Rational Heuristics for One-Shot Games[*]

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

March 29, 2022

Abstract

We present a theory of human behavior in one-shot interactions based on the assumption that people use heuristics that optimally trade off expected payoff and cognitive costs. The theory predicts that people's behavior will depend on their past experience; specifically, they will make choices using heuristics that would have 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, outperforming 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 a grant to TLG by the Templeton foundation, as well as grants from 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 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 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, so that there is an opportunity for learning, both theoretical and experimental work suggests that Nash equilibria can often be sensible long-run predictions (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 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. 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 (Wright and Leyton-Brown, 2017), behavior depends on aspects of the game that these models are indifferent to (Bardsley et al., 2010; Heap, Arjona and Sugden, 2014), and there is evidence of earlier games played having an effect on behavior, something that any static model will 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 construct 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 they might not choose the best action in a given game, they will learn which heuristics generally work well (c.f. *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 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, 2020; Griffiths, Lieder and

Goodman, 2015; c.f. Howes, Lewis and Vera, 2009; Lewis, Howes and Singh, 2014; Gershman, Horvitz and Tenenbaum, 2015).

It is instructive to compare resource-rational analysis with two other approaches to explaining observed deviations from perfectly rational behavior. Like *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), resource-rational models assume that the costs and benefits of information processing are optimally traded off. However, while rational inattention models typically assume domain-general cost functions (e.g., based on entropy reduction), resource-rational models typically make stronger assumptions about the specific computational processes and costs that are likely to be involved in a given domain. In this way, resource-rationality is more similar *ecological rationality*, a framework based on the idea that people use computationally frugal heuristics, which are highly effective for the kinds of problem 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-rational models. This makes it possible to predict when people will use one heuristic versus another (Lieder and Griffiths, 2017) and even to automatically discover novel heuristics (Lieder, Krueger and Griffiths, 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 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 have common interests, it might be better to look for outcomes that would be good for everyone (c.f. 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-interest* environment, there is a

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

positive correlation between the payoffs of the two players over the set of strategy profiles—outcomes that are good for one player tend to be good for the other as well. In the *competing-interest* environment, the payoff correlation is negative—one player's loss is the other's gain, 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 (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 provide further support for the claim that participant behavior is consistent with an optimal tradeoff between payoff and cognitive cost, we define a parameterized family 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 cost/benefit tradeoff, and ask how well they predict 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. 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 version of this model predicted the modal action with an accuracy of 88%, compared to 80% for a quantal cognitive hierarchy model.

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 a parameterized family of heuristics and associated cognitive costs, which we call *metaheuristics.* In Section 4, we present our experimental results, confirming that the two different environments indeed lead to large differences in behavior. Furthermore, we show that the differences in behavior can be accurately predicted out-of-sample by assuming that the participants use the optimal metaheuristics for the respective environments. We also reconfirm the predictions of the general model using a nonparametric representation of the possible heuristics. Lastly, in Section 5, we compare our model to alternative models, including quantal cognitive hierarchy and prosocial preferences, finding that our model predicts behavior better than 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, $c(h) \in \mathbb{R}_+$.[2] 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

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

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 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 opponent heuristic combinations a player is likely to face.

[2] 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 3×3, we drop that dependency.

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

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

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 (1), i.e.,

$$h^* = \operatorname*{argmax}_{h \in \mathcal{H}} V(h, \mathcal{E}), \tag{2}$$

or in slightly expanded form

$$h^* = \operatorname*{argmax}_{h \in \mathcal{H}} \mathbb{E}_{\mathcal{E}}\left[\pi_i\left(h_i(G), h_{-i}(G^T)\right) - c(h_i)\right]. \tag{3}$$

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 trade-off 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 *learn* an optimal heuristic (but not an optimal action) even if the individual has limited experience with each specific game. In Appendix D, we show that a simple learning model can reproduce the performance of the full optimizing model.

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 confirm—that this family is expressive enough to illustrate the general theory's predictions and provide a strong quantitative account 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 the first entry 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 $f : \mathbb{R}^{m_{-i}} \to \mathbb{R}$. 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 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. 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 **1** (the top row) when γ is large and positive, **2** when γ is large and negative, and **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 a human decision maker. 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

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

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

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

where $C_{\text{row}} > 0$ 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 $(\pi_i(s_i, s_{-i}), \pi_{-i}(s_i, s_{-i}))$. 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 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

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

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 1), the jointmax heuristic would assign the highest probability to row $\mathbf{1}$ 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

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 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 best responses is potentially a good decision heuristic. We here describe how we parameterize such heuristics, which we call *simulation heuristics*.

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 players 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 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,

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

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.

3.4 Selection Rule

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 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, hence the term "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

$$\begin{aligned}\mathbb{P}\left[\{\text{use } h^n \text{ in } G\}\right] &= \frac{\exp\left[(a_n + V(h^n, \mathcal{E}, G))/\lambda\right]}{\sum_{j=1}^{N} \exp\left[(a_j + V(h^j, \mathcal{E}, G))/\lambda\right]} \\ &= \frac{p_n \exp\left[V(h^n, \mathcal{E}, G)/\lambda\right]}{\sum_{j=1}^{N} p_j \exp\left[V(h^j, \mathcal{E}, G)/\lambda\right]} \end{aligned} \tag{4}$$

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

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 write down the performance of a metaheuristic in an environment $\mathcal{E}$, analogously to (1) for heuristics, as

$$V^{meta}(m, \mathcal{E}) = \sum_{G \in \mathcal{G}} \sum_{h \in H} V(h^n, \mathcal{E}, G) \cdot \frac{p_n \exp\left[V(h^n, \mathcal{E}, G)/\lambda\right]}{\sum_{j=1}^{N} p_j \exp\left[(V(h^j, \mathcal{E}, G))/\lambda\right]} \cdot P(G) \tag{5}$$

The optimization problem faced by the individual, subject to the adjustment cost λ, is then to maximize (5), 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 large. Therefore, we will assume a small set of heuristics and jointly find optimal parameters of those heuristics and priors P.

3.5 Example and Model Predictions

The main assumption of our theory is captured in Equation 2. Given the set of possible heuristics, $\mathcal{H}$, the environment $\mathcal{E}$, and the cognitive cost function, c, an individual will use the heuristic that maximizes the expected payoff minus cognitive costs in that environment. Here, we demonstrate the consequences of the theory with a simple example.

Consider two possible environments: one consists entirely of coordination games (where the players want to coordinate on the same action), and another consists entirely of constant-sum games (where the players interests are exactly opposed). In both environments, all other players follow the row-mean heuristic (or level-1 in the language of level-k) without noise. 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**, since 8 is larger than 6, so row **1** is the optimal play. In the constant sum game, the column player would choose column **2** since $7 + 6 > 4 + 6$, so 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 highest minimum value for each player (the jointmax cell heuristic). In the constant sum game, the best action is the one that has the highest guaranteed payoff (the maximin row 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 only 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, 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. In 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 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 to select one's own action can sometimes be approximated by simpler heuristic strategies. In some environments, this approximation may be highly accurate; in this case, 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.

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. 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.

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.[3]

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 (c.f. Spiliopoulos and Hertwig, 2020). 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.[4] 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.

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

[4] 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, 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-interest 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-interest treatment.

4.1.1 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 a row that gives 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 with different efficiency. 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 recursive reasoning model would predict (5,5) being played.

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.

The third comparison game is a game with two NE, where one is the pure NE $(\mathbf{3}, \mathbf{3})$, 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.

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**.

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 on which model parameters are estimated and a test set on which predictive performance is evaluated. We used the first 30 games from each session as the training set and the remaining 16 treatment games as the test set. We chose this split so that we would test the predictions on the later games when people would be most likely to be using a consistent decision strategy. We consider each game as two observations, the 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 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 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 letting the opponents 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-interest environment $\mathcal{E}^-$ in the corresponding way. Lastly, we divide the games into 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 p is the observed distribution of play for for some role in some game, and q is the predicted distribution of play from some model, the negative log-likelihood (NLL) is defined

$$\mathrm{NLL}(q, p) = -\sum_{s} p_s \cdot \log(q_s).$$

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

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

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 (4).

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 jointmax cell heuristic, a row heuristic, and a simulation heuristic. The model thus has twelve free parameters: six that specify the behavior of each primitive heuristic, four for the cognitive costs, and two for the selection rule's prior.

The cost parameters are fixed from the decision-maker's perspective, reflecting constraints imposed by their cognitive abilities. We thus fit the cost parameters to data. In contrast, the heuristics' parameters and selection rule prior are under the decision maker's control. We consider two methods for estimating these parameters: 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-interest metaheuristic is given by

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

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-interest metaheuristic, for cognitive costs C, is instead given by

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

The 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-interest environment are defined analogously.

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 that minimize the average NLL of the optimized or fitted heuristics on the training data. So

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

and

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

Notice the crucial difference between the fitted and optimized metaheuristics. For the fitted metaheuristics, we fit both the 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 four pre-registered hypotheses. The first two are model-free and concern the behavior in the comparison games. The next two are model-based and concern the behavior in the treatment games.

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

Table 4: χ^2 tests for each comparison games. All of the them significant at the preregistered 0.05 level.

4.3.1 Model-free analysis of comparison games

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 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 4 confirms that these differences are statistically significant (χ^2-tests, as preregistered).

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

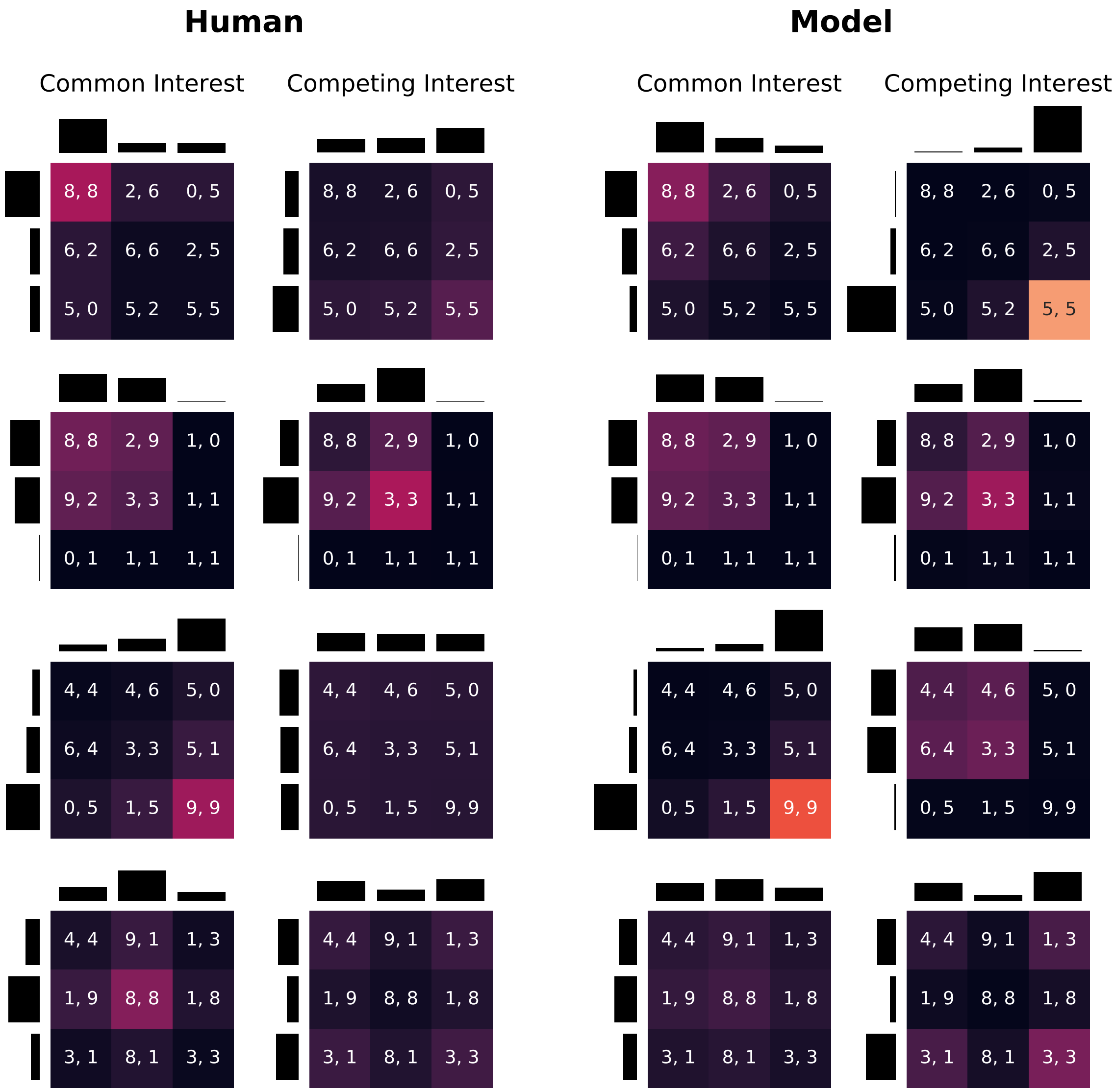


Figure 2: Distribution of play in the four comparison games. Each panel shows the joint and marginal distributions of row/column plays in a single game. The cells are annotated with each player's payoffs for the given outcome. The two columns to the left show the actual behavior in the two environments,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

Treatment average payoff

	Common interest	Competing interest	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 average payoff between the two treatments in the comparison games.

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 5 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 $p < .05$).

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 differs in a way that the model can capture. Concretely, the common-interest 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-interest heuristics. This should hold for both the fitted and the optimized heuristics.*

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 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 $NLL(y, y)$.[5] 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 3 shows the relative prediction loss on the held out test data in each treatment achieved by each possible method of fitting the model. We clearly see that the models which 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 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.

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 uniform guess and the theoretical minimum. Table 6 in the Appendix show accuracy and average NLL for all models we consider in the paper. There, we see that the average accuracy

[5] 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.

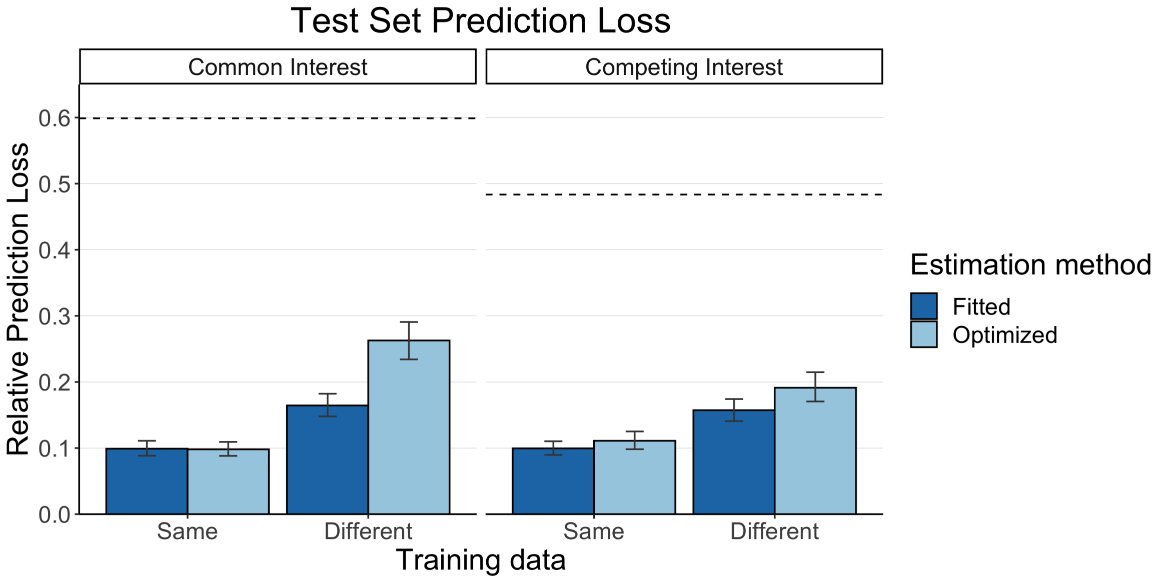


Figure 3: Model predictive performance. Each panel shows the relative prediction loss (average negative log likelihood minus the lowest possible value) of the held-out test data for one treatment (competing interest vs. common interest). Models are fitted or optimized to either the competing-interest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

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. So in the games where the proposed model fails to assign the highest probability to the modal action, play is quite even and therefore difficult to predict.

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. To account for differences in maximal payoff in the different games, we measure model performance

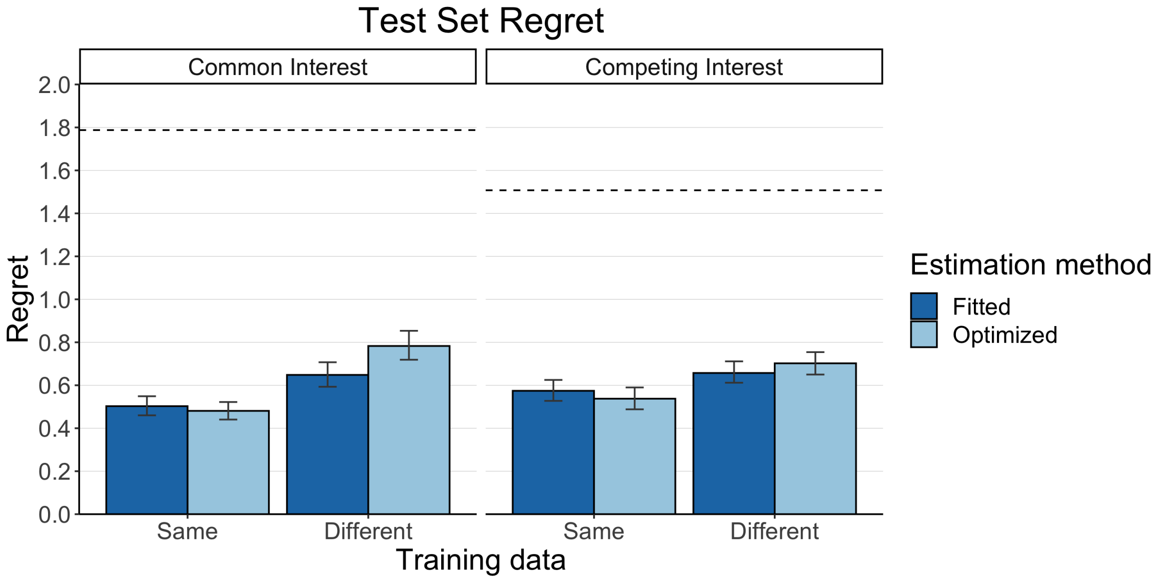


Figure 4: Model payoff performance. Each panel shows the regret (best possible expected payoff minus true expected payoff) attained by models that are trained and tested in different combinations of common-interest and competing-interest environments. The dashed line shows the performance of uniform-random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

in terms of regret, the difference between the maximum expected payoff in each game, and the expected payoff given the predicted behavior.

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

In Appendix C.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 at least the 0.01 level.[6]

4.4 Deep Heuristics

A drawback of using explicitly formulated heuristics, as we have done above, is that the results are dependent on somewhat arbitrary decisions made by the researchers (in particular, the consideration set of heuristics, $\mathcal{H}$). 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 being as interpretable as the

[6] 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.

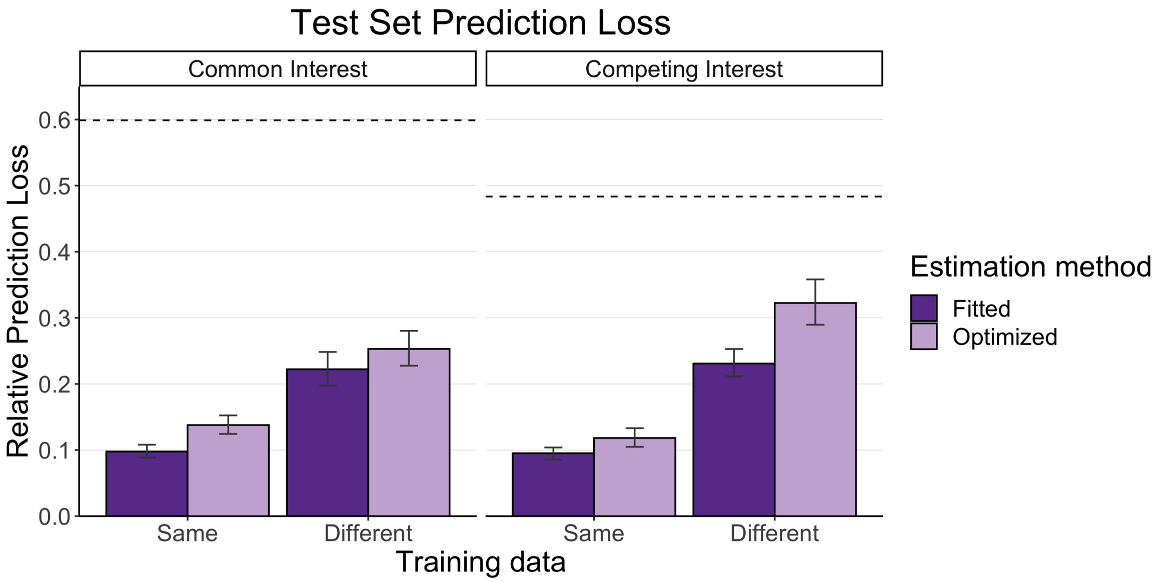


Figure 5: Deep heuristics predictive performance.

metaheuristics, this new class includes a much larger set 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 allow for modeling cognitive costs. This architecture 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, vastly reducing the size of the parameter space (c.f. 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 different cognitive costs to each type of reasoning. A detailed description of the architecture is given in Appendix B.

By applying the same estimation method to the deep heuristics as we did to the metaheuristics, 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 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.

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. As before, we see that the fitted models achieved lower regret

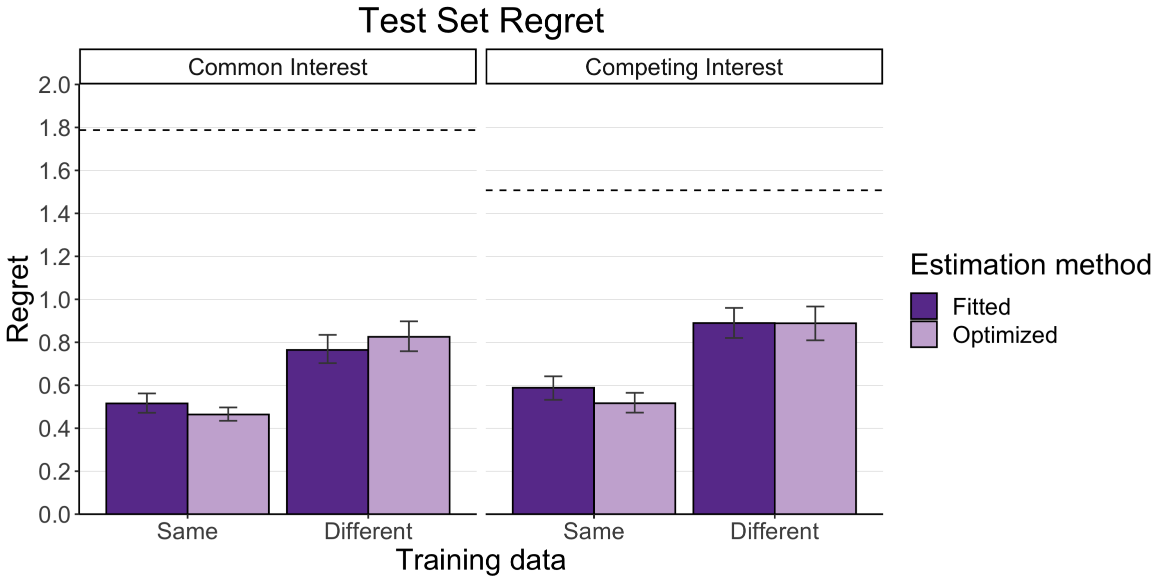


Figure 6: Deep heuristics payoff performance.

in the treatment on which they were trained, again suggesting that the heuristics people use are well-adapted to their environment.

5 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 families of heuristics. To further show the appropriateness of the chosen spaces of heuristics, and the strength of the predictions, we consider three alternative models of behavior: 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, 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. We found that adding higher levels of play did not improve predictive performance.

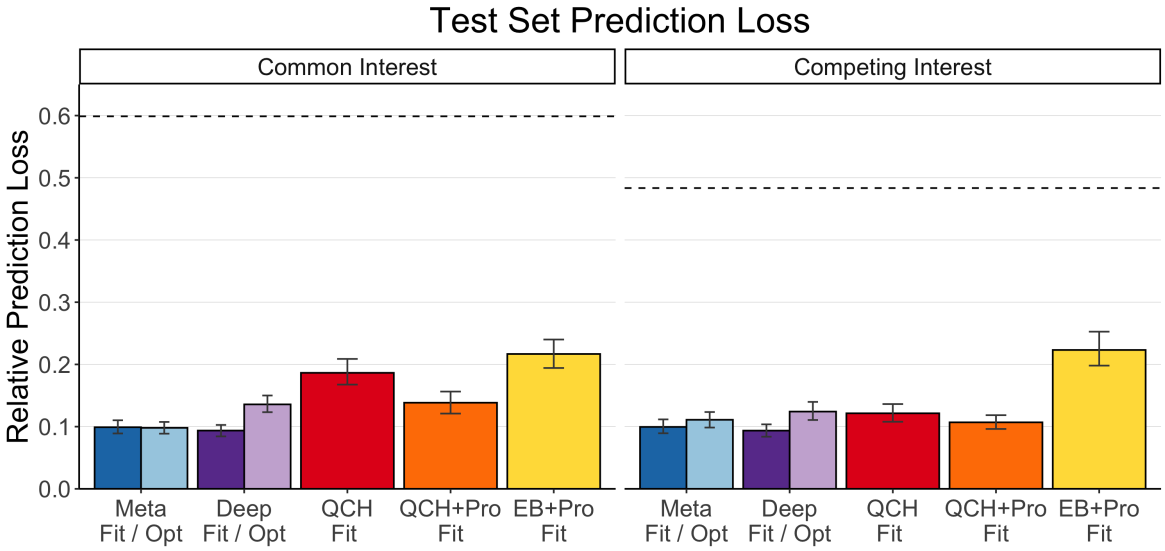


Figure 7: Out-of-sample relative prediction loss for alternative models of behavior. All the models are estimated on the training games of the same environment as the test games. The error bars show a 95% confidence interval. Legend: QCH = quantal cognitive hierarchy, Pro = prosocial preferences, EB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

Prosocial preferences

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

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

$$u_i(s_i, s_{-i}) = (1 - \alpha s - \beta r) \times \pi_i(s_i, s_{-i}) + (\alpha s + \beta r) \times \pi_{-i}(s_i, s_{-i}) \tag{6}$$

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. This thus adds two parameters to the QCH model, α and β. In this model, the beliefs are formed with a standard QCH model, but the payoffs are transformed according to the prosocial preferences (Equation 6) before the last quantal best response step.[7] This model can account for differences in behavior in the two treatments both by assuming different levels of prosociality and also 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 the populations' differ in behavior, participants may form different beliefs about what they expect the other player to do. In particular, participants in the common-interest treatment may expect the other player to cooperate by selecting an action with a jointly beneficial outcome, while participants in the competing-interest 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 only estimated by fitting the parameters to match behavior, we also include the optimized versions of our two specifications.

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), as well as a model with both prosocial preferences and recursive reasoning (Pro+QCH). On the competing-interest games, the model with prosociality and empirical beliefs is still clearly performing worst, but the performance of the QCH and Pro+QCH models are not significantly different from the metaheuristic models (see Appendix C.2 for pairwise tests). This suggests that behavior in the competing-interest environment is closer to that of a QCH type model than in the common-interest 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 fitted metaheuristics and deep heuristics is very close, even though the deep heuristics encapsulates a much larger space of heuristics. This suggests that we managed to capture the relevant

[7] We also considered another model combining QCH and prosocial preferences, in which the player also has some beliefs about the other players prosociality that informs the beliefs formed during the QCH steps. This didn't make a substantial difference in fit, as reported in Table 9.

space of heuristic strategies with our specification of the metaheuristics.

6 Discussion

In the theory presented we combine two perspectives. On the one hand we assume that people use simple cognitive strategies, working directly on the level of the payoff information, to choose actions that 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 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 the 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 decision. Secondly, that heuristic has to perform well in the larger environment so that the decision maker learns to use it. If either there doesn't exist a heuristic that identifies the optimal action, or the heuristic that normally works well in the environment leads to the wrong decision in the specific situation, 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. So, a person used to common-interest games might miss an opportunity for personal gain at another's expense while a person used to competing-interest games might fail to notice an outcome that is actually best for everyone.

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

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 payoffs and 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 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.

Bruhin, Adrian, Ernst Fehr, and Daniel 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., 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.

Dhami, Sanjit. 2016. *The foundations of behavioral economic analysis.* Oxford University Press.

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.

Fehr, Ernst, and Klaus M Schmidt. 1999. "A theory of fairness, competition, and cooperation." *The quarterly journal of economics*, 114(3): 817–868.

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): 75–90.

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.

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

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.

Krueger, Paul, Frederick Callaway, Sayan Gul, Tom Griffiths, and Falk Lieder. 2022. "Discovering Rational Heuristics for Risky Choice."

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. 2020. "Resource-rational analysis: Understanding human cognition as the optimal use of limited computational resources." *Behavioral and Brain Sciences*, 43: e1.

Lieder, Falk, Paul M. Krueger, and Thomas L. 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, 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): 631–647.

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.* John Wiley & Sons.

Simon, Herbert A. 1976. “From substantive to procedural rationality.” In *25 years of economic theory.* 65–86. Springer.

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

Spiliopoulos, Leonidas, and Ralph Hertwig. 2020. “A Map of Ecologically Rational Heuristics for Uncertain Strategic Worlds.” *Psychological Review*, 127(2): 245–280.

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

Stahl, Dale O., and Paul W. Wilson. 1995. “On players’ models of other players: Theory and experimental evidence.”

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.

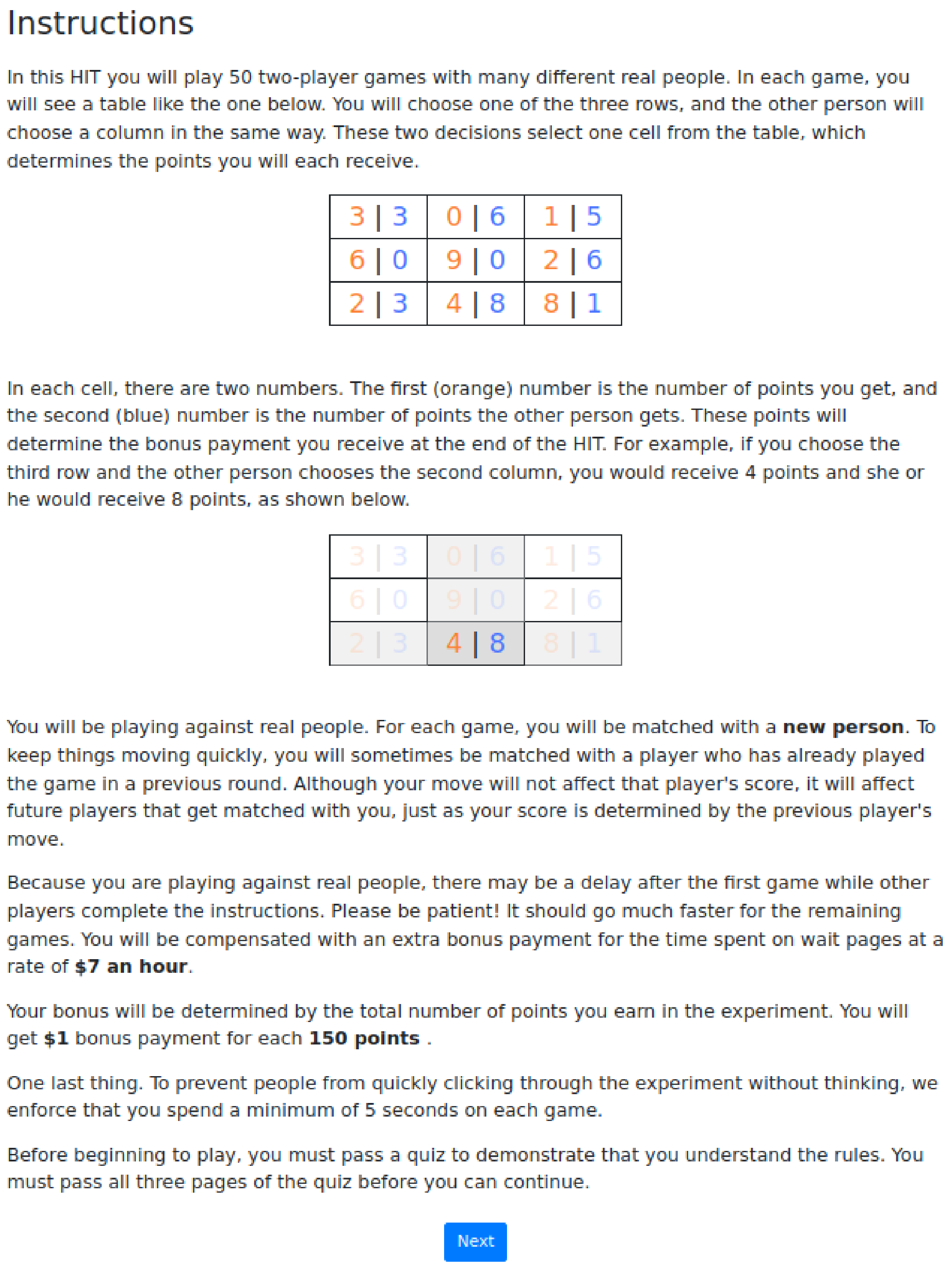
Sugden, Robert. 2003. “The logic of team reasoning.” *Philosophical explorations*, 6(3): 165–181.

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.

<table>
<tr><td>3 | 3</td><td>0 | 6</td><td>1 | 5</td></tr>
<tr><td>6 | 0</td><td>9 | 0</td><td>2 | 6</td></tr>
<tr><td>2 | 3</td><td>4 | 8</td><td>8 | 1</td></tr>
</table>

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.

<table>
<tr><td>3 | 3</td><td>0 | 6</td><td>1 | 5</td></tr>
<tr><td>6 | 0</td><td>9 | 0</td><td>2 | 6</td></tr>
<tr><td>2 | 3</td><td>4 | 8</td><td>8 | 1</td></tr>
</table>

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.

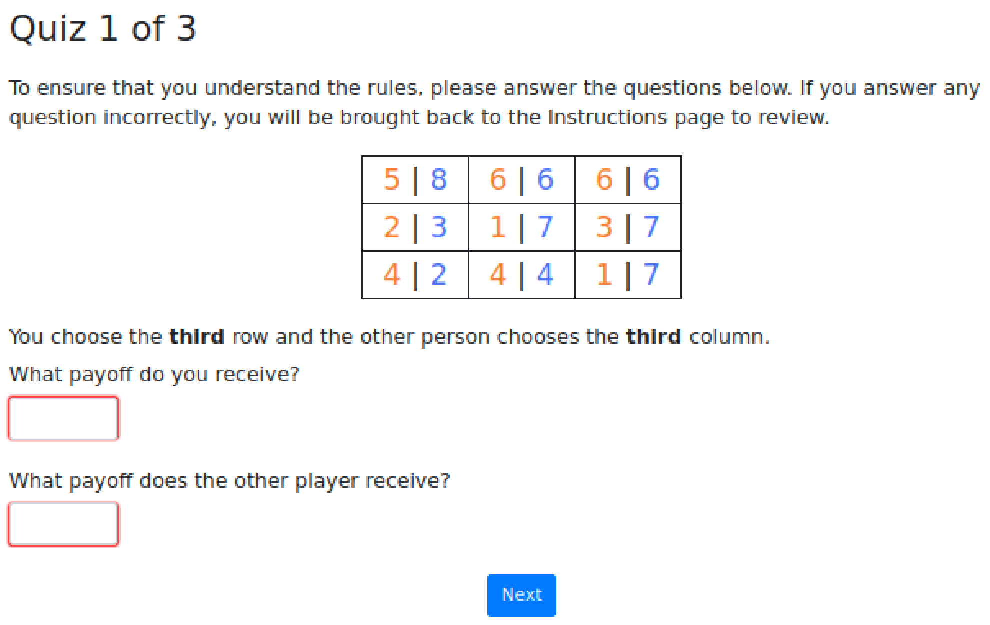
Because you are playing against real people, there may be a delay after the first game while other players complete the instructions. Please be patient! It should go much faster for the remaining games. You will be compensated with an extra bonus payment for the time spent on wait pages at a rate of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

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



Quiz 1 of 3

To ensure that you understand the rules, please answer the questions below. If you answer any question incorrectly, you will be brought back to the Instructions page to review.

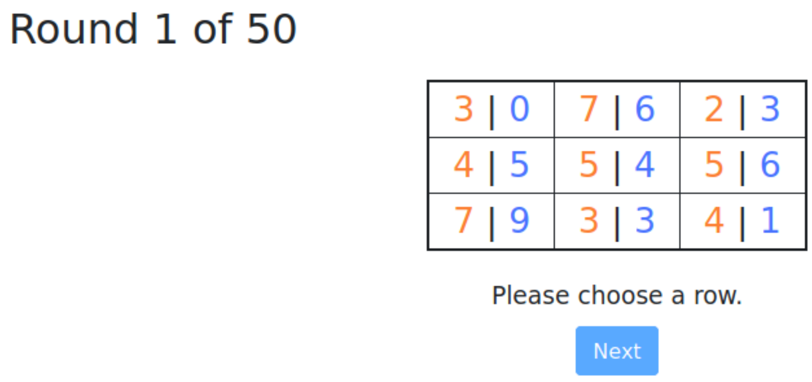
5 | 8	6 | 6	6 | 6
2 | 3	1 | 7	3 | 7
4 | 2	4 | 4	1 | 7

You choose the **third** row and the other person chooses the **third** column.

What payoff do you receive?

What payoff does the other player receive?

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

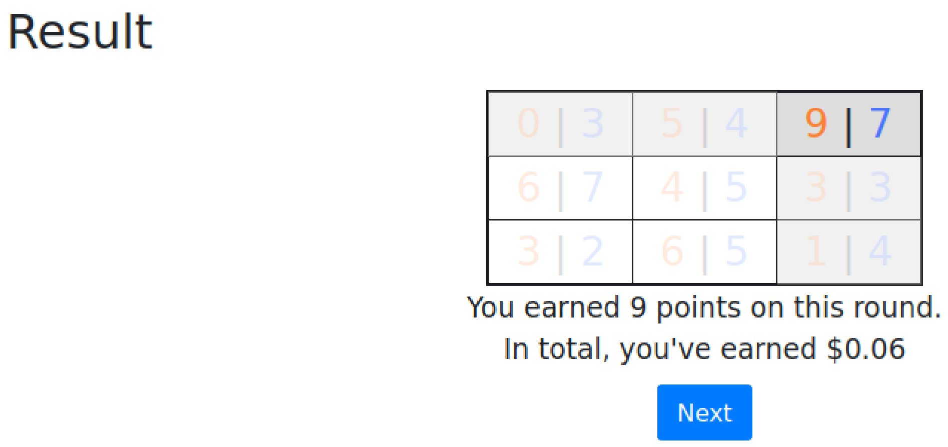


Round 1 of 50

3 | 0	7 | 6	2 | 3
4 | 5	5 | 4	5 | 6
7 | 9	3 | 3	4 | 1

Please choose a row.

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.



Result

0 | 3	5 | 4	9 | 7
6 | 7	4 | 5	3 | 3
3 | 2	6 | 5	1 | 4

You earned 9 points on this round.
In total, you've earned $0.06

Figure 11: 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 Deep Heuristics

Our neural network architecture is based on that developed in Hartford, Wright and Leyton-Brown (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 behavior in separate "action response" layers. The different action response layers are then mixed into a response distribution. 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 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 12.

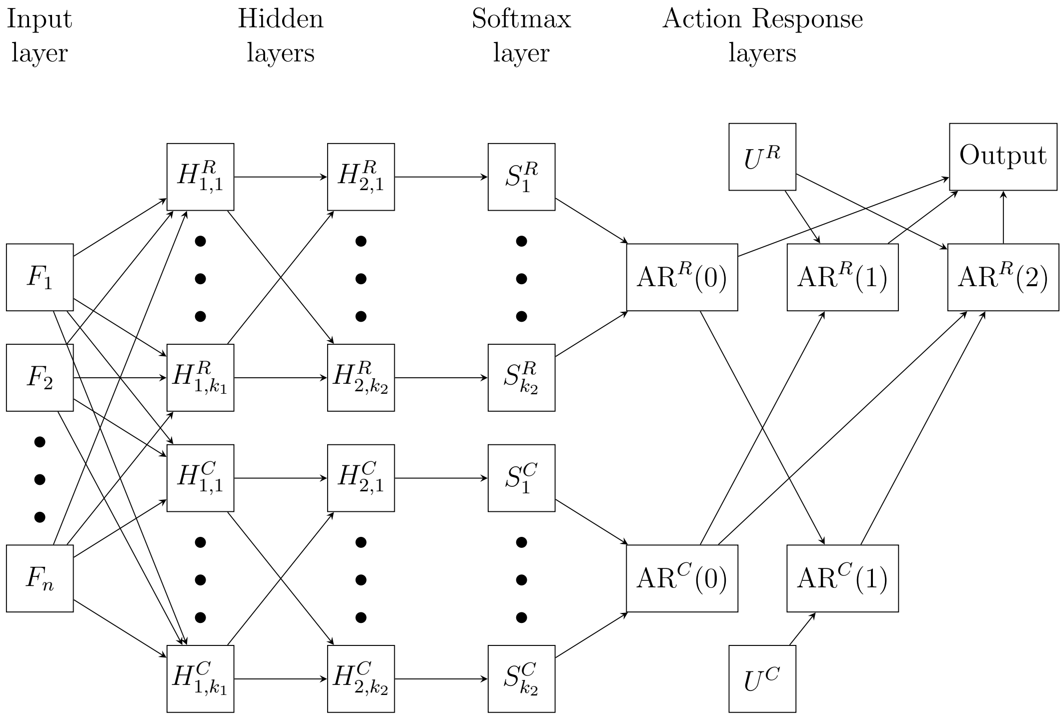


Figure 12: Architecture of the deep heuristic.

B.1 Feature Layers

The hidden layers are updated according to

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

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.

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

Figure 13: Examples of input feature matrices.

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

B.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 $\text{AR}^R(1)$ calculates the expected value from each action, conditioned on that expected play, and takes a softmax over those payoffs.

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

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

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

B.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.

B.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 $\text{AR}^R(1)$, 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.

C Detailed Results

C.1 Accuracies and Prediction Losses

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.

C.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 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, Wilcoxon rank test. As can bee seen in the tables below, all of theses tests are significant.

In Table 9 we see pairwise test for the difference in predictive ability of the optimized metaheuristic and alternatives. Prosocial EB is a model with prosocial preferences and

		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.

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 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. 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	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 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.

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 difference in prediction loss between the optimized metaheuristic model and alternatives 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 difference in prediction loss between the optimized metaheuristic model and alternatives for the common-interest games.

correct beliefs. We see that the optimized metaheuristic 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-interest treatment. For the competing-interest treatment, the difference is not significant for either QCH with prosocial preferences or the standard QCH.

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 difference in prediction loss between the optimized metaheuristic model and alternatives for the competing-interest games.

D 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 the 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 both parameters of the primitive heuristics and priors.

For each experimental session 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

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

where $m(\cdot \mid \theta)$ is the metaheuristic with parameters θ, $G_{\xi,t}$ is the game played in round t of sessions 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

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

A simple learning model is that 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

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

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

To simplify the model, we consider one population level model for each session. The behavior at 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 $G_{\xi,t}$. So, 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.

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.

Model	Estimation	Common	Competing
Meta heuristic	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 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, i.e. $\theta(0)$ 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 session in both treatments. We then predict the last 16 treatment games of each session in both treatments.

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

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