Playing to win or playing to learn? Human performance in a social card game task

Alexandra Witt* **(alexandra.witt@gmx.net)**
University of Tübingen, Maria-von-Linden-Str. 6
Tübingen, Baden-Württemberg 72076 Germany

Joel Vasama* **(joel.vasama@student.uni-tuebingen.de)**
University of Tübingen, Maria-von-Linden-Str. 6
Tübingen, Baden-Württemberg 72076 Germany

Natalia Vélez (nvelez@fas.harvard.edu)
Harvard University, 52 Oxford Street
Cambridge, MA 02138 USA

Charley M. Wu (charley.wu@uni-tuebingen.de)
University of Tübingen, Maria-von-Linden-Str. 6
Tübingen, Baden-Württemberg 72076 Germany

* These authors contributed equally

Abstract

Humans use a variety of cognitive capacities and strategies to learn from others, ranging from faithfully imitating other people's behavior to inferring the mental states that produced the behavior. Prior work has suggested people flexibly arbitrate between imitating others' choices (Imitation; I) and inferring the value of their chosen option (Value Inference; VI). However, it remains an open question how people balance these strategies against drawing ever richer social inferences about the structure of the environment (Model-Based Inference; MBI). Using a task designed to dissociate the three strategies, we find evidence for the adaptive use of imitation, as well as preliminary evidence for MBI-level performance. Our results provide a methodological framework to understand how humans learn from others, with future work using computational modeling of choices expected to provide important insights into the arbitration of social learning mechanisms.

Keywords: Social Learning, Arbitration, Transitive Inference, Imitation, Theory-of-Mind

Introduction

Humans learn from others in a variety of ways, including faithfully imitating actions (Tennie, Call, & Tomasello, 2009), inferring what others value (Collette, Pauli, Bossaerts, & O'Doherty, 2017), and detecting how their beliefs about the world may differ from our own (Berke & Jara-Ettinger, 2022). With this range of strategies at our disposal, an important question is how we decide which to employ in which situation.

Past work suggests that when we learn from others, we take the most reliable strategy (Charpentier, Iigaya, & O'Doherty, 2020). In a three-armed bandit task, participants flexibly alternated between simply copying the demonstrator's action (Imitation; I) or inferring which token was most valuable (here, we call this strategy Value Inference; VI).

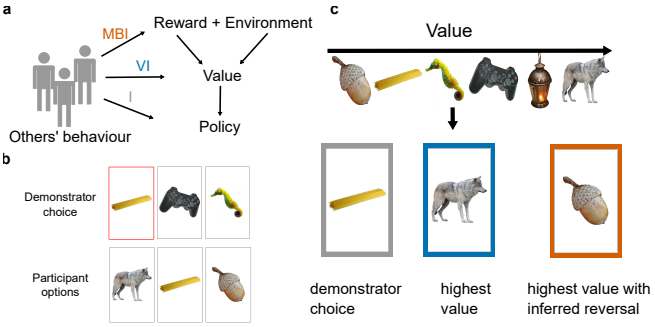

Figure 1: Social learning mechanisms and Experiment. **a**) Three levels of social inference: imitation (I) copies the policy, value inference (VI) infers the value function, and model-based inference (MBI) infers the demonstrator's reward function and beliefs about the environment (e.g., whether a reversal has occured) from behavior. **b**) The social transitive inference task (STIT). Participants observe an omniscient demonstrator's choice, then make their own. **c**) At some points during the task, the reward hierarchy is reversed. In those instances, all three models can predict different choices.

Collectively, these strategies trade off the *cost* of performing the computation against the *flexibility* of its outputs (Wu, Vélez, & Cushman, 2022, Fig. 1a). VI may be more demanding than copying the demonstrator, but it allows greater flexibility to generalize (e.g., when choices available to the demonstrator are different from those available to you).

However, beyond inferring what others find valuable, humans are also capable of inferring other people's beliefs about the structure of the world (Baker, Jara-Ettinger, Saxe, & Tenenbaum, 2017). This capacity has variously been called Theory of Mind, or mentalizing. By analogy to asocial learning mechanisms, we refer to it as Model-Based Inference (MBI). Developmental work suggests that VI and MBI may arise at distinct points in development (Gergely & Csibra, 2003). For example, toddlers understand that people can have personal preferences (Repacholi & Gopnik, 1997) years before they understand that other people can have different beliefs (Scott &

Baillargeon, 2017). Yet, it remains unknown whether humans also arbitrate between MBI and other social learning mechanisms. Here, we tackle this question using a transitive inference task designed to dissociate the three strategies.

Experiment: Social transitive inference task

We designed the Social Transitive Inference Task (STIT; Fig. 1b) where participants could learn an underlying value hierarchy from an omniscient demonstrator's choices. Sometimes, this value hierarchy was reversed, which could be inferred from the demonstrator's behavior. Thus, in trials after the reversal, each strategy can make distinct predictions (Fig. 1c), allowing us to distinguish between strategies.

Participants and design. We recruited 71 participants ($M_{age} = 39.7$; $SD = 12.0$; 34 female, 2 other) on Prolific in a within-subject design, where we manipulated the set size $s \in [6,9,12]$ of cards in each game to alter the cognitive costs of more complex strategies. On average, participants spent 31.8 $\pm$ 16.6 minutes and earned £6.0 $\pm$ £0.70.

Materials and procedure. Participants were instructed to earn as many points as possible by selecting the best out of three randomly drawn cards (Fig. 1b). Each choice was preceded by a selection made by an omniscient demonstrator. The demonstrator's chosen card was always available to the participant to allow for exact imitation, although the other two alternatives could differ, making imitation suboptimal in some cases. Additionally, participants were told the value ranking of cards would be reversed once during each round, and to pay close attention to the demonstrator's choices to identify when this occurred.

After the instructions, an interactive tutorial, and a comprehension check, participants began the main experiment. The task was performed over 6 rounds, with two repetitions of each set size $s \in 6,9,12$, and in randomized order. The number of trials in a round was defined relative to the set size ($T = s * 5$), to ensure enough time to learn the underlying reward hierarchy. Each set size had one early-switch (after 40% of the trials) and one late switch (60% of the trials) block. After each block, participants received feedback on the percentage of correct choices and the corresponding bonus, and were shown the unreversed card hierarchy. Card images were obtained from the BOSS inventory (Brodeur, Guérard, & Bouras, 2014), with an equal number of natural and artificial stimuli in each block.

Social learning models

To perform **Imitation** (I), we assume participants simply copy the demonstrator's choice ($x_i = x_d$) in a slightly noisy fashion. We implement this as a softmax policy:

$$\pi_I(x_i) \propto \exp(\delta(x_i = x_d)/\tau) \tag{1}$$

where $\delta(x_i = x_d) = 1$ for the demonstrator's choice and 0 otherwise, and the temperature τ controls decision noise.

Value Inference (VI) can be described as Inverse Reinforcement Learning (IRL; Jara-Ettinger, Gweon, Schulz, & Tenenbaum, 2016) using Bayesian formalism to infer another individual's value function V from their actions A:

$$P(V|A) \propto P(A|V)P(V) \tag{2}$$

We implement VI as a particle filter (Speekenbrink, 2016), where each of the m particles $z_m^t = \{v_1^t, \ldots, v_s^t\}$ represents a hypothesis about the value V of each card. We initialize $v_i^0 \sim \mathcal{N}(0,1)$ as a draw from a normal distribution. After each observation, particles are re-weighted and resampled proportional to their likelihood $P(A|V)$, assumed to be a softmax policy (similar to Eq. 1 using v_i instead of $\delta(x_i = x_d)$) predicting the demonstrator's choice. Thus, value estimates consistent with demonstrator choices are more likely to be resampled and proliferate. Because the static value function V is prone to sample impoverishment , we rejuvenate particles by adding Gaussian noise $\tilde{z}_m^t \sim \mathcal{N}(z_m^t, 1)$. We then use Metropolis-Hastings (Murphy, 2012) to probabilistically accept the rejuvenated particles with $p(\text{accept}) = \min\left(1, \frac{p(A|\tilde{z}_m^t)p(z_m^t|\tilde{z}_m^t)}{p(A|z_m^t)p(\tilde{z}_m^t|z_m^t)}\right)$. The posterior distribution in Eq. 2 is then approximated as the average over particles, which is used in a softmax policy to model participant choices $\pi_{VI}(x_i) \propto \exp(v_i/\tau)$.

We assume **Model-based Inference** (MBI) decomposes the demonstrator's value function into their knowledge about the reward rankings R and their belief about the environmental state B (i.e., whether the hierarchy was reversed or not):

$$P(R,B|A) \propto P(A|R,B)P(B)P(R) \tag{3}$$

MBI is also implemented as a particle filter, but with particles including both R and $B \in [-1;1]$. Reward rankings R are initialized, resampled, and rejuvenated the same way as value V (see above). The belief component B is initialized as 1 (no reversal) and has an asymmetric transition probability of flipping from 1 to -1 with $P(B_{t+1} = -1|B_t = 1) = \frac{1}{\text{T-t}}$ to indicate an inferred reversal. As with VI, the posterior distribution in Eq. 3 is approximated as the average across all particles, which is then used to inform the model choices with a softmax policy $\pi_{MBI}(x_i) \propto \exp(r_i * B/\tau)$.

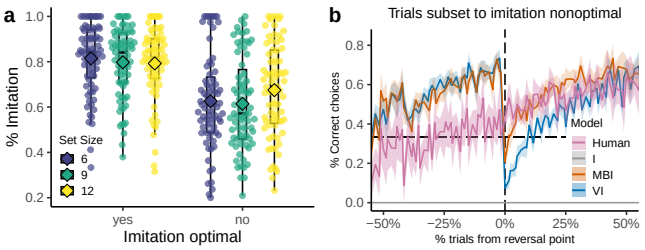

Figure 2: Behavioural results. **a**) Percentage of imitation in trials where imitation is vs. is not the optimal, split by set size. Participants imitate less when it is not optimal, but more when set sizes are higher. **b**) Learning curves (ribbons indicate 95%-CI) for simulated models and participant data centered around the reversal point (dashed vertical line). The dashed horizontal line is chance performance.

Results

Participants chose the best card above chance (33%; $t(70) = 18.0$, $p < .001$, $M_{\text{correct}} = 56.9\%$, $SD = 11\%$), but at a slightly worse rate than pure imitation ($t(70) = -2.0$, $p = .048$). Overall choice accuracy did not differ based on set size ($F_{1,70} = 2.917$, $p = .092$).

We find preliminary evidence for adaptive imitation, with participants imitating more when it was the optimal strategy ($F_{1,70} = 127.2$, $p < .001$; Fig. 2a). Although the rate of imitation was not influenced by set size ($F_{1,70} = 0.799$, $p = .374$), we find an interaction between set size and optimal imitation ($F_{1,70} = 7.813$, $p = .007$), with participants imitating more when suboptimal in larger set sizes. This may suggest an increased use of simpler strategies for more complex problems, although VI and MBI also predict imitation when it is optimal.

To find evidence for VI or MBI, we ran 333 simulations per model with the number of particles set to $m = 500$ and softmax temperatures τ sampled from a log-normal prior with $M \approx 0.004$ and $SD \approx 0.005$. We remove trials where imitation is optimal, and plot human and model performance relative to the reversal point (Fig. 2b), reporting aggregated learning curves due to a lack of differences across set sizes. After the reversal point, MBI recovers faster than VI, with participant performance most resembling MBI ($t(70) = -0.9$, $p = .389$) and reliably better than VI ($t(70) = 5.1$, $p < .001$). Note that model performance can be arbitrarily degraded by reducing the number of particles, and future work fitting models to choices can provide stronger evidence for strategy use.

Conclusion

Humans employ a variety of social learning mechanisms. Using a novel task to tease apart these mechanisms, we find evidence of the adaptive use of imitation and MBI-level performance after reversals. This could be underpinned by either an arbitration mechanism, or MBI implemented in a resource-limited fashion, with model fitting required to disentangle these hypotheses. Our results add to a growing body of work on the sophistication of human social learning (Witt, Toyokawa, Lala, Gaissmaier, & Wu, in press; Charpentier et al., 2020; Vélez & Gweon, 2021), and lays the foundations for future investigations into arbitration between social learning mechanisms.

Acknowledgements

This work is supported by the German Federal Ministry of Education and Research (BMBF): Tübingen AI Center, FKZ: 01IS18039A and funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy–EXC2064/1–390727645. The authors thank the International Max Planck Research School for Intelligent Systems (IMPRS-IS) for supporting Alexandra Witt.

References

Baker, C. L., Jara-Ettinger, J., Saxe, R., & Tenenbaum, J. B. (2017). Rational quantitative attribution of beliefs, desires and percepts in human mentalizing. *Nature Human Behaviour*, *1*(4), 0064.

Berke, M., & Jara-Ettinger, J. (2022). Integrating experience into bayesian theory of mind. In *Proceedings of the annual meeting of the cognitive science society* (Vol. 44).

Brodeur, M. B., Guérard, K., & Bouras, M. (2014). Bank of standardized stimuli (boss) phase ii: 930 new normative photos. *PloS one*, *9*(9), e106953.

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.

Collette, S., Pauli, W. M., Bossaerts, P., & O'Doherty, J. (2017). Neural computations underlying inverse reinforcement learning in the human brain. *Elife*, *6*, e29718.

Gergely, G., & Csibra, G. (2003). Teleological reasoning in infancy: The naıve theory of rational action. *Trends in cognitive sciences*, *7*(7), 287–292.

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.

Murphy, K. P. (2012). *Machine learning: a probabilistic perspective*. MIT press.

Repacholi, B. M., & Gopnik, A. (1997). Early reasoning about desires: evidence from 14-and 18-month-olds. *Developmental psychology*, *33*(1), 12.

Scott, R. M., & Baillargeon, R. (2017). Early false-belief understanding. *Trends in Cognitive Sciences*, *21*(4), 237–249.

Speekenbrink, M. (2016). A tutorial on particle filters. *Journal of Mathematical Psychology*, *73*, 140–152.

Tennie, C., Call, J., & Tomasello, M. (2009). Ratcheting up the ratchet: on the evolution of cumulative culture. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1528), 2405–2415.

Vélez, N., & Gweon, H. (2021). Learning from other minds: An optimistic critique of reinforcement learning models of social learning. *Current opinion in behavioral sciences*, *38*, 110–115.

Witt, A., Toyokawa, W., Lala, K., Gaissmaier, W., & Wu, C. M. (in press). Social learning with a grain of salt. In M. Goldwater, F. Anggoro, B. Hayes, & D. Ong (Eds.), *Proceedings of the 45th Annual Conference of the Cognitive Science Society.* Sydney, Australia: Cognitive Science Society.

Wu, C. M., Vélez, N., & Cushman, F. A. (2022). Representational exchange in human social learning: Balancing efficiency and flexibility. In I. C. Dezza, E. Schulz, & C. M. Wu (Eds.), *The Drive for Knowledge: The Science of Human Information-Seeking.* Cambridge: Cambridge University Press.