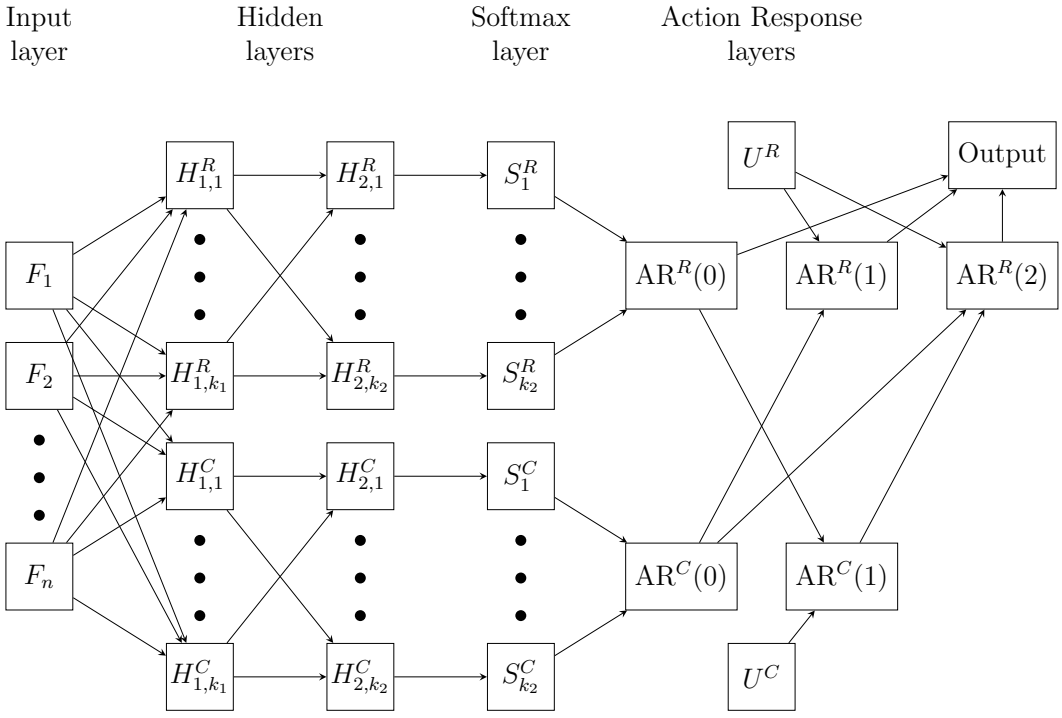


Figure 5: Architecture of the deep heuristic.

5.1 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 $H^R_{0,i} = H^C_{0,i} = F_i$, so the two disjoint parts start with the same feature matrices, but then have 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 U^R and U^C, respectively. More specifically, we 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 follow three examples of such feature matrices.

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

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

Figure 6: Examples of input feature matrices.

5.2 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. So for example, the distribution $S^R_1 \in \Delta^{m_R}$ is give 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 is taken over the columns, so

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

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, 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 for the play of the column player are formed, the $\mathrm{AR}^R(1)$ calculates the expected value from each action, conditioned on that expected play, and takes a softmax over those payoffs.

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

Higher-level action response layers form its beliefs about the other player by taking a weighted average of the lower layers and respond in the same fashion.

5.3 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.

5.4 Cognitive costs

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

5.5 Results

By applying the same estimation method to the deep heuristics as we did to the meta-heuristics, we can test if hypotheses 3 and 4 also hold for a completely different specification of the space of heuristics and cognitive costs. In Figure 7, we see that Hypothesis 3 holds for this specification as well. We also see that the predictive performance of the optimal heuristic is close to the fitted heuristic, given optimized cognitive costs.

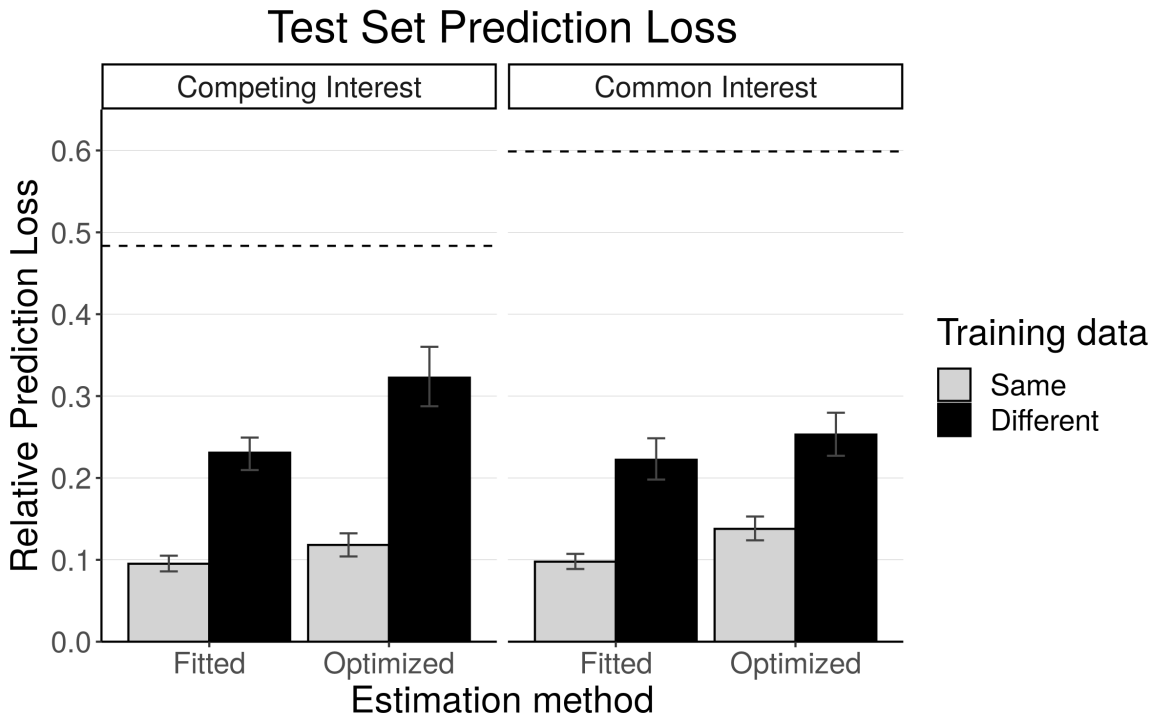


Figure 7: Testing of hypothesis 3 for the deep heuristics. Each panel shows the relative prediction loss of the held-out test data for one treatment (competing interest vs common interest). Models are fitted or optimized to either the competing interest training games or the common interest training games. The error bars show 95% confidence interval. The dotted line corresponds to uniformed random guess, which assigns each action the same probability in each game.

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 populations in the two different treatments, and see that this also holds for the deep heuristics.

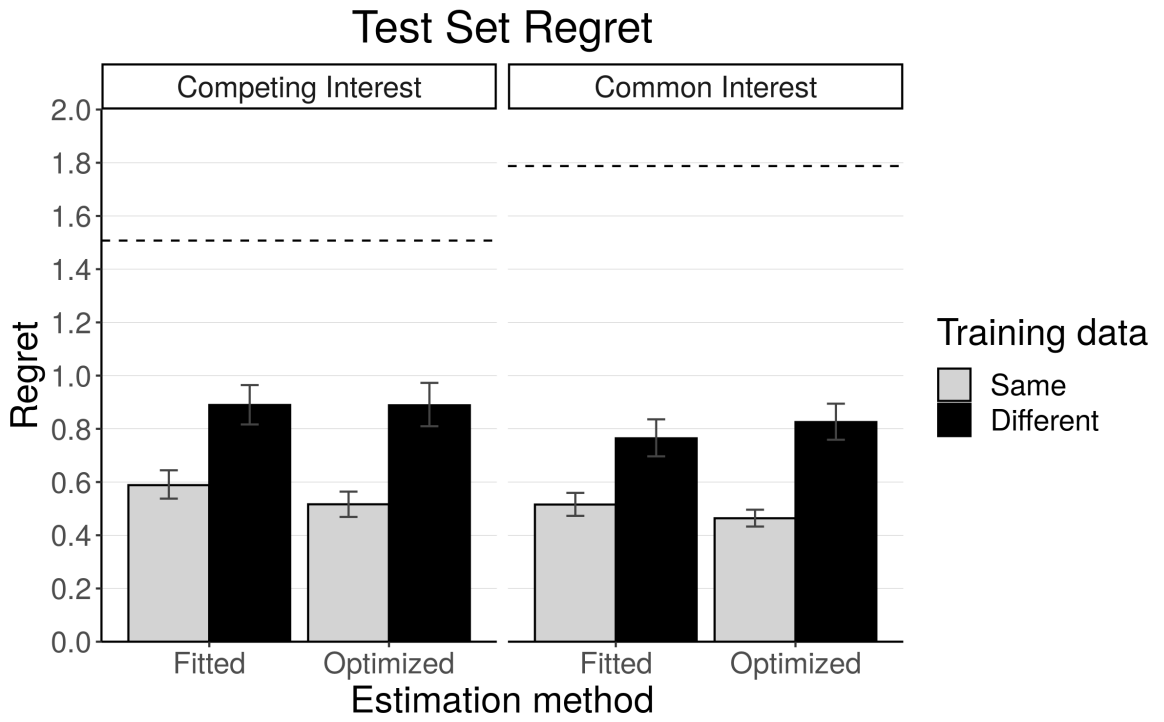


Figure 8: Testing of hypothesis 4 using the deep heuristics. Each panel shows the relative prediction loss of the held-out test data for one treatment (competing vs common interest). Models are fitted or optimized to either the competing interest training games or the common interest training games. The error bars show a 95% confidence interval.

6 Alternative Models

In the previous sections we have seen that rational use of heuristics can explain and predict behavior in one-shot games, and that this is a result we can reproduce with two very different specifications of the space of heuristics. To further show the appropriateness of the chosen spaces of heuristics, and the strength of the predictions, we consider two alternative models of behavior: quantal cognitive hierarchy and prosocial preferences.

Quantal Cognitive Hierarchy. In previous comparisons, variations of cognitive hierarchy models are usually the best performing, Camerer, Ho and Chong (2004); Wright and Leyton-Brown (2017). In such a model, we consider agents of different cognitive levels. In the quantal cognitive hierarchy model we consider here, a level-0 agent plays the uniformly random strategy, playing each action with an equal probability. Level-1 plays a quantal best response to a level-0, and a level-2 player best responds to a combination of level-0 and level-1. In total this model has 4 parameters, the share of level-0 and level-1 players (and thus also the share of level-2), the sensitivity λ_1 of

level-1 players and the sensitivity λ_2 of level-2 players.

Prosocial preferences. In the comparison games where we see that the difference in behavior could potentially be explained by simply having correct and differing beliefs and potentially some kind of pro-social preferences. Since the populations behavior differ, one possible explanation to observed differences in behavior might simply be that the two populations "coordinate" on different norms without necessarily relying explicitly on optimal heuristics. We therefore test a model with noisy best reply to the correct beliefs with a prosocial utility function. The prosocial utility function we consider is

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

where s indicates if $\pi_i(s_i, s_{-i}) < \pi_{-i}(s_i, s_{-i})$ and r indicates if $\pi_i(s_i, s_{-i}) > \pi_{-i}(s_i, s_{-i})$. In other words α determines how much player i values the payoff of player $-i$ when i gains less, and β how much player i values the payoff of player $-i$ when i gain more.

In Figure 6 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 only estimated by fitting the parameters to match behavior, we also include the optimized versions of our two specifications.

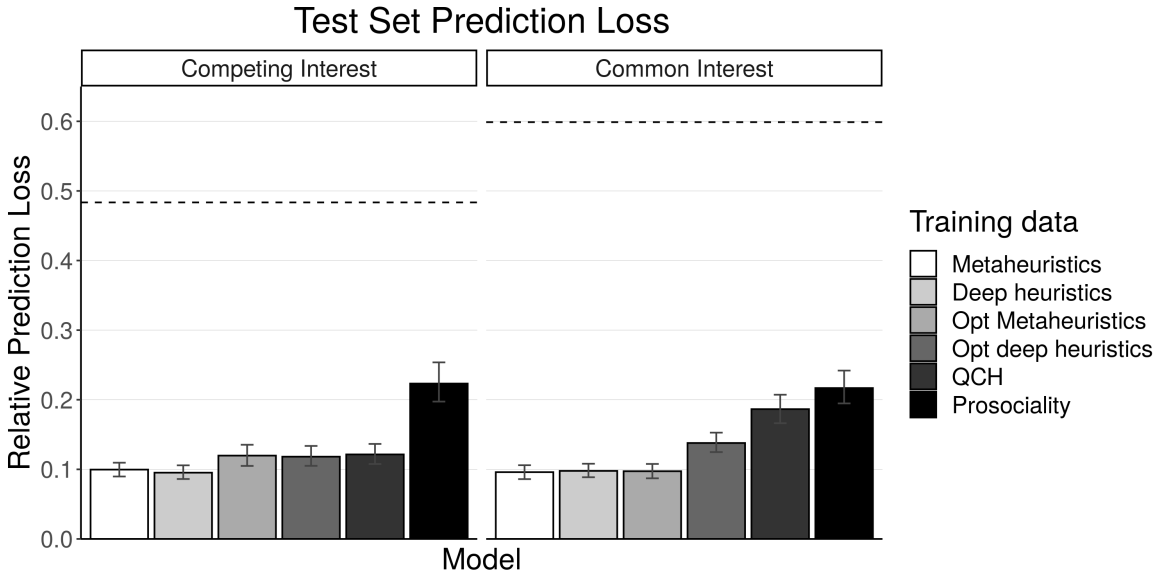


Figure 9: Out of sample relative prediction loss for alternative models of behavior. All the models are estimated on the training games of the same environment as the test games. The error bars show a 95% confidence interval.

For the common interest games, it is clear that both the fitted and optimized versions of our models outperform both the quantal cognitive hierarchy model (QCH) and noisy best response with prosocial preferences (Prosociality). On the competing interest

games, the model with prosociality is still clearly performing worst, but the QCH model is closer in performance to our models. The fitted versions of our models are still better, but the QCH is on par with the optimized versions of our models. This is perhaps not surprising, since the estimated metaheuristic in this treatment is closer to a QCH model than the estimated metaheuristic in the common interest treatment. Taken together, it is clear that our proposed models are better at predicting behavior than alternative models. Furthermore, the rational use of heuristics is better at predicting behavior than the current best performing model from the literature (QCH).

We also see clearly in figure that the predictive performance of our metaheuristics and deep heuristics is very close, even though the deep heuristics encapsulates a much larger space of heuristics. This suggests that we manage to capture the relevant heuristics with our specification of the metaheuristics.

7 Discussion

In the theory presented we combine two perspectives. On the one hand we assume that people use simple cognitive process, working directly on the level of the payoff information, to choose actions which are often inconsistent with rational behavior in any give 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 specific heuristics people use are chosen according to a resource-rational analysis such that they strike an optimal balance between expected payoffs and cognitive costs. We have seen that by combining these two perspectives, we can get better predictions of behavior and understand the influence of the larger environment on the behavior in a given game.

This approach gives a lens through which we can understand the connection between classical rational models of behavior and more behavioral approaches. More generally, it can help us predict the when we should expect behavior to coincide with rational behavior and when we might see systematic deviations from a rational benchmark. If the considered decision situation is such that there exists a simple heuristic which coincides with the optimal behavior, and this heuristic is good in the general environment, we can expect behavior to appear rational. If there is no simple heuristic that leads to optimal behavior in an environment, or if the situation differs from the general environment in such a way that the usually optimal heuristic is performing poorly, we should expect consistent deviations from the rational benchmark. The comparison games illustrate an important consequence of the optimal use of heuristics. The optimal heuristic will focus on the features of the games that are often of importance, but miss opportunities that are rare. So a person used to common interest games might miss the beneficial deviations from the efficient outcome, while persons used to a competing interest games fail to see and take advantage of a common interest.

Our findings relate to those of Peysakhovich and Rand (2016), where varying the sus-

tainability of cooperation in an initial session of repeated prisoner's dilemma affected how much pro-social behavior and trust was shown in later games, including 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 pro-social heuristics (such as jointmax) perform well led them to choose pro-social actions in the comparison games, in some cases, even selecting dominated options. In contrast, putting people in an environment where pro-social actions often result in low payoffs prevented people from achieving efficient outcomes, even when they were Nash Equilibria. Consistent with our theory, the authors interpreted their findings as the result of heuristic decision making. We build on this intuitively appealing notion by specifying a formal model of heuristics in one-shot games that makes quantitative predictions. We also emphasize the influence of cognitive costs (in addition to payoffs) on the heuristics people use.

Lastly, a point relating to the larger literature. In our theory, the generalization between games happens on the level of reasoning; the individuals are not learning which actions are good, but rather how they should reason when choosing an action. This contrasts with theories where the generalization happens on the level of actions, as in Jehiel (2005) or Grimm and Mengel (2012). Furthermore, since our games are randomly generated, no such action learning should take place or alter behavior systematically in our experiment.

8 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 an optimal tradeoff between payoffs and cognitive cost.

In a large preregistered experiment, we confirmed one of the primary qualitative predictions of the theory: people learn what 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 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 noisy best response. Strikingly, we found that a resource-rational model, in which behavior in both treatments is predicted using a single set of fitted cost parameters (with the heuristic parameters set to optimize the resultant payoff-cost tradeoff), achieved nearly the same accuracy as the fully-parameterized model, in which the heuristic parameters are fit directly and separately to 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 payoff and cognitive costs. In a followup analysis, we 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 wider perspective, our model speaks to a decades-long debate on the rationality of human decision making. With classical models based on optimization and utility maximization failing 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 that this approach will be valuable in working towards a more unified understanding of economic decision making.

References

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

Bardsley, Nicholas, Judith Mehta, Chris Starmer, and Robert Sugden. 2010. "Explaining focal points: Cognitive hierarchy theory versus team reasoning." *Economic Journal*, 120(543): 40–79.

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

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

Caplin, Andrew, and Mark Dean. 2013. "Behavioral Implications of Rational Inattention with Shannon Entropy." *NBER Working Paper*, , (August): 1–40.

Chen, Daniel L., Martin Schonger, and Chris Wickens. 2016. "oTree—An Open-Source Platform for Laboratory, Online, and Field Experiments." *Journal of Behavioral and Experimental Finance*, 9: 88–97.

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

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

Devetag, Giovanna, Sibilla Di Guida, and Luca Polonio. 2016. "An eye-tracking study of feature-based choice in one-shot games." *Experimental Economics*, 19(1): 177–201.

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

Fudenberg, Drew, and Annie Liang. 2019. "Predicting and Understanding Initial Play." *American Economic Review*, 109(12): 4112–4141.

Fudenberg, Drew, Fudenberg Drew, David K Levine, and David K Levine. 1998. *The theory of learning in games.* Vol. 2, MIT press.

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, Gerd, and Peter M Todd. 1999. *Simple Heuristics That Make Us Smart.* Oxford University Press, USA.

Goeree, Jacob K., and Charles A. Holt. 2004. "A model of noisy introspection." *Games and Economic Behavior*, 46(2): 365–382.

Goldstein, Daniel G., and Gerd Gigerenzer. 2002. "Models of Ecological Rationality: The Recognition Heuristic." *Psychological review*, 109(1): 75.

Griffiths, Thomas L, Falk Lieder, and Noah 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, Veronika, and Friederike Mengel. 2012. "An experiment on learning in a multiple games environment." *Journal of Economic Theory*, 147(6): 2220–2259.

Hartford, Jason S., James R. Wright, and Kevin Leyton-Brown. 2016. "Deep Learning for Predicting Human Strategic Behavior." *Advances in Neural Information Processing Systems*, , (Nips): 2424–2432.

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

Hebert, Benjamin, and Michael Woodford. 2019. "Rational Inattention When Decisions Take Time." *Journal of Chemical Information and Modeling*, 53(9): 1689–1699.

Imai, Taisuke, Tom A Rutter, and Colin F Camerer. 2020. "Meta-Analysis of Present-Bias Estimation Using Convex Time Budgets*." *The Economic Journal*, 186(2): 227–236.

Izard, Véronique, and Stanislas Dehaene. 2008. "Calibrating the Mental Number Line." *Cognition*, 106(3): 1221–1247.

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

Kahneman, Daniel. 2011. *Thinking, fast and slow.* Macmillan.

Lewis, Richard L., Andrew Howes, and Satinder Singh. 2014. "Computational Rationality: Linking Mechanism and Behavior through Bounded Utility Maximization." *Topics in Cognitive Science*, 6(2): 279–311.

Lieder, Falk, and Thomas 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, Falk, and Thomas L. Griffiths. 2017. "Strategy Selection as Rational Metareasoning." *Psychological Review*, 124(6): 762–794.

Lieder, Falk, and Thomas L. Griffiths. 2019. "Resource-Rational Analysis: Understanding Human Cognition as the Optimal Use of Limited Computational Resources." *Behavioral and Brain Sciences*.

Lieder, Falk, Paul M Krueger, and Tom Griffiths. 2017. "An automatic method for discovering rational heuristics for risky choice."

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

Mengel, Friederike, and Emanuela Sciubba. 2014. "Extrapolation and structural similarity in games." *Economics Letters*, 125(3): 381–385.

Misyak, Jennifer B., and Nick Chater. 2014. "Virtual Bargaining: A Theory of Social Decision-Making." *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369(1655).

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

Peysakhovich, Alexander, and David G. Rand. 2016. "Habits of virtue: Creating norms of cooperation and defection in the laboratory." *Management Science*, 62(3).

Polonio, Luca, Sibilla Di Guida, and Giorgio Coricelli. 2015. "Strategic sophistication and attention in games: An eye-tracking study." *Games and Economic Behavior*, 94: 80–96.

Savage, Leonard J. 1954. *The Foundations of Statistics. The Foundations of Statistics*, Oxford, England:John Wiley & Sons.

Simon, Herbert A. 1976. "From substantive to procedural rationality." *25 Years of Economic Theory*, 65–86.

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

Smith, Vernon L. 2003. "Constructivist and Ecological Rationality in Economics." *American economic review*, 93(3): 465–508.

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

Stahl, Dale O., and Paul W. Wilson. 1995. "On players' models of other players: Theory and experimental evidence."

Steiner, Jakub, Colin Stewart, and Filip Matějka. 2017. "Rational Inattention Dynamics: Inertia and Delay in Decision-Making." *Econometrica*, 85(2): 521–553.

Stewart, Neil, Simon Gächter, Takao Noguchi, and Timothy L Mullett. 2016. "Eye Movements in Strategic Choice." 156(October 2015): 137–156.

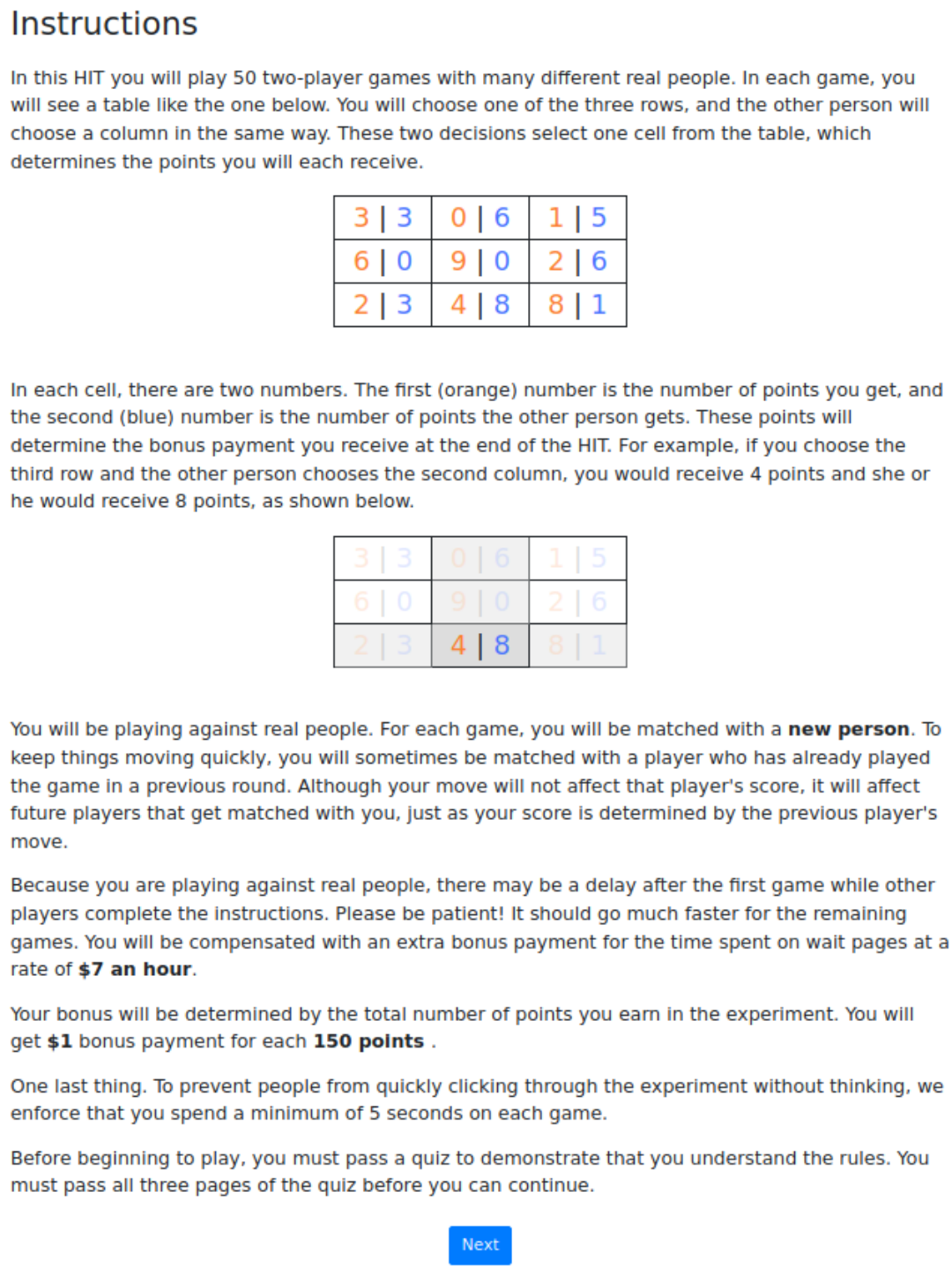
Sugden, Robert. 2003. "The logic of team reasoning." *Philosophical explorations*, 6(3): 165–181.

Todd, Peter M., and Gerd Ed Gigerenzer. 2012. *Ecological Rationality: Intelligence in the World.* Oxford University Press.

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

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

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

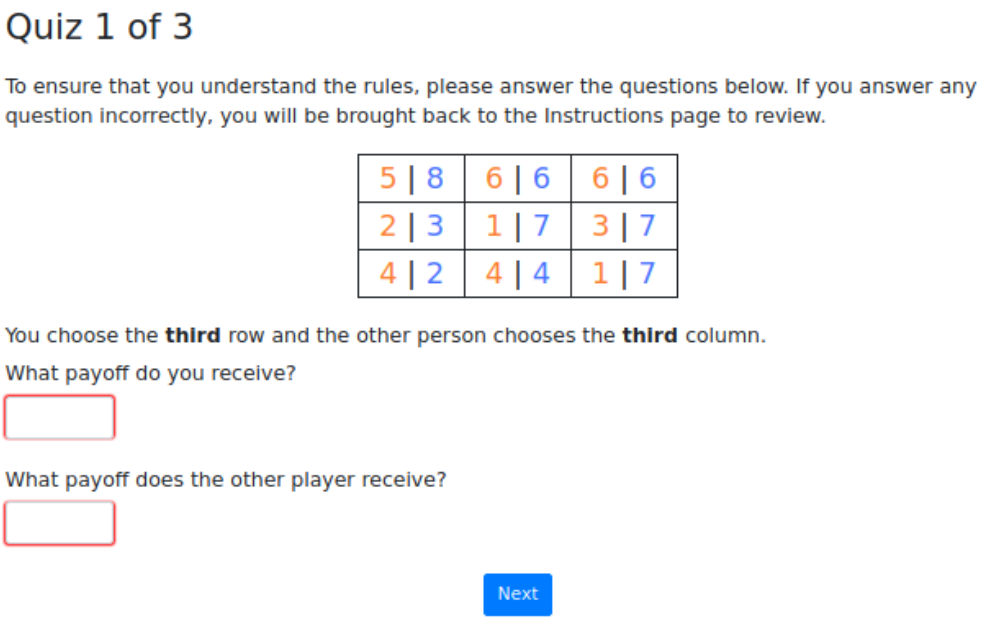


Figure 11: The participants have to complete three questions like this in a row in order to be allowed to participate in the experiment.

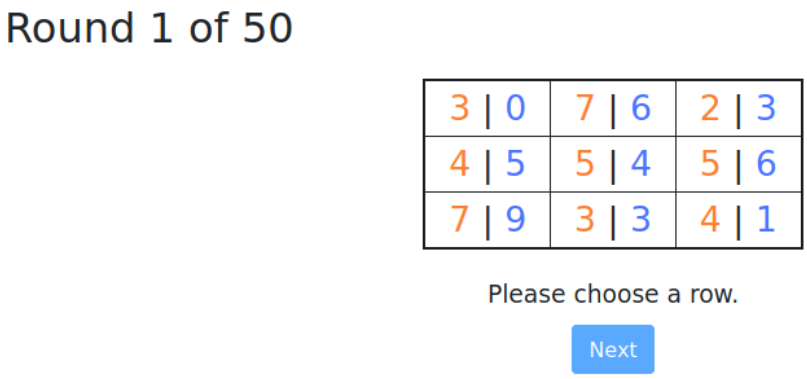


Figure 12: In each round, the participant chose a row by clicking on it. Once it is clicked it is highlighted and they have to click the next button to proceed.

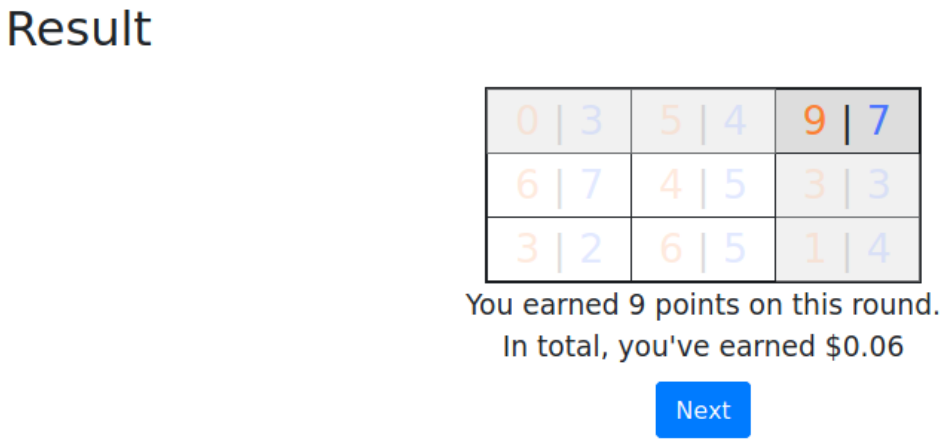


Figure 13: Once the behavior of the matched participant, either by her making a decision or by sampling from previous decisions in the game from the same population, the result is shown.

B 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 prediction loss or payoff between the relevant models. For each game we get two observations, one for each role. For each of these comparisons we perform both a t-test and a non-parametric, Wilcox rank test. As can bee seen in the tables below, all of theses tests are significant.

Model	Test set	Estimation	Difference	T-test	Wilcox
Metaheuristics	Competing Interest	Fitted	-0.052	$p < 0.001$	$p < 0.001$
Metaheuristics	Competing Interest	Optimized	-0.074	$p < 0.001$	$p < 0.001$
Metaheuristics	Common Interest	Fitted	-0.073	$p < 0.001$	$p < 0.001$
Metaheuristics	Common Interest	Optimized	-0.145	$p < 0.001$	$p < 0.001$
Deep heuristics	Competing Interest	Fitted	-0.136	$p < 0.001$	$p < 0.001$
Deep heuristics	Competing Interest	Optimized	-0.204	$p < 0.001$	$p < 0.001$
Deep heuristics	Common Interest	Fitted	-0.124	$p < 0.001$	$p < 0.001$
Deep heuristics	Common Interest	Optimized	-0.115	$p < 0.001$	$p < 0.001$

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

Model	Test set	Estimation	Difference	T-test	Wilcox
Metaheuristics	Competing Interest	Fitted	-0.043	$p = 0.002$	$p < 0.001$
Metaheuristics	Competing Interest	Optimized	-0.275	$p < 0.001$	$p < 0.001$
Metaheuristics	Common Interest	Fitted	-0.180	$p < 0.001$	$p < 0.001$
Metaheuristics	Common Interest	Optimized	-0.207	$p < 0.001$	$p < 0.001$
Deep heuristics	Competing Interest	Fitted	-0.301	$p < 0.001$	$p < 0.001$
Deep heuristics	Competing Interest	Optimized	-0.372	$p < 0.001$	$p < 0.001$
Deep heuristics	Common Interest	Fitted	-0.249	$p < 0.001$	$p < 0.001$
Deep heuristics	Common Interest	Optimized	-0.362	$p < 0.001$	$p < 0.001$

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