No Evidence for Cost-Benefit Arbitration Between Social Learning Strategies

Ariel Levy, Xavier Roberts-Gaal & Fiery Cushman
ariellevy@g.harvard.edu, xavierrobertsgaal@g.harvard.edu, cushman@fas.harvard.edu

Department of Psychology, Harvard University
William James Hall, 33 Kirkland Street, Cambridge MA

Abstract

When learning a task by observing another person performing it, an individual can either focus on imitating the other's behavior (policy imitation), or attempt to infer the other's goals and beliefs and adjust their own behavior accordingly (goal emulation). Imitation is considered to be computationally cheap but less accurate, while emulation is considered to be computationally costly but more accurate. Drawing upon research on computational resource rationality, we ask whether individuals incorporate cost-benefit considerations when choosing whether to imitate actions or emulate goals. To answer this question, we used an observational-learning extension of a two-step bandit task, and manipulated the reward at stake. Participants' behavior was best fit by a dual-process model of goal emulation and one-step imitation, consistent with findings from previous research. However, contrary to our hypothesis and inconsistent with cost-benefit arbitration, we found no evidence that rewards at stake influenced participants' social learning strategies.

Keywords: social learning; goal emulation; policy imitation; resource-rationality; computational modeling;

Introduction

Human behavior sometimes exhibits remarkable rationality and outstanding inferential skills, yet at other times succumbs to equally striking instances of folly, driven by systematic biases and errors. The AI-inspired framework of computational resource-rationality (Gershman et al., 2015; Lieder & Griffiths, 2020) provides one explanation for this inconsistency: we are limited in the cognitive resources we can bring to bear on each decision, bounded both by our neurobiology and by the time pressure under which we must act. One way to operate under these constraints is through cognitive meta-control, weighing the expected cost of performing exhaustive, accurate computations against the utility such computations might yield. When the cost outweighs the potential benefit, we rely on computationally inexpensive but typically less accurate heuristics or decision-making strategies.

Resource-rationality has particularly influenced research on how humans arbitrate between different learning strategies when they must learn individually through trial-and-error. Abundant evidence has established that humans (like animals) employ two primary classes of reinforcement learning strategies when they learn individually: quick and inflexible model-free learning and costly but accurate model-based learning (Daw et al., 2005, 2011). Subsequent research employing the resource-rationality framework has demonstrated that cost-benefit considerations guide how ee arbitrate between these two strategies (Kool et al., 2016, 2017). In a reinforcement learning task with varying stakes, individuals were more likely to engage in costly model-based learning when stakes were higher. Importantly, this effect was observed only in environments where model-based control was expected to yield greater rewards than model-free control. These results are consistent with the existence of a cognitive meta-control mechanism that estimates the expected rewards of each learning strategy to decide which to employ.

Like individual reinforcement learning, social learning is often conceptualized as arbitration between two learning mechanisms with distinct costs and benefits (Charpentier et al., 2020; Roberts-Gaal & Cushman, 2023; Wu et al., 2022). When learning by observing another person performing a task, an individual can either prioritize imitating the other's behavior exactly (henceforth, policy imitation), or else concentrate on inferring the other's goals and beliefs from their actions so as to find the best way to achieve the same outcome (goal emulation).

Like model-free reinforcement learning, imitation is computationally cheap, as it does not require representing a causal model of the environment. Imitation can be less accurate, though, as observed behavior may be effective only in a narrow range of circumstances. In contrast, mirroring model-based reinforcement learning, goal emulation is computationally demanding. It relies on Bayesian inference to attribute mental states from overt behavior (Baker et al., 2009; Jara-Ettinger et al., 2016), and requires the costly representation of a causal model of the environment for planning. Nevertheless, it is more flexible, and it can lead to higher accuracy even in unstable environments. Applying the resource-rational framework to social learning results in a straightforward hypothesis: individuals are likely to adopt a costly goal emulation strategy only when its anticipated benefits exceed its computational costs. At other times, they will opt for a simpler imitation strategy.

However, transferring the concept of resource-rationality from individual learning to social learning is not trivial, as some social cognition literature explicitly contradicts the way we framed the contrast between imitation and emulation. First, a line of research characterizes theory of mind as an automatic, rather than deliberative, process—such that people often represent the mental state of others without consciously directed effort, even when it is irrelevant to the task at hand (Kleiman et al., 2022; Kulke et al., 2018; Samson et al., 2010; Schneider et al., 2017). Second, in studies comparing the social learning abilities of humans and other primates (e.g., Call et al., 2005; Nagell et al., 1993), non-

In D. Barner, N.R. Bramley, A. Ruggeri and C.M. Walker (Eds.), *Proceedings of the 47th Annual Conference of the Cognitive Science Society*

human primates principally employ goal emulation, while high fidelity imitation is unique to humans. From this perspective, precise imitation of the fine details of others' actions may be what gives humans a distinctive advantage in social learning over other species. These perspectives challenge the idea that emulation is necessarily computationally costlier than imitation.

In the present study, we ask whether participants treat goal emulation as more costly than policy imitation and, if so, whether they arbitrate between these strategies by considering their respective costs and benefits, opting for the more cognitively demanding strategy when it is expected to pay off. To answer this question, we employed an observational learning paradigm—loosely adapted from Charpentier et al. (2020)—specifically designed to differentiate between goal emulation and policy imitation. We created a task environment where emulation yields a higher reward than imitation and verified this property by simulations. Additionally, we incorporated a within-participants manipulation of stakes, defined as the magnitude of the potential reward available on a trial.

Computational Models

The experimental paradigm we used to investigate arbitration between goal emulation and policy imitation was modeled after Charpentier et al. (2020). This is an observational learning extension of a two-step bandit task, framed, in our case, as a fishing game. In some of the trials, participants observe the actions of a virtual mentor ("observe" trials). In the remaining trials, participants decide between actions themselves ("play" trials). Each trial begins in an initial state s_0, with three possible actions $a \in \{a_1, a_2, a_3\}$. However, only two of these three actions are available for participants to choose from on any given trial. This aspect is necessary to tease apart the strategies, as goal emulation can only produce non-imitative decisions when the mentor and the learner have different options to choose from. Each action leads to a transition to one of three possible end states $s \in \{s_1, s_2, s_3\}$ according to a probabilistic transition function $T(s|s_0, a)$. During each trial, only one end state s_v is valuable and earns points. Which state is valuable changes periodically during the experiment. The mentor has full knowledge of the valuable state and the transition structure of the environment and always chooses the optimal action. Participants are given direct access to the transition structure of the environment, but not to the valuable state.

To model participants' choices in the task, we adapted three simple (one process) models from Charpentier et al., 2020. We additionally derived mixture (dual process) models that provide a weighted combination of pairs of the simple models. Models 1-3 are models of goal emulation, reinforcement learning imitation, and one-step imitation, respectively. Models 4-7 are mixture models of emulation with each imitation strategy; models 4 and 6 are stakes-independent, and models 5 and 7 are stakes-dependent.

Model 1: Goal Emulation In this model, the participant uses sequential Bayesian updating to infer the current valuable state from the mentors' choices. The probability a state s_i is valuable at timepoint t, after observing an action a performed by the mentor, is given by:

$$P_t(s_i = s_v|a) = \frac{P(a|s_i = s_v) \cdot {P_t}'(s_i = s_v)}{\sum_j P(a|s_j = s_v) \cdot {P_t}'(s_j = s_v)}$$

where the likelihood, $P(a|s_i = s_v)$ is always 0 or 1, as the mentor is always correct. The prior ${P_t}'(s_i = s_v)$ is given by:

$${P_t}'(s_i = s_v) = \lambda \cdot P_{t-1}(s_i = s_v) + (1 - \lambda) \cdot \sum_{j \neq i} \frac{P_{t-1}(s_j = s_v)}{2}$$

where λ is a free parameter capturing the possibility of changes in the valuable fish. Specifically, there is λ probability that the valuable fish remained as in the previous trial, and $1 - \lambda$ probability that a switch has occurred. Action values are computed as the dot product of the vector of the probabilities of each state being valuable ($\overrightarrow{P_t}$) and the vector of the transition structure of the environment for a given action ($\vec{T}(a)$):

$$Q_t^{em}(a) = \overrightarrow{P_t}^{\top} \cdot \vec{T}(a)$$

Finally, the action probability, $P_t^{em}(a)$ is computed by applying a softmax function with a free parameter β to the action values.

Model 2: Reinforcement Learning Imitation In this model, the value of actions is directly updated by observing the mentor's actions, where the value of the chosen action a is positively reinforced, and the value of the unchosen action $\bar{a}$ is negatively reinforced, using a learning rate parameter α as follows:

$$Q_t^{im}(a) = Q_{t-1}^{im}(a) + \alpha \cdot \left(1 - Q_{t-1}^{im}(a)\right)$$

$$Q_t^{im}(\bar{a}) = Q_{t-1}^{im}(\bar{a}) + \alpha \cdot \left(-1 - Q_{t-1}^{im}(\bar{a})\right)$$

Like model 1, the decision rule determines action probabilities $P_t^{im}(a)$ by applying a softmax function with a free parameter β to the action values.

Model 3: One-step Imitation This is a simple imitation model in which the action chosen by the mentor in the previous trial is assigned a value of 1, while all other actions are assigned a value of 0. The decision rule applies a softmax function with a free parameter β to these action values. Thus, when the action that was chosen in the previous trial is unavailable in the current trial, the decision between the available actions is random.

Mixture Models In order to estimate the degree of goal-emulation relative to policy imitation on high-and low-stakes trials, we additionally created four mixture models inspired by dual-systems RL models (Daw et al., 2011; Kool et al., 2016, 2017). **Model 4** is a stakes-independent mixture model that determines the relative contribution of goal emulation over policy imitation to the eventual participants' choice as follows:

$$P_t(a) = \omega \cdot P_t^{em}(a) + (1 - \omega) \cdot P_t^{im}(a)$$

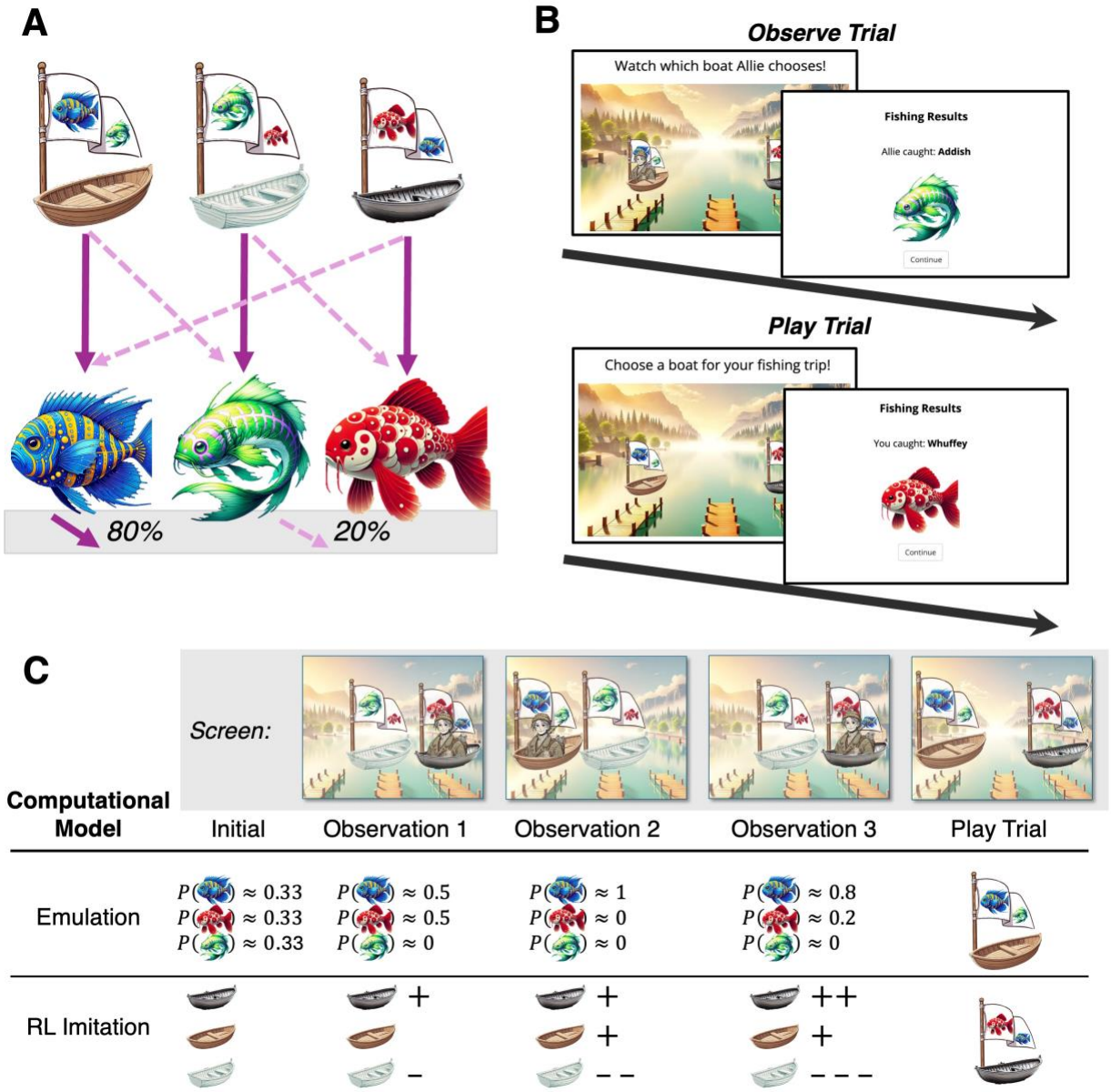

Figure 1: Observational learning fishing task. **A)** Transition structure from boats to fish. **B)** An example of a trial sequence, in play trials and observe trials. In both cases, a boat choice was followed by information about the fish caught. **C)** An example of a critical sequence: the boat that typically catches the valuable fish is unavailable during the first and third trials. In the emulation model, the probability of each fish being valuable is sequentially updated based on the mentor's choices and the context of the available boats (the values shown in the figure correspond to a λ of 0.8, but the principle applies to all λ above 0.5). In RL imitation, the value of the chosen boat is positively reinforced, and the value of the available unchosen boat is negatively reinforced. These mechanisms result in different behavioral predictions for RL imitation and emulation.

where $P_t^{em}(a)$ and $P_t^{im}(a)$ are the outputs of the goal emulation process (model 1) and RL imitation process (model 2), respectively, and the weighting parameter, ω , determines the relative weight of emulation over imitation in the eventual decision. **Model 5** is similar, but it allows weights to vary between stakes, so that there are two weighting parameters ω_{high} for high-stakes blocks and ω_{low} for low-stakes blocks. **Model 6** is a stakes-independent mixture of goal emulation (model 1) and one-step imitation (model 3). **Model 7** is a mixture of goal emulation and one-step imitation that allows weights to vary by stakes. Note that in all mixture models, we only fit one inverse-temperature (β) for both imitation and emulation instead of separated βs.

Simulations

We conducted simulations of the single-process imitation and emulation models (models 1-3) with two main objectives. First, we aimed to identify scenarios where the emulation and imitation strategies produce similar outcomes, and those where goal emulation offers higher performance than policy imitation. This approach builds on the principle that a resource-rational agent will adopt a costly strategy only when it enhances expected outcomes (Kool et al., 2016, 2017). Second, we designed the task to include critical sequences, enabling us to generate diagnostic and falsifiable predictions about participants' behavior, rather than relying exclusively

on model fitting (Palminteri et al., 2017). Our goal was to confirm that, across various parameter sets, these diagnostic critical sequences yield distinct predictions for imitation and emulation.

When Does Goal Emulation Pay Off? We systematically manipulated two important features of the task: volatility, i.e., the rate of shifts in the identity of the valuable state, and uncertainty, i.e., the extent to which the mappings between actions to states is stochastic. Charpentier et al. (2020) showed that these two factors affect arbitration between policy imitation and goal emulation.

For each combination of volatility, uncertainty, and computational model parameters (λ in model 1, α in model 2), we ran 250 simulations of a 150-trial study, and extracted the reward difference between goal emulation and policy imitation. Figure 2 shows the relationship between volatility and reward difference for different uncertainty vectors. This result is aggregated over all λ and α values, and the pattern of results remains the same when only looking at optimal parameters. When contrasting emulation with RL imitation, we identified an inverted U-shape relationship between volatility and reward difference, so that in extremely high and low volatility, the reward difference is low. This arises because, in very stable environments, the action values learned through RL imitation have time to converge and accurately reflect the reward probabilities. Thus emulation and imitation agents make the same (correct) decisions. However when the environment is too volatile, neither social learning strategy is reliable. Thus, emulation significantly outperformed RL imitation only at moderate volatility values, ranging between 0.1 and 0.2 (i.e., rewards structure changes every 5 to 10 days).

When contrasting emulation with one-step imitation, there was a monotonic negative relationship between volatility and reward difference. In both cases, there was also a monotonic relationship between uncertainty and reward difference, so that the advantage of emulation increased as the environment became more deterministic.

Overall, we can conclude that goal emulation is advantageous over policy imitation in deterministic environments that are moderately volatile.

Diagnostic Critical Sequences We introduced a critical sequence at the beginning of each block. This sequence distinguishes between goal emulation and policy imitation via three observation trials followed by a diagnostic play trial (see Figure 1 C). First, participants observe a mentor that is choosing an action with a 20% chance of transitioning to the valuable state when only actions with 20% and 0% chances are available. Next, participants observe the mentor choosing an action with an 80% chance of yielding the valuable state when actions with 80% and 0% chances are available. In the following trial, participants observe the mentor again choosing an action with a 20% chance of reaching the valuable state when only actions with 20% and 0% chances are available. In the subsequent diagnostic play trial, participants choose between actions with 20% and 80% chances of transitioning to the valuable state.

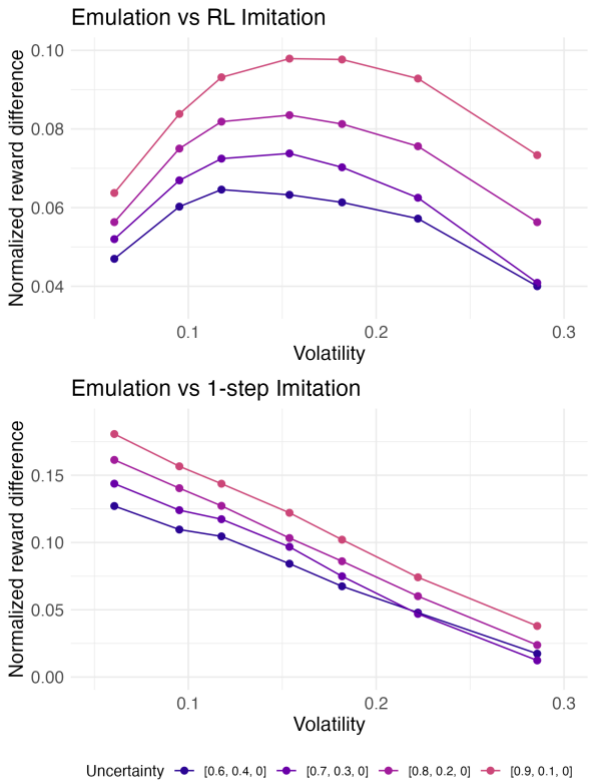

Figure 2: Relationship between volatility (x-axis) and uncertainty (colors) to the reward difference between emulation and RL imitation (top panel), or the reward difference between emulation and one-step imitation (bottom panel).

Simulations of critical sequences confirmed that the goal emulation model predicts choosing the 80% action, and both the imitation models predict choosing the 20% action. This is true for all possible values of α, and for all values of λ within the reasonable $0.5 < \lambda < 1$ range[1].

Experiment

Methods

Participants Sample size, as well as all other data processing details and the main analysis, was pre-registered: https://aspredicted.org/qyy3-vvfd.pdf. Three hundred US-based Prolific participants (39% female, $\text{Range}_{\text{age}}$= 18-45, M_{age} = 31.61, SD_{age} = 7.48) completed the task for $4.25 plus a task performance bonus up to $3.5. Sample size was determined via power analysis that was based on a pilot study we ran, which indicated that at a sample size of 240

[1]The practical implication of $\lambda < 0.5$ is that the agent is consistently more confident in the opposite of their priors than in the priors themselves, making such values implausible for a cognitive model.

participants, we have power of 90% to detect a significant effect of stakes on choice in the critical trial. The final sample size was determined to reach this number of participants after applying our exclusion criteria.

Based on the pre-registered exclusion criteria, the data of 25 participants who achieved very low accuracy (accuracy < 0.45) throughout the task, suggesting random responses, and the data of an additional 44 participants who failed the comprehension checks more than twice were excluded from the analysis.[2] Data from an additional five participants who completed the study were not recorded due to a server malfunction, leaving a final sample of 226 participants.

Materials and Procedure The experimental task was loosely based on Charpentier et al. (2020) and was programmed using jsPsych (de Leeuw, 2015). This two-step fishing game involves three types of boats used to catch three types of fish. Participants are informed that each fishing boat travels to a specific part of the lake, where it predominantly catches one type of fish and occasionally another type. In each trial, only one type of fish is valuable and earns points, with the valuable fish changing periodically. The rate of change (every 6–7 days) is set to maximize the advantage of emulation over imitation based on simulation results. Participants are informed that the valuable fish changes periodically, but were not told the specific rate of change or the valuable fish identity in any given trial. In each trial, two boats are available. In ~65% of trials, participants observe a virtual fishing mentor's choice ("observe" trials), while in ~35%, they select a boat themselves ("play" trials). Participants are informed that the mentor always knows the valuable fish and selects the best available boat to catch it. Participants do not receive immediate feedback on their earnings and must rely on observational learning, with feedback provided only at the end of each block.

The study included 8 experimental blocks, each containing of 17 trials, resulting in 136 trials overall. The first four trials of each block were a critical sequence, carefully structured to differentiate between imitation and emulation (see simulations section and Figure 1C).

The stakes manipulation was framed by marking half of the experimental blocks as "festival days". During festival days, one can earn 10 times the reward for catching the valuable fish compared to routine days. The condition of the first block is randomly determined, and then the task alternates between high and low stakes. We have taken several steps to emphasize the stakes manipulation: First, we added a comprehension question about the value of fish during festival days. Second, the information screen about festival days featured a salient visual representation of a festival, and decision screens in these blocks were highlighted with a pink frame and festival icons. Third, the end-of-block feedback included both the number of coins collected and their equivalent monetary reward in USD, emphasizing the differential stakes for blocks.

Before the task, participants completed practice trials, including observation and play trials where the valuable fish was revealed, to familiarize them with the task structure. They then answered three comprehension questions.

Model Fitting and Parameter Estimation

We applied hierarchical Bayesian analysis, modeling individual parameters as samples from a population-level Gaussian distribution. Group-level parameters were estimated using an Expectation-Maximization algorithm with Laplace approximation to minimize negative log-likelihood. Model comparisons were conducted using integrated Bayesian Information Criteria (iBIC), which accounts for group-level fit by integrating over individual parameters (Charpentier et al., 2020; Huys et al., 2011).

Results

Model Fitting Results

The results of fitting all 7 models to the participants' data are displayed in table 1. Among the single process models, participants' behavior in the task was better explained by a goal emulation model (model 1) than by any of the policy imitation models (2-3). The mixture models of one-step imitation and goal emulation (models 6-7) outperformed all other models, including single process models and the mixture of RL-imitation and emulation, in predicting participants' choices. This result holds both with and without penalization for adding parameters (i.e., when looking at NLL or iBIC). These results are consistent with the results obtained by Charpentier et al. (2020) that showed support for arbitration between emulation and simple, 1 step imitation.

Contrary to our hypothesis, we found that, when penalizing for adding parameters, there is no advantage for stakes-dependent over stakes-independent mixture models, as iBICs for the stakes-independent mixture models (models 4, 6) are lower than iBICs for stake-dependent mixture models (models 5,7). In addition, parameter estimates of weights did not differ significantly between high and low stakes in the stake-dependent models (Model 5: $\omega_{low} = 0.64 \pm 0.11, \omega_{high} = 0.65 \pm 0.11$, Model 7: $\omega_{low} = 0.64 \pm 0.13, \omega_{high} = 0.63 \pm 0.14$; see Figure 3A).

Table 1: Model Fitting Results.

Model	Parameters	Negative-log-likelihood	iBIC
1	λ, β	7194.28	15101.56
2	α, β	7568.10	15961.37
3	β	8182.05	16671.02
4	$\alpha, \lambda, \beta, \omega$	7049.58	15106.47
5	$\alpha, \lambda, \beta, \omega_{high}, \omega_l$	7043.22	15209.33
6	$\boldsymbol{\lambda, \beta, \omega}$	**6830.86**	**14670.31**
7	$\lambda, \beta, \omega_{high}, \omega_{low}$	6808.98	14795.21

[2] The pre-registered exclusion criteria ended up excluding more participants than planned. The pattern of results remains identical when softening the exclusion criteria, or when analyzing all the data without exclusions, further details can be found in the GitHub repository: https://github.com/arielevy8/obs-learning-fishing.

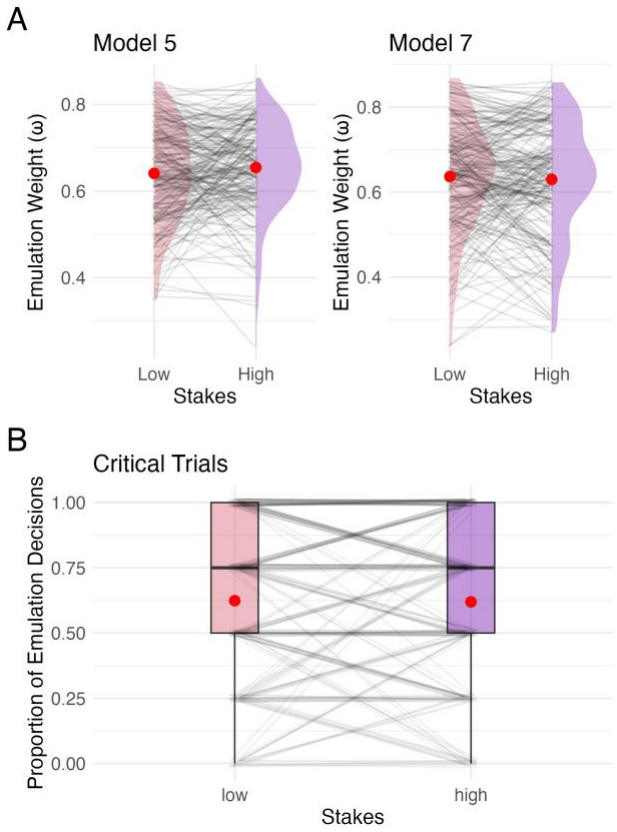

Figure 3: **A)** Individual parameter estimates for stakes-dependent parameters in models 5 and 7. Each line is one participant. **B)** Box plot of proportion of emulation-based decisions in critical trials, for low vs. high stakes.

Analysis of Critical Trials

Overall, behavior during critical trials was more consistent with emulation (mean proportion = 0.62, median = 0.75) than with imitation. To test our hypothesis—that participants will exhibit more emulation-based behavior in high-stakes situations—we estimated a logistic mixed model with an indicator for emulation-based behavior as the dependent variable and stakes as the sole predictor, along with random intercepts and slopes for participants. Analysis was performed using the lme4 package (Bates et al., 2014). Contrary to our hypothesis, we did not observe a statistically significant effect of stakes: the odds of making an emulation-based choice in high-stakes versus low-stakes blocks were not significantly different, $\beta = -0.02$, $SE = 0.12$, $z = -0.20$, $p = .842$ (see figure 3B).

Discussion

In this study, we explored how individuals arbitrate between social learning strategies and whether this process involves cost-benefit considerations—specifically, whether they choose the more cognitively demanding strategy only when it is expected to pay off. To address these questions, we used a model comparison approach and a critical trials approach, backed up by simulations. We found that participants predominantly employed goal emulation, which we initially considered to be more cognitively costly: The single-process model of emulation outperformed imitation models, the emulation weight parameter in all mixture models was significantly higher than 0.5, and participants made emulation-based decisions in diagnostic trials significantly more frequently than imitation-based decisions. Out of the models that were tested, a mixture of goal emulation and one-step imitation best explains our data. These results are consistent with the possibility that participants primarily rely on goal emulation, attempting to infer and replicate the mentor's intentions, but sometimes resort back to a simple, memoryless approach of one-step imitation—either due to lapses in attention or an inability to understand the mentor's behavior. Contrary to our hypothesis, we found no evidence for cost-benefit arbitration between these strategies. Specifically, the probability of choosing an emulation-based option in diagnostic trials was not affected by stakes, and stakes-independent models provided a better fit to the data than stakes-dependent models.

Charpentier et al. (2020) proposed an arbitration model between imitation and emulation based solely on the reliability of the strategies, in which participants follow simple imitation if they cannot identify a reliable goal-based explanation for the mentor's behavior. The goal of the present work was to anchor this preference within a broader framework of resource-rational meta-control, where the relative reliability of goal emulation only matters when its added value outweighs its cost. We did not find empirical support for this possibility. One potential reason, though in our view a less plausible one, is that cognitive meta-control only applies to individual learning strategies (Kool et al., 2017) but not to social learning. This is somewhat consistent with the finding of Charpentier et al. (2024) that reward magnitude manipulations do not affect the balance between individual and social learning.

An alternative, plausible explanation for our results is the differences between classic reinforcement learning tasks and the observational learning task used in this study. Model-based control is more costly than model-free control because it requires constructing a detailed causal model of the environment, and then using that model to plan what to do. However, in the present experiment, the transition structure is simple and transparent to the participants, leaving only the reward structure of the environment to be learned from the choices of others. We modeled this learning process as implemented by Bayesian inference. Considering that people may engage in simple mental state inferences without exerting effort (Schneider et al., 2017), it is plausible that inferring the mentor's desires from their choices in our task is not sufficiently costly to require arbitration via a meta-control mechanism. Future work can use a two-step task that makes inference of the mentor's goal more cognitively demanding (for example, by introducing a more complex reward structure), or make planning according to this goal more complex (for example, by making the transition structure less obvious). Such conditions might make goal emulation too complex to implement when the stakes are low, enhancing the effectiveness of the stakes manipulation.

Acknowledgments

We thank all members of the Laboratory for Social Cognitive Science in the Harvard Department of Psychology for their insightful feedback throughout the development of this project. XRG is supported by the Department of Defense through the National Defense Science and Engineering Graduate Fellowship Program. This work was supported by award number N00014-22-1-2205 to FC from the Office of Naval Research.

References

Baker, C. L., Saxe, R., & Tenenbaum, J. B. (2009). Action understanding as inverse planning. Cognition, 113(3), 329–349. https://doi.org/10.1016/j.cognition.2009.07.005

Call, J., Carpenter, M., & Tomasello, M. (2005). Copying results and copying actions in the process of social learning: Chimpanzees (Pan troglodytes) and human children (Homo sapiens). Animal Cognition, 8(3), 151–163. https://doi.org/10.1007/s10071-004-0237-8

Charpentier, C. J., Iigaya, K., & O'Doherty, J. P. (2020). A Neuro-computational Account of Arbitration between Choice Imitation and Goal Emulation during Human Observational Learning. Neuron, 106(4), 687-699.e7. https://doi.org/10.1016/j.neuron.2020.02.028

Charpentier, C. J., Wu, Q., Min, S., Ding, W., Cockburn, J., & O'Doherty, J. P. (2024). Heterogeneity in strategy use during arbitration between experiential and observational learning. Nature Communications, 15(1), 4436. https://doi.org/10.1038/s41467-024-48548-y

Daw, N. D., Gershman, S. J., Seymour, B., Dayan, P., & Dolan, R. J. (2011). Model-Based Influences on Humans' Choices and Striatal Prediction Errors. Neuron, 69(6), 1204–1215. https://doi.org/10.1016/j.neuron.2011.02.027

Daw, N. D., Niv, Y., & Dayan, P. (2005). Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. Nature Neuroscience, 8(12), Article 12. https://doi.org/10.1038/nn1560

de Leeuw, J. R. (2015). jsPsych: A JavaScript library for creating behavioral experiments in a Web browser. Behavior Research Methods, 47(1), 1–12. https://doi.org/10.3758/s13428-014-0458-y

Gershman, S. J., Horvitz, E. J., & Tenenbaum, J. B. (2015). Computational rationality: A converging paradigm for intelligence in brains, minds, and machines. Science, 349(6245), 273–278. https://doi.org/10.1126/science.aac6076

Huys, Q. J. M., Cools, R., Gölzer, M., Friedel, E., Heinz, A., Dolan, R. J., & Dayan, P. (2011). Disentangling the Roles of Approach, Activation and Valence in Instrumental and Pavlovian Responding. PLoS Computational Biology, 7(4), e1002028. https://doi.org/10.1371/journal.pcbi.1002028

Jara-Ettinger, J., Gweon, H., Schulz, L. E., & Tenenbaum, J. B. (2016). The Naïve Utility Calculus: Computational Principles Underlying Commonsense Psychology. Trends in Cognitive Sciences, 20(8), 589–604. https://doi.org/10.1016/j.tics.2016.05.011

Kleiman, T., Meiran, N., & Eyal, T. (2022). Perspectives, they might be a-changin': A proactive-control take on the cognitive cost of maintaining one's own perspective. Journal of Experimental Psychology: General, 151(6), 1473.

Kool, W., Cushman, F. A., & Gershman, S. J. (2016). When Does Model-Based Control Pay Off? PLOS Computational Biology, 12(8), e1005090. https://doi.org/10.1371/journal.pcbi.1005090

Kool, W., Gershman, S. J., & Cushman, F. A. (2017). Cost-Benefit Arbitration Between Multiple Reinforcement-Learning Systems. Psychological Science, 28(9), 1321–1333. https://doi.org/10.1177/0956797617708288

Kulke, L., Von Duhn, B., Schneider, D., & Rakoczy, H. (2018). Is Implicit Theory of Mind a Real and Robust Phenomenon? Results From a Systematic Replication Study. Psychological Science, 29(6), 888–900. https://doi.org/10.1177/0956797617747090

Lieder, F., & Griffiths, T. L. (2020). Resource-rational analysis: Understanding human cognition as the optimal use of limited computational resources. Behavioral and Brain Sciences, 43, e1. https://doi.org/10.1017/S0140525X1900061X

Nagell, K., Olguin, R. S., & Tomasello, M. (1993). Processes of social learning in the tool use of chimpanzees (Pan troglodytes) and human children (Homo sapiens). Journal of Comparative Psychology, 107(2), 174–186. https://doi.org/10.1037/0735-7036.107.2.174

Palminteri, S., Wyart, V., & Koechlin, E. (2017). The Importance of Falsification in Computational Cognitive Modeling. Trends in Cognitive Sciences, 21(6), 425–433. https://doi.org/10.1016/j.tics.2017.03.011

Roberts-Gaal, X., & Cushman, F. (2023). Computational principles underlying the evolution of cultural learning mechanisms. Proceedings of the Annual Meeting of the Cognitive Science Society, 45(45). https://escholarship.org/uc/item/78c4w10j

Samson, D., Apperly, I. A., Braithwaite, J. J., Andrews, B. J., & Bodley Scott, S. E. (2010). Seeing it their way: Evidence for rapid and involuntary computation of what other people see. Journal of Experimental Psychology: Human Perception and Performance, 36(5), 1255.

Schneider, D., Slaughter, V. P., & Dux, P. E. (2017). Current evidence for automatic Theory of Mind processing in adults. Cognition, 162, 27–31. https://doi.org/10.1016/j.cognition.2017.01.018

Shafto, P., Goodman, N. D., & Griffiths, T. L. (2014). A rational account of pedagogical reasoning: Teaching by, and learning from, examples. Cognitive Psychology, 71, 55–89. https://doi.org/10.1016/j.cogpsych.2013.12.004

Wu, C. M., Vélez, N., Cushman, F. A., Dezza, I. C., Schulz, E., & Wu, C. M. (2022). Representational exchange in human social learning. The Drive for Knowledge, 169–192.