

Level-0 Meta-Models for Predicting Human Behavior in Games

JAMES R. WRIGHT, University of British Columbia
KEVIN LEYTON-BROWN, University of British Columbia

Behavioral game theory seeks to describe the way actual people (as compared to idealized, “rational” agents) act in strategic situations. Our own recent work has identified iterative models (such as quantal cognitive hierarchy) as the state of the art for predicting human play in unrepeated, simultaneous-move games [Wright and Leyton-Brown 2012]. Iterative models predict that agents reason iteratively about their opponents, building up from a specification of nonstrategic behavior called level-0. The modeler is in principle free to choose any description of level-0 behavior that makes sense for the given setting; however, in practice almost all existing work specifies this behavior as a uniform distribution over actions. In most games it is not plausible that even nonstrategic agents would choose an action uniformly at random, nor that other agents would expect them to do so. A more accurate model for level-0 behavior has the potential to dramatically improve predictions of human behavior, since a substantial fraction of agents may play level-0 strategies directly, and furthermore since iterative models ground all higher-level strategies in responses to the level-0 strategy. Our work considers “meta-models” of level-0 behavior: models of the way in which level-0 agents construct a probability distribution over actions, given an arbitrary game. We evaluated many such meta-models, each of which makes its prediction based only on general features that can be computed from any normal form game. We evaluated the effects of combining each new level-0 meta-model with various iterative models, and in many cases observed large improvements in the models’ predictive accuracies. In the end, we recommend a meta-model that achieved excellent performance across the board: a linear weighting of features that requires the estimation of five weights.

Categories and Subject Descriptors: I.2.11 [Artificial Intelligence]: Distributed Artificial Intelligence—*Multiagent systems*

Additional Key Words and Phrases: Behavioral game theory; bounded rationality; game theory; cognitive models; cognitive hierarchy; level- k ; prediction

1. INTRODUCTION

It is well known that the standard game-theoretic assumption that agents will adopt *Nash equilibrium* strategies—where each agent simultaneously responds optimally to all the others—is often a poor predictor of actual human behavior [e.g., see Goeree and Holt 2001]. The field of *behavioral game theory* aims to develop models that more accurately describe such behavior, as evaluated using experimental data [Camerer 2003]. Our own recent work has identified one particular model, *quantal cognitive hierarchy*, as the state of the art behavioral model for predicting human play in unrepeated, simultaneous-move games [Wright and Leyton-Brown 2012, 2013]. The quantal cognitive hierarchy (QCH) model has two key components. The first component is *quantal response*; that is, agents respond stochastically to their incentives—playing high utility actions with high probability and low utility actions with low probability—rather than best responding. This expresses the intuition that two actions that yield roughly

Author’s addresses: J. Wright and K. Leyton-Brown, Computer Science Department, University of British Columbia; email: {jrwright,kevinlb}@cs.ubc.ca.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

EC’14, June 8–12, 2014, Stanford, CA, USA.

ACM 978-1-4503-2565-3/14/06 ...\$15.00.

Copyright is held by the owner/author(s). Publication rights licensed to ACM.

<http://dx.doi.org/10.1145/2600057.2602907>

equal utilities have a roughly equal chance of being chosen. The second component is *iterative reasoning*; that is, agents do not reason arbitrarily deeply about their opponents' beliefs about beliefs about beliefs, but instead start from a simple nonstrategic strategy¹ (the *level-0* behavior), and then reason for some fixed number of iterations about responses to that strategy (e.g., a “level-2” agent quantally best responds to the combined behaviors of level-1 and level-0 agents).

Thus, in order to make use of a quantal cognitive hierarchy model one must first commit to a specification of level-0 behavior. Indeed, this is true of iterative models in general, such as cognitive hierarchy [Camerer et al. 2004] and level- k [Stahl and Wilson 1994; Nagel 1995; Costa-Gomes et al. 2001]. It is important to get this specification right, for two reasons. First, there is growing evidence that a substantial fraction of human players do act nonstrategically [Burchardi and Penczynski 2012; Wright and Leyton-Brown 2013]. Second, the level-0 model also effectively determines the predictions that will be made about strategic agents, because higher-level agents are assumed to act by responding strategically to lower-level agents' behavior.

Almost all work in the literature that uses iterative models adopts the specification that level-0 agents play a uniform distribution over actions. (In Section 5 we discuss the few exceptions of which we are aware, each of which is based on intuition about a specific setting of interest.) The uniform-distribution approach has the advantage that it does not require insight into a game's structure, and hence can be applied to any game. However, in many games it is not plausible that an agent would choose an action uniformly at random, nor that any other agent would expect them to do so. (For example, consider a dominated action that always yields very low payoffs for all players.)

In this paper we consider the question of how to do better. Specifically, we investigate general rules (“meta-models”) that can be used to induce a level-0 specification from the normal-form description of an arbitrary game. In the next section we formally define our setting, and describe the data, methods, and model that we used in our work. In Section 3, we propose a richer level-0 meta-model, computable directly from the normal form of a game. (We also investigated and ruled out a wide range of alternative meta-models; for completeness, these are described in an appendix.) In Section 4 we present the results of an experimental evaluation of the level- k , cognitive hierarchy, and quantal cognitive hierarchy models, augmented to use our level-0 specification. We found that performance always improved relative to the uniform level-0 specification, often dramatically. We then briefly survey some related work in Section 5 before concluding in Section 6.

2. FRAMEWORK

We focus on unrepeated, simultaneous-move *normal-form games*. Formally, a normal-form game G is a tuple (N, A, u) , where N is a finite set of *agents*; $A = \prod_{i \in N} A_i$ is the set of possible *action profiles*; A_i is the finite set of *actions* available to agent i ; $u = \{u_i\}_{i \in N}$ is a set of *utility functions* $u_i : A \rightarrow \mathbb{R}$, each of which maps from an action profile to a utility for agent i . Let $\Delta(X)$ denote the set of probability distributions over a finite set X . Overloading notation, we represent the expected utility of a profile of mixed strategies $s \in S = \prod_{i \in N} \Delta(A_i)$ by $u_i(s)$. We use the notation a_{-i} to refer to the joint actions of all agents except for i . Our objective is to find a behavioral model which maps from a game description G and a (human) agent i to a probability distribution

¹In this work, we refer to agents that form explicit beliefs about the beliefs and/or behaviors of the other agents as “strategic,” and agents that do not reason about other agents in this way as “nonstrategic”, hence the oxymoronic-seeming “nonstrategic strategy”. Nonstrategic should not be taken as a synonym for unsophisticated or thoughtless. Some of the level-0 behavior that we describe below is rather sophisticated.

over i 's action set A_i that predicts i 's behavior in G . In this section we first describe the model that we have shown achieves state-of-the-art performance in predicting human behavior, Spike-Poisson QCH [Wright and Leyton-Brown 2013]. We then describe the dataset and methods that we used to learn parameters and evaluate the performance of extensions to Spike-Poisson QCH.

2.1. Quantal Cognitive Hierarchy

The first key component of the quantal cognitive hierarchy model is quantal best response. Like most of the behavioral literature, we use the logit specification of this concept.

Definition 2.1 (Quantal best response). Let $u_i(a_i, s_{-i})$ be agent i 's expected utility when playing action a_i against strategy profile s_{-i} . Then a *quantal best response* $QBR_i(s_{-i}; \lambda)$ by agent i to s_{-i} is a mixed strategy s_i such that

$$s_i(a_i) = \frac{\exp[\lambda \cdot u_i(a_i, s_{-i})]}{\sum_{a'_i} \exp[\lambda \cdot u_i(a'_i, s_{-i})]}, \quad (1)$$

where λ (the *precision*) represents agents' sensitivity to utility differences. \square

Note that unlike (classical) best response, which is a set-valued function, quantal best response always returns a single mixed strategy. When $\lambda = 0$, quantal response mixes uniformly over all of the agents' actions; as $\lambda \rightarrow \infty$, quantal best response approaches best response.

The second key component of the quantal cognitive hierarchy model is iterative response, in which higher-level agents reason about and respond to lower-level agents. Describing the distribution of different levels in the population of agents is a crucial decision. We use the Spike-Poisson specification of this distribution [Wright and Leyton-Brown 2013].

Definition 2.2 (Spike-Poisson QCH model). Let $\text{Poisson}(m; \tau)$ denote a Poisson distribution with mean τ evaluated at index m . Let

$$g(m) = \begin{cases} \epsilon + (1 - \epsilon)\text{Poisson}(m; \tau) & \text{if } m = 0, \\ (1 - \epsilon)\text{Poisson}(m; \tau) & \text{otherwise.} \end{cases}$$

Let $QBR_i(s_{-i}; \lambda)$ denote i 's quantal best response to the strategy profile s_{-i} , given precision parameter λ . Let

$$\pi_{i,0:m} = \sum_{\ell=0}^m g(\ell) \frac{\pi_{i,\ell}}{\sum_{\ell'=0}^m g(\ell')}$$

be the "truncated" distribution over actions predicted for an agent conditional on that agent's having level $0 \leq \ell \leq m$. Now we can define $\pi_{i,m} \in \Delta(A_i)$, the distribution over actions predicted for an agent i with level m by the Spike-Poisson QCH model:

$$\begin{aligned} \pi_{i,0}^\emptyset(a_i) &= |A_i|^{-1} \\ \pi_{i,m}(a_i) &= QBR_i(\pi_{i,0:m-1}; \lambda). \end{aligned}$$

Notice that $\pi_{i,0}^\emptyset$ is just the uniform distribution; that is, level-0 agents randomize uniformly among their actions. In Section 3, we will construct more plausible definitions of level-0 behavior by replacing $\pi_{i,0}^\emptyset$ with richer distributions. This baseline model thus has three parameters: the mean of the Poisson distribution τ , the spike probability ϵ , and the precision λ . The overall predicted distribution of actions is a weighted sum of

Table I. Our datasets; units are in expected value.

Name	Source	Games	n	Units
SW94	Stahl and Wilson [1994]	10	400	\$0.025
SW95	Stahl and Wilson [1995]	12	576	\$0.02
CGCB98	Costa-Gomes et al. [1998]	18	1566	\$0.022
GH01	Goeree and Holt [2001]	10	500	\$0.01
CVH03	Cooper and Van Huyck [2003]	8	2992	\$0.10
HSW01	Haruvy et al. [2001]	15	869	\$0.02
HS07	Haruvy and Stahl [2007]	20	2940	\$0.02
SH08	Stahl and Haruvy [2008]	18	1288	\$0.02
COMBO8	400 samples from each dataset above	111	3200	\$0.01
RPC09	Rogers et al. [2009]	17	1210	\$0.01

the distributions for each level:

$$\Pr(a_i | G, \tau, \epsilon, \lambda) = \sum_{\ell=0}^{\infty} g(\ell) \pi_{i,\ell}(a_i). \quad \square$$

In previous work, we found that the distributions of levels with the highest posterior probability tended to have large proportions of level-0 agents (the “spike”) and level-2 agents, but very few level-1 agents [Wright and Leyton-Brown 2012, 2013]. (We also found that the number of higher-level agents decreased steadily after level 2). The spike-Poisson specification allows us to represent a distribution with such a shape much more closely than the more common Poisson distribution, which is unimodal. At the same time, it requires only two parameters, unlike a categorical distribution, which requires a number of parameters linear in the maximum level.

2.2. Data

We analyzed data from the nine experimental studies summarized in Table I. In Stahl and Wilson [1994] experimental subjects played 10 normal-form games, where every point represented a 1% chance (per game) of winning \$2.50. In Stahl and Wilson [1995], subjects played 12 normal-form games, where each point of payoff gave a 1% chance (per game) of winning \$2.00. In Costa-Gomes et al. [1998] subjects played 18 normal-form games, with each point of payoff worth 40 cents. However, subjects were paid based on the outcome of only one randomly-selected game. Goeree and Holt [2001] presented 10 games in which subjects’ behavior was close to that predicted by Nash equilibrium, and 10 other small variations on the same games in which subjects’ behavior was *not* well-predicted by Nash equilibrium. The payoffs for each game were denominated in pennies. We included the 10 games that were in normal form. In Cooper and Van Huyck [2003], agents played the normal forms of 8 games, followed by extensive form games with the same induced normal forms; we include only the data from the normal-form games. Payoffs were denominated in 10 cent units. In Haruvy et al. [2001], subjects played 15 symmetric 3×3 normal form games. The payoffs were “points” representing a percentage chance of winning \$2.00 for each game. In Haruvy and Stahl [2007], subjects played 20 games, again for payoff points representing a percentage chance of winning \$2.00 per game. Stahl and Haruvy [2008] present new data on 15 games that contain strategies that are dominated in ways that are “obvious” to varying degrees, again for percentage chances of winning \$2.00 per game. Finally, in Rogers et al. [2009], subjects played 17 normal-form games, with payoffs denominated in pennies.

We represent each observation of an action by an experimental subject as a pair (G, a_i) , where a_i is the action that the subject took when playing as player i in game G . All games had two players, so each single play of a game generated two observations.

Rather than analyze each dataset separately, we combined eight² of them into a single omnibus dataset (COMBO8) containing observations from each source dataset. To ensure that each was equally represented, despite their differences in size, we included exactly 400 observations from each dataset (sampled uniformly without replacement). We also renormalized all games so that their payoffs were in expected cents. This was important because the precision parameter for quantal response is not *scale invariant*: the correct value of λ can differ depending upon the units in which payoffs are expressed.

2.3. Methods

To evaluate a given model’s performance, we performed 10 rounds of 10-fold cross-validation, following our past work [Wright and Leyton-Brown 2010]. Specifically, for each round, we randomly divided the dataset into 10 parts. For each of the 10 ways of selecting 9 parts from the 10, we computed the maximum likelihood estimate of the model’s parameters based on those 9 parts, using the Covariance Matrix Adaptation Evolution Strategy (CMA-ES) algorithm [Hansen and Ostermeier 2001]. We then determined the log likelihood of the remaining part given the prediction. We call the average of this quantity across all 10 parts the *cross-validated log likelihood*. The average (across rounds) of the cross-validated log likelihoods is distributed according to a Student’s- t distribution [see, e.g., Witten and Frank 2000]. We compared the predictive power of different behavioral models by comparing the average cross-validated log likelihood of the dataset under each model. We say that one model predicted significantly better than another when the 95% confidence intervals for the average cross-validated log likelihoods do not overlap.

We use likelihood as our scoring criterion because we are attempting to measure how closely the predicted distribution of play matches an underlying “true” distribution of play, based on a finite sample from this distribution. Although “accuracy” has a more intuitive interpretation, it has the undesirable property of only being sensitive to the most likely action; in contrast, we are interested in accurately modeling the entire distribution of play.

In Section 4.3, we analyze posterior distributions over the parameters of the models we consider. To compute these posteriors, we used a sequential Monte Carlo technique called annealed importance sampling, or AIS [Neal 2001]. AIS allows for efficient sampling from high dimensional distributions, similarly to Markov Chain Monte Carlo (MCMC) techniques. We used the proposal distributions and hyperparameters described in Wright and Leyton-Brown [2012]. We used the following uninformative prior distributions for the parameters of the QCH model and weighted linear meta-model (see Section 3):

$$\begin{aligned}\epsilon &\sim U[0, 1] \\ \tau &\sim U[0, 20] \\ \lambda &\sim U[0, 20] \\ \mathbf{w} &\sim \text{Dirichlet}(1, 1, 1, 1, 1, 1)\end{aligned}$$

The Dirichlet distribution on the weight vector amounts to a uniform distribution over all coefficient vectors that sum to 1. The prior distribution for λ warrants particular at-

²We held out the RPC09 dataset to test whether we had overfit the COMBO8 data during model selection (see Section 2.3). We chose RPC09 for this task because it contains the widest variety of game types.

tention. A more conventional choice of prior for a scaling parameter such as λ would be an exponential distribution. We chose to use a uniform distribution with wide support instead, to ensure that we would not privilege smaller values of λ , as the maximum likelihood fits for λ on COMBO8 tended to be rather small. The posterior medians were essentially identical to the maximum likelihood fits in the end, even with a prior that placed most of its weight on values above 1.

We performed our experiments on the *hermes* cluster of WestGrid (www.westgrid.ca), which consists of 2112 cores, each having either 2GB or 4GB of RAM. In total, the results reported in this paper required approximately 16 CPU years of machine time, primarily for model fitting.

3. LEVEL-0 META-MODEL

In this section we present a “meta-model” for computing a distribution of play for level-0 agents. This model can be applied to any normal form game, without relying on any game-specific domain expertise. We settled on this model after exploring a much larger range of other models not described here; see Appendix A for details. The model is based on features computed for each of an agent’s action. We first describe each feature, and then describe the way that we combine them to obtain a level-0 prediction.

3.1. Level-0 Features

The core idea of our work is that nonstrategic behavior need not be uniform. How then might a nonstrategic agent behave? We argue that agents consider simple rules (“features”) that recommend one or more actions. (Thus, all of our features have range $\{0, 1\}$.) To be eligible for inclusion in our level-0 specification, we required that rules not depend on beliefs about the actions of other agents in a game. The rules we propose below depend only on a game’s payoff structure. Like the meta-model itself, we obtained this list by exploring a much larger space of candidate features (many of which took on real rather than Boolean values); again, see Appendix A for details.

For each feature, we briefly describe its motivation and then formally define it. Many of our features have been investigated in both the classical and behavioral game theory literature in other contexts. In particular, the *maxmax* payoff, *maxmin* payoff, and *maxmax efficiency* features correspond to the *Optimistic*, *Pessimistic*, and *Altruistic* nonstrategic types in Costa-Gomes et al. [2001].

3.1.1. Maxmin payoff. A *maxmin* action for agent i is the action with the best worst-case guarantee. That is,

$$f^{\text{maxmin}}(a_i) = \begin{cases} 1 & \text{if } a_i \in \arg \max_{a'_i \in A_i} \min_{a_{-i} \in A_{-i}} u_i(a'_i, a_{-i}), \\ 0 & \text{otherwise.} \end{cases}$$

This is the safest action to play against hostile agents.

3.1.2. Maxmax payoff. In contrast, a *maxmax* action for agent i is the action with the best best case. That is,

$$f^{\text{maxmax}}(a_i) = \begin{cases} 1 & \text{if } a_i \in \arg \max_{a'_i \in A_i} \max_{a_{-i} \in A_{-i}} u_i(a'_i, a_{-i}), \\ 0 & \text{otherwise.} \end{cases}$$

An agent who wishes to maximize his possible payoff will play a *maxmax* action.

3.1.3. Minimax regret. Savage [1951] proposed the *minimax regret* criterion for making decisions in the absence of probabilistic beliefs. In a game theoretic context, it works as follows. For each action profile, an agent has a possible *regret*: how much more utility could the agent have gained by playing the best response to the other agents’ actions?

Each of the agent's actions is therefore associated with a vector of possible regrets, one for each possible profile of the other agents' actions. A minimax regret action is an action whose maximum regret (in the vector of possible regrets) is minimal. That is, if

$$r(a_i, a_{-i}) = u_i(a_i, a_{-i}) - \max_{a_i^* \in A_i} u_i(a_i^*, a_{-i})$$

is the regret of agent i in action profile (a_i, a_{-i}) , then

$$f^{\text{mmr}}(a_i) = \begin{cases} 1 & \text{if } a_i \in \arg \min_{a_i' \in A_i} \max_{a_{-i} \in A_{-i}} r(a_i', a_{-i}), \\ 0 & \text{otherwise.} \end{cases}$$

3.1.4. Fairness. A concern for the fairness of outcomes is a common feature of human play in strategic situations, as has been confirmed in multiple behavioral studies, most famously in the Ultimatum game [Thaler 1988; Camerer and Thaler 1995]. Let the unfairness of an action profile be the difference between the maximum and minimum payoffs among the agents under that action profile:

$$d(a) = \max_{i, j \in N} u_i(a) - u_j(a).$$

Then a "fair" outcome minimizes this difference in utilities. The *fairness* feature selects every action which is part of a minimally unfair action profile.

$$f^{\text{fair}}(a_i) = \begin{cases} 1 & \text{if } a_i \in \arg \min_{a_i' \in A_i} \min_{a_{-i} \in A_{-i}} d(a_i', a_{-i}), \\ 0 & \text{otherwise.} \end{cases}$$

3.1.5. Efficiency. A final reason that a nonstrategic agent might find an action profile desirable is that it produces the best overall benefit to the collection of agents. The *efficiency* feature selects every action that is part of some action profile that maximizes the sum of utilities:

$$f^{\text{efficient}}(a_i) = \begin{cases} 1 & \text{if } a_i \in \arg \max_{a_i' \in A_i} \max_{a_{-i} \in A_{-i}} \sum_{j \in N} u_j(a_i', a_{-i}), \\ 0 & \text{otherwise.} \end{cases}$$

3.2. Weighted Linear Meta-Model

The model of level-0 behavior that we recommend is a linear weighting of "informative" features for a game. We call a feature informative when it is able to distinguish at least one pair of actions.

Definition 3.1 (Informative features). A feature f is *informative* in a game G if there exists $i \in N$ and $a_i, a_i' \in A_i$ such that $f(a_i) \neq f(a_i')$. Let I be an indicator function for informative features:

$$I(f) = \begin{cases} 1 & \text{if } \exists i \in N, a_i, a_i' \in A_i : f(a_i) \neq f(a_i'), \\ 0 & \text{otherwise.} \end{cases}$$

The weighted linear specification of level-0 behavior produces a prediction over actions for a given agent by taking a weighted sum of feature outputs for each action, and then normalizing to produce a distribution.

Definition 3.2 (Weighted linear level-0 specification). Let F be a set of features mapping from an action to $\{0, 1\}$.³ For each feature $f \in F$, let $w_f \in [0, 1]$ be a weight

³We restrict attention to binary features, but this model is easily extended to real-valued features.

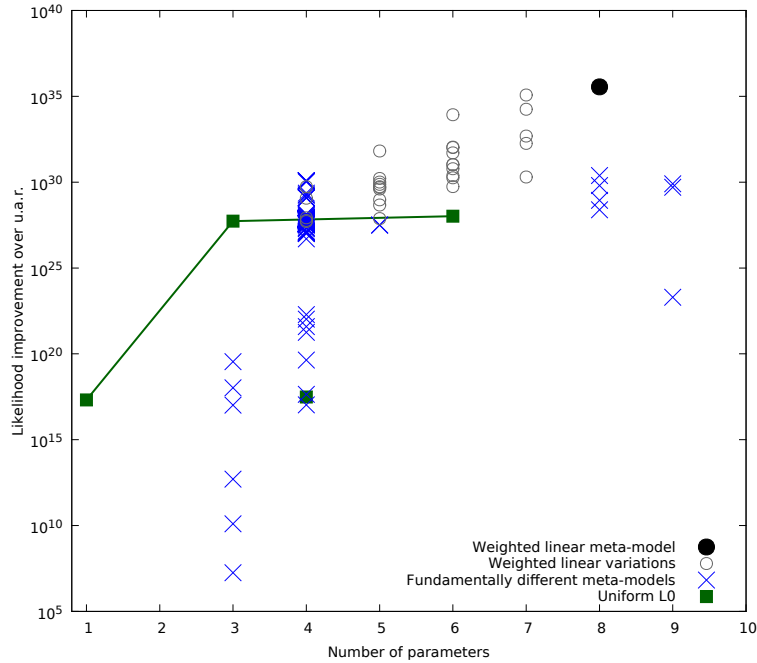


Fig. 1. Average likelihood ratios of model predictions to random predictions across different models. The “efficient frontier” of models from our past work [Wright and Leyton-Brown 2012] is indicated in green; the rightmost “uniform L0” datapoint is QCH5, the best-performing model from that work. The solid black dot is the weighted linear meta-model that we describe in later sections. The open circles are the weighted linear model using different subsets of features. The blue X’s are alternative meta-models that we considered; observe that many of them perform worse than the uniform meta-model.

parameter. Let $\sum_{f \in F} w_f \leq 1$, and let $w_0 = 1 - \sum_{f \in F} w_f$. Then the *weighted linear level-0 specification* predicts the following distribution of actions for level-0 agents:

$$\pi_{i,0}^F(a_i) = \frac{w_0 + \sum_{f \in F} w_f I(f) f(a_i)}{\sum_{a'_i \in A_i} w_0 + \sum_{f \in F} w_f I(f) f(a'_i)}.$$

Notice that $\pi_{i,0}^F$ reduces to the uniform distribution $\pi_{i,0}^\emptyset$ when no features are informative.

Given an iterative model, we can construct an extended iterative model by replacing its level-0 specification. For example, we extended the Spike-Poisson QCH model of Section 2.1 by replacing $\pi_{i,0}^\emptyset$ with $\pi_{i,0}^F$, setting F to be the set of features presented in Section 3.1.

4. RESULTS

In this section we present the results of our experiments. First, we justify our choice of the weighted linear meta-model by illustrating the outcome of our model selection procedure. Second, we more deeply investigate the performance of our selected level-0 meta-model when combined with multiple iterative models. Finally, we interpret the values we fit for the weighted linear meta-model’s parameters.

4.1. Model Selection

In our attempts to design a highly predictive level-0 meta-model, we explored a wide range of alternative meta-model and feature specifications (see Appendix A). This section describes the range of empirical performance we obtained across the different meta-models we investigated, and thus helps to explain our focus on the weighted linear meta-model. We also investigated *variants* of this meta-model, evaluating all subsets of our five features. In the end, we selected weighted linear with all five (binary) features.

Our methodology for model selection was as follows. We evaluated candidate meta-models in terms of their cross-validated likelihood on the COMBO8 dataset when used in conjunction with the spike-Poisson QCH model. We used cross-validation to guard against overfitting the parameters of the models that we evaluated. To ensure that our results were not skewed because we had selecting a model that was overly specialized to a single dataset or because we had overfit this dataset via model selection itself, we also held out one dataset (RPC09) for final testing, which we did not consider until the experiments reported in Section 4.2.

Figure 1 shows the performance of all the models that we considered. The x -axis indicates the total number of parameters when the given meta-model is combined with (3-parameter) spike-Poisson; the y -axis indicates the cross-validated performance of the model on COMBO8. The solid black datapoint is the 8-parameter weighted linear meta-model that we selected. The gray circles are variations on the weighted linear model (feature subsets). To help give meaning to the likelihood improvement scale, we also include the performance of four iterative models from the literature, all of which use uniform level-0 specifications. The green line connects Poisson-CH, spike-Poisson QCH and QCH5, three models which our past work [Wright and Leyton-Brown 2013] identified as lying on a Pareto frontier trading off complexity with performance. This allows us to put our dramatic performance improvements using richer level-0 models in the context of earlier improvements: QCH5 achieved less than a factor of 10 better performance than spike-Poisson QCH in exchange for doubling the number of parameters; in contrast weighted linear improves upon spike-Poisson's performance by a factor of 100,000,000.⁴

We also explored several real-valued versions of our set of features in addition to the binary valued versions presented in Section 3.1. We were surprised to discover that they tended to yield worse performance than the same meta-model with binary-only features.⁵

4.2. Model Performance

We compared the predictive performance of three iterative models using two different specifications of level-0 behavior. The results are displayed in Figure 2. The y -axis gives the average ratio of the cross-validated likelihood of the extended models' predictions divided by the likelihood of a uniform random prediction; in other words, each bar is the factor by which an extended model predicts better than a purely random guess. We present results for both cross-validation (i.e., testing on held-out data from the same

⁴In all cases, note that because we consider cross-validated likelihood ratios, we do not need to worry that more flexible models achieve better performance because they overfit their training data.

⁵We also explored several features based on the idea that agents may be drawn to "salient" actions for reasons unrelated to the structure of the game (e.g., based on payoff values' position in the game matrix or number of digits). Some of these features did prove useful, but we nevertheless omitted them from our final analysis. This was because we wanted to focus on meta-models that were completely general, in the sense of being based entirely on the mathematical structure of the game, without requiring details of the game's presentation; and indeed, in our dataset it was not always clear how a given game had been presented to subjects, making the features ambiguous.

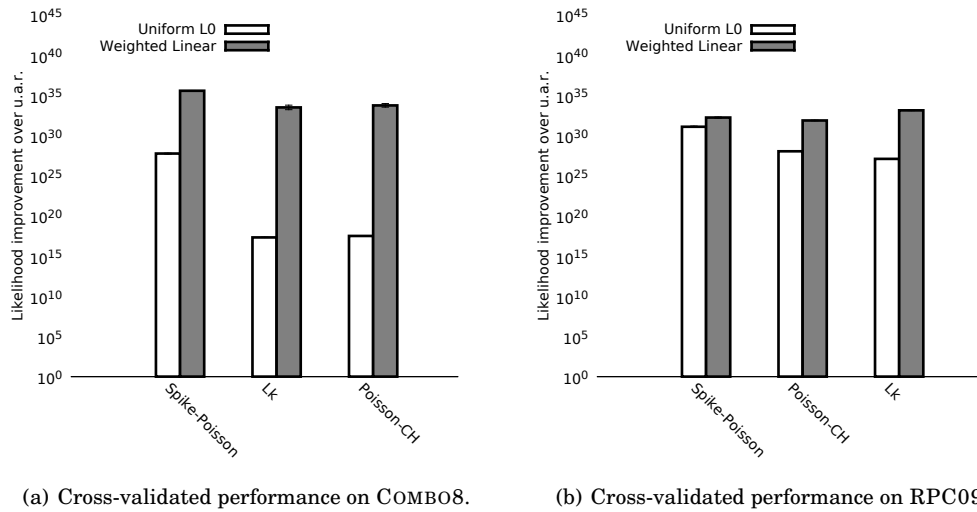


Fig. 2. Average likelihood ratios of model predictions to random predictions, with 95% confidence intervals. Results are shown for three different iterative models (Poisson cognitive hierarchy [Camerer et al. 2004], level- k [Costa-Gomes et al. 2001], and spike-Poisson quantal cognitive hierarchy [see Section 2.1 and Wright and Leyton-Brown 2013]) using two different level-0 meta-models (uniform randomization and weighted linear (Section 3.2)).

distribution used to train the model) and RPC09, a dataset that we did not use for any purpose—including the evaluation of different candidate meta-models and features—before generating the figure presented here. Observe that while the differences seem smaller on RPC09, this is largely an artifact of the way likelihood ratios are computed: RPC09 has about half as many data points as COMBO8, and all things being equal, likelihood differences grow with dataset size. Notice also that the difference between the three models under a uniform level-0 specification are small on RPC09; this may indicate that there is just less room for improvement in this dataset for a more sophisticated model to extract.

The weighted linear model used all five features productively, but two of them had much higher weights (see Section 4.3). Nevertheless, omitting the remaining three features consistently worsened prediction performance.

The weighted linear meta-model achieved the best performance when used to augment each of the three iterative models. The worst-performing models benefited disproportionately from the weighted linear meta-model, especially on COMBO8. Spike-Poisson QCH performed decisively better than the other two models under a uniform level-0 meta-model, but the three models had much more similar (and improved) performance under a weighted linear meta-model.

4.3. Parameter Analysis

In this section, we examine and interpret the posterior distributions for the spike-Poisson model’s parameters under the two different level-0 specifications, following methodology described in our past work [Wright and Leyton-Brown 2012]. Here we concentrate on the COMBO9 dataset; our findings for RPC09 were qualitatively similar except for one exception, noted below.

Figure 3 shows the marginal posterior distribution for each level in the population (up to level 3), for both the weighted linear meta-model and the uniform meta-model. There were two broad patterns. The first was a striking shift of probability mass from

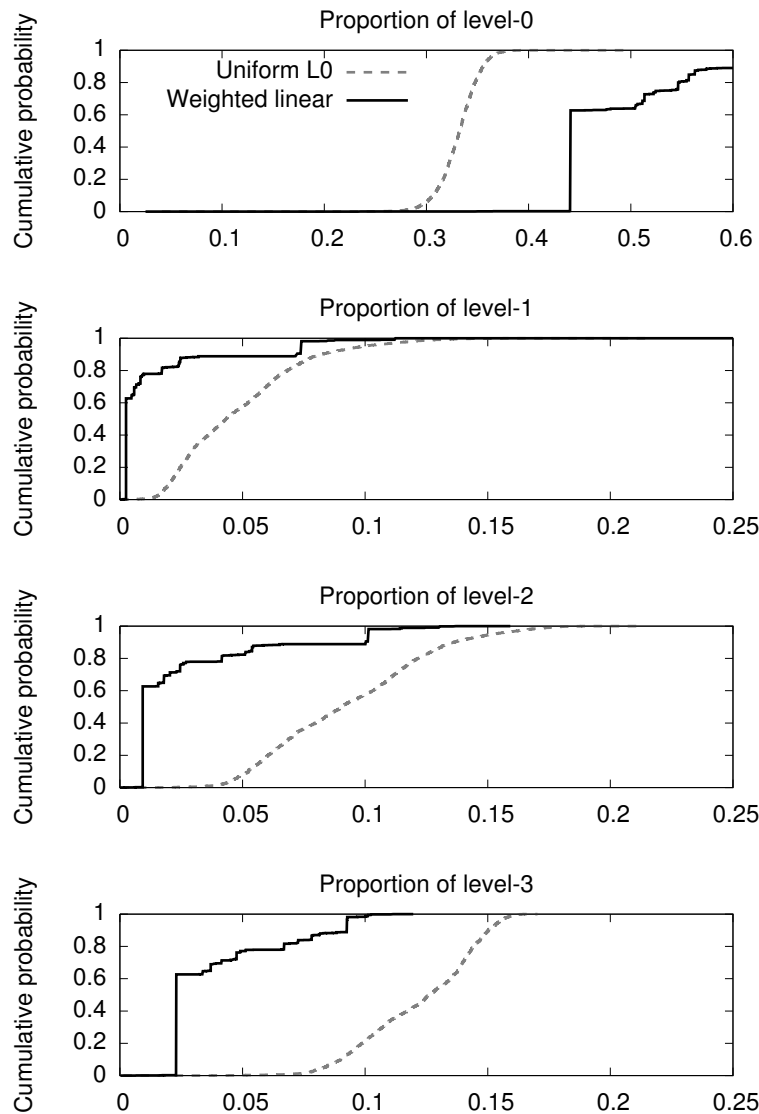


Fig. 3. Marginal cumulative posterior distributions of levels of reasoning in the COMBO8 dataset, for spike-Poisson QCH with weighted linear and uniform meta-models.

higher-level agents to level-0 agents under the weighted linear meta-model; the median posterior estimate for the proportion of level-0 agents was 0.44, compared with 0.33 for the uniform meta-model. For every nonzero level, the proportion estimated by the weighted linear meta-model was lower than that estimated by the uniform meta-model. As described in the introduction, there are two ways that an improved level-0 meta-model could improve an iterative model's predictions. First, it could simply describe the actions of level-0 agents more accurately. Second, it could also allow the model to more accurately describe the actions of higher-level agents, because higher-level agents form their strategies by best responding to the strategies of lower-level

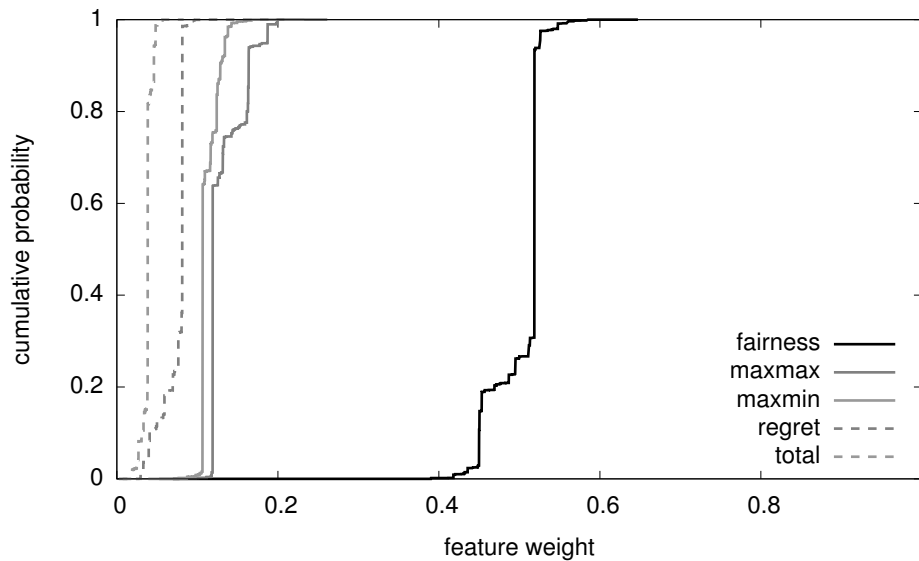


Fig. 4. Marginal cumulative posterior distributions over weight parameters of the weighted linear meta-model, for spike-Poisson QCH on the COMBO8 dataset.

agents, ultimately “grounding out” in a best response to the level-0 strategy. The large increase in the estimated number of level-0 agents emphasizes the importance of understanding these agents’ behavior, both for predicting their actions *and* for predicting the behavior of the agents who respond to them.

The second pattern was a much sharper identification of the posterior estimated proportions. The vertical portions of the cumulative distributions represent modes of the density function. Each level has a very sharp modal estimate under the weighted linear meta-model. In contrast, the marginal distributions under the uniform meta-model, although roughly unimodal, do not have pronounced modes. Put another way, the data argues strongly for much more specific estimates under the weighted linear meta-model than under the uniform meta-model.

Figure 4 shows the marginal posterior distribution for the weights of the linear weighted meta-model on the COMBO8 dataset. As with the distribution over levels, the posterior distributions on the weight parameters had very tall modes and narrow support, indicating that the data argued strongly for specific ranges of parameter values; the fairness parameter seems to have two modes of differing heights, but both of these modes were in qualitatively the same range of the space (i.e., much larger than any of the other features). The fairness parameter had by far the highest median posterior weight; the remaining weight was split approximately evenly among the uniform noise, maxmax payoff, and maxmin payoff features. While both minimax regret and efficiency had very small median posterior weights, we observed a substantial performance degradation during model selection when we omitted them altogether.

The precision parameter, λ , was stable across both meta-models, with average fitted values of 0.24 ± 0.001 for the uniform meta-model and 0.19 ± 0.0001 for the weighted-linear meta-model. This parameter was the only one for which we saw qualitatively different values on RPC09. On that distribution λ took very large values (60 and above), which again occurred across both meta-models.

5. RELATED WORK

Almost every study that employs iterative reasoning models of either the level- k or cognitive hierarchy types assumes a uniform distribution of play for level-0 agents. However, there are a few exceptions. Crawford and Iriberry [2007b] specified truth-telling as the single salient action in first-price auctions. Crawford and Iriberry [2007a] manually designated certain actions as “salient” (based on visual features such as “left-most”) in a hide-and-seek game. They then estimated an iterative model with a level-0 specification in which level-0 agents play salient actions, with the strengths of each action’s salience estimated under the assumption that no agent truly plays a level-0 distribution. Arad and Rubinstein [2009] specified a single action of reinforcing all battlefields equally in a Col. Blotto game as the sole level-0 action. Arad and Rubinstein [2012] specified the highest possible request as the level-0 action in a money-request game where players receive the amount of money they request, but also receive a relatively large bonus for requesting exactly 1 shekel less than the other player. Arad [2012] manually specified two “anchor” strategies for a Col. Blotto-like game in which players simultaneously assign four representatives to four separate contests in order of the representatives’ ability.

In spite of the crucial dependence of iterative models upon the specification of the level-0 distribution, few studies have empirically investigated level-0 play. Agranov et al. [2010] incentivized subjects to choose an action quickly (and then to revise it after thinking) by imposing a randomized time limit when playing the beauty-contest game of Nagel [1995]. They hypothesized that early actions represent level-0 choices and that later actions represent higher-level choices. Based on this assumption, they found that level-0 behavior did not differ significantly from a uniform distribution. In contrast, Burchardi and Penczynski [2012] incentivized players to reveal their reasoning by allowing a one-time simultaneous exchange of messages between teammates playing a beauty-contest game. Teams of two simultaneously sent each other a single message containing arguments in favor of a given action, and then simultaneously chose an action, with the team’s action being chosen randomly from the two choices. Burchardi and Penczynski then classified each argument according to level of reasoning, and extracted the level-0 behavior from both level-0 and higher-level arguments. They found that level-0 play was significantly different from uniform. Intriguingly, they also found that the level-0 behavior hypothesized by higher-level players was very similar to the level-0 behavior that was actually proposed.

6. CONCLUSIONS

Our main result is a specification of level-0 behavior that dramatically improves the performance of each of the iterative models we evaluated—level- k , cognitive hierarchy, and quantal cognitive hierarchy. This specification depends only upon the payoffs of the game, and is thus generally applicable to any domain, even ones in which human intuition gives little guidance about the level-0 specification. A linear weighting of five nonstrategic features (maxmax payoff, maxmin payoff, minmin unfairness, minimax regret, and maxmax payoff sum) improved all three models’ performances to the extent that it nearly erased differences between the original models. Omitting any of the features degraded the performance of this model, even though the median posterior weights of some features were much larger than others.

The conventional wisdom in economics is that level-0 agents exist only in the minds of higher level agents; that is, that a level-0 specification acts solely as a starting point for higher level agents’ strategies. Our results argue against this point of view: the best performing model estimated that over half of the agents were level-0. These results are strong evidence that nonstrategic behavior is an important aspect of human behavior,

even in strategic settings. Further refining our understanding of nonstrategic behavior is an important direction for future work, both in terms of the factors that are considered by nonstrategic agents, and in terms of the details of incorporating these factors into predictive behavioral models.

Acknowledgements

Thanks to Colin Camerer for very helpful early discussions about features that could form the basis of richer level-0 meta-models.

REFERENCES

- AGRANOV, M., CAPLIN, A., AND TERGIMAN, C. 2010. The process of choice in guessing games. Social science working paper 1334r, California Institute of Technology. Available at <http://www.wordsmatter.caltech.edu/SSPapers/sswp1334R.pdf>, accessed Feb. 9, 2014.
- ARAD, A. 2012. The tennis coach problem: A game-theoretic and experimental study. *The BE Journal of Theoretical Economics* 12, 1.
- ARAD, A. AND RUBINSTEIN, A. 2009. Colonel Blotto's top secret files. Working paper. Available at <http://www.dklevine.com/archive/refs4814577000000000432.pdf>, accessed Feb. 9, 2014.
- ARAD, A. AND RUBINSTEIN, A. 2012. The 11–20 money request game: A level- k reasoning study. *The American Economic Review* 102, 7, 3561–3573.
- ARIELY, D. AND WALLSTEN, T. S. 1995. Seeking subjective dominance in multidimensional space: An explanation of the asymmetric dominance effect. *Organizational Behavior and Human Decision Processes* 63, 3, 223–232.
- BURCHARDI, K. AND PENCZYNSKI, S. 2012. Out of your mind: Eliciting individual reasoning in one shot games. Working paper, London School of Economics. Available at <http://people.su.se/~kburc/research/BurchardiPenczynski2012.pdf>, accessed Feb. 9, 2014.
- CAMERER, C., HO, T., AND CHONG, J. 2004. A cognitive hierarchy model of games. *Quarterly Journal of Economics* 119, 3, 861–898.
- CAMERER, C. AND THALER, R. H. 1995. Anomalies: Ultimatums, dictators and manners. *The Journal of Economic Perspectives* 9, 2, 209–219.
- CAMERER, C. F. 2003. *Behavioral Game Theory: Experiments in Strategic Interaction*. Princeton University Press.
- COOPER, D. AND VAN HUYCK, J. 2003. Evidence on the equivalence of the strategic and extensive form representation of games. *Journal of Economic Theory* 110, 2, 290–308.
- COSTA-GOMES, M., CRAWFORD, V., AND BROSETA, B. 1998. Cognition and behavior in normal-form games: an experimental study. Discussion paper 98-22, UCSD. Available at <http://weber.ucsd.edu/~vcrawfor/ucsd9822.pdf>, accessed Feb. 9, 2014.
- COSTA-GOMES, M., CRAWFORD, V., AND BROSETA, B. 2001. Cognition and behavior in normal-form games: An experimental study. *Econometrica* 69, 5, 1193–1235.
- CRAWFORD, V. AND IRIBERRI, N. 2007a. Fatal attraction: Salience, naïveté, and sophistication in experimental “hide-and-seek” games. *American Economic Review* 97, 5, 1731–1750.
- CRAWFORD, V. AND IRIBERRI, N. 2007b. Level- k auctions: Can a nonequilibrium model of strategic thinking explain the winner's curse and overbidding in private-value auctions? *Econometrica* 75, 6, 1721–1770.
- GOEREE, J. K. AND HOLT, C. A. 2001. Ten little treasures of game theory and ten intuitive contradictions. *American Economic Review* 91, 5, 1402–1422.

- HANSEN, N. AND OSTERMEIER, A. 2001. Completely derandomized self-adaptation in evolution strategies. *Evolutionary Computation* 9, 2, 159–195.
- HARUVY, E. AND STAHL, D. 2007. Equilibrium selection and bounded rationality in symmetric normal-form games. *Journal of Economic Behavior and Organization* 62, 1, 98–119.
- HARUVY, E., STAHL, D., AND WILSON, P. 2001. Modeling and testing for heterogeneity in observed strategic behavior. *Review of Economics and Statistics* 83, 1, 146–157.
- NAGEL, R. 1995. Unraveling in guessing games: An experimental study. *American Economic Review* 85, 5, 1313–1326.
- NEAL, R. M. 2001. Annealed importance sampling. *Statistics and Computing* 11, 2, 125–139.
- ROGERS, B. W., PALFREY, T. R., AND CAMERER, C. F. 2009. Heterogeneous quantal response equilibrium and cognitive hierarchies. *Journal of Economic Theory* 144, 4, 1440–1467.
- SAVAGE, L. 1951. The Theory of Statistical Decision. *Journal of the American Statistical Association* 46, 253, 55–67.
- SIMONSON, I. 1989. Choice based on reasons: The case of attraction and compromise effects. *Journal of Consumer Research*, 158–174.
- STAHL, D. AND HARUVY, E. 2008. Level- n bounded rationality and dominated strategies in normal-form games. *Journal of Economic Behavior and Organization* 66, 2, 226–232.
- STAHL, D. AND WILSON, P. 1994. Experimental evidence on players' models of other players. *Journal of Economic Behavior and Organization* 25, 3, 309–327.
- STAHL, D. AND WILSON, P. 1995. On players' models of other players: Theory and experimental evidence. *Games and Economic Behavior* 10, 1, 218–254.
- THALER, R. H. 1988. Anomalies: The ultimatum game. *The Journal of Economic Perspectives* 2, 4, 195–206.
- WITTEN, I. H. AND FRANK, E. 2000. *Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations*. Morgan Kaufmann.
- WRIGHT, J. R. AND LEYTON-BROWN, K. 2010. Beyond equilibrium: Predicting human behavior in normal-form games. In *AAAI-10*. 901–907.
- WRIGHT, J. R. AND LEYTON-BROWN, K. 2012. Behavioral game-theoretic models: A Bayesian framework for parameter analysis. In *AAMAS-2012*. Vol. 2. 921–928.
- WRIGHT, J. R. AND LEYTON-BROWN, K. 2013. Predicting human behavior in un-repeated normal-form games. Working paper, University of British Columbia. arXiv:1306.0918 [cs.GT].

A. OTHER MODELS AND FEATURES

In this appendix we provide a brief overview of some of the different functional forms and features that we rejected in favor of the weighted linear meta-model.

Logit and log link functions. In addition to the weighted linear form, we also investigated generalized linear models with a logit link function—i.e., we computed weighted linear sums and then normalized over the exponentials of the sums. We also looked at the opposite: normalizing over the log of the sums. The rationale for the logit link was that, as with quantal response, broadly similar sums ought to have broadly similar probabilities. The rationale for the log translation was that, as values grow large, larger differences would be required to yield the same impact on level-0 probabilities. We used these functional forms particularly for non-binary feature specifications, in which an “almost” maxmax payoff action might be expected nearly as often as the maxmax payoff action.

Ordered veto. In an attempt to reduce the number of parameters required to specify a weighted linear meta-model, we deeply investigated a family of models that we refer to as “ordered veto”. In this specification, a single weight is learned for a fixed, ordered list of features. The prediction of the first informative feature (in the sense described in Section 3.2) is combined with the uniform prediction according to the single weight to produce the level-0 prediction.

Using the ordered veto level-0 specification adds only a single parameter (w_1) to any iterative model. It is also more cognitively plausible, as multiple behavioral studies have proposed that agents attempt to simplify decision tasks by searching for a decision that can be justified by a single attribute comparison without the need for explicit tradeoffs [e.g., see Simonson 1989; Ariely and Wallsten 1995]. The optimal feature set for ordered veto contained only two features (fairness and maxmax payoff). Although the ordered veto model yielded incremental improvements to prediction performance over the uniform-random meta-model, we still obtained substantially worse performance than that of the linear meta-model.

Veto voting. The fact that the optimal feature set for ordered veto contains only two features was surprising to us; for the weighted linear meta-models, adding a new element to the feature set always improved performance. On the theory that there must be information in the additional features as well, we modified the ordered veto model as follows. First, divide the features into two sets: the veto features, and the voting features. Predict using the first informative veto feature, as in ordered veto. If neither of the veto features is informative, take an average among the voting features.

Naïve Bayesian. We evaluated two specifications of a Bayesian model, in which we conditioned on features of an action to compute the probability that a level-0 agent would play the action. If an action had a feature, its probability would be adjusted upward by a parameterized factor. The two specifications coincided for binary features, but differed for real valued features.

Informativeness. In addition to binary informativeness, which can only be applied to binary features, we evaluated two other approaches. The first was to omit informativeness checks. This led to serious degradations in performance when used with binary features. The second was to generalize the notion of informativeness. We attempted to capture the degree to which a feature distinguished among actions; we measured this degree of distinction by the Shannon entropy of the (normalized) distribution of feature values among the actions.

Soft features. The binary features indicate whether a given action maximizes or minimizes a particular quantity (e.g., maximum regret, minimum payoff). Rather than use binary features indicating whether an action attains an extreme value of these quantities, we also evaluated features that used the quantities directly. Another real-valued version of the features was a logit translation, which was intended to express how likely an agent would be to believe that the action attained the extreme value of the quantity (e.g., had the best best case among all the actions) under a particular error structure.

Saliency features. In addition to the features that we chose, we also evaluated features that aimed to capture an action's "saliency"; that is, how easily the action would come to an agent's attention, based on the way in which it had been presented. Many of these features broadly performed well, but in the interests of a more generally applicable model, we eliminated them in order to focus on features based solely on the strategic structure of the game. We evaluated the following.

- (1) "First" action: The "first" action is the visually first item presented to the experimental subject; in matrix games, this is the topmost action for the row player and the leftmost action for the column player. Empirically there is evidence that the first element of a set is more available to an agent for processing.
- (2) Round numbers: Actions that had a payoff ending in 0 were flagged by this feature, on the theory that round numbers are more focal. There were two variations depending on whether other agents' payoffs were considered or not.
- (3) Digit count: Actions that had payoffs with unusual numbers of digits (longer or shorter) were flagged by this feature. There were two variations depending on whether other agents' payoffs were considered or not.