Arbitrating Computational Models of Observational Learning

Bryan Gonzalez[1]
Luke J. Chang[1,*]

[1]Department of Psychological & Brain Sciences
Dartmouth College
6207 Moore Hall
Hanover, NH 03755

*Corresponding Author
luke.j.chang@dartmouth.edu
603.646.2056

Abstract

How do we learn in the absence of direct experience? In this paper, we discuss a recent paper by Charpentier et al., 2000 that proposes a new computational account of observational learning, which arbitrates between choice imitation and goal emulation.

Humans have a remarkable ability to learn how to navigate an environment in the absence of direct experience by simply observing others (Olsson et al., 2020). For example, imagine traveling to a foreign country and trying to order food without being able to understand the menu. How would you accomplish this? One strategy is to simply copy others ahead of you in line - a form of *imitation learning*. This will likely result in successfully getting something to eat, but does not ensure that you will enjoy it. An alternative strategy is to instead infer other people's goals and **emulate** the one that is most consistent with your own. This requires the additional computation of inferring a model of another person's mental state (Gonzalez and Chang, In press).

Recently, Charpentier et al., (2020) sought to explore this question to better understand how humans learn from observations. Using a novel behavioral task, participants observed another agent choose between two of three presented slot machines. Each slot machine paid out a token color (e.g., green, red or blue), but only one color could be exchanged for money, which was unknown to the participants. Thus, participants could only learn which machine to pick by observing the other agent's choices. Participants were told that the other agent knew which color yielded money, and that the winning color could change across trials. Participants had information about the token color probabilities for each slot machine, and were able to see the outcome (e.g., token color) after the agent's choice. Showing participants the token color returned by the chosen slot machine, and not its explicit value, required participants to estimate which color was valuable based solely on the agent's actions.

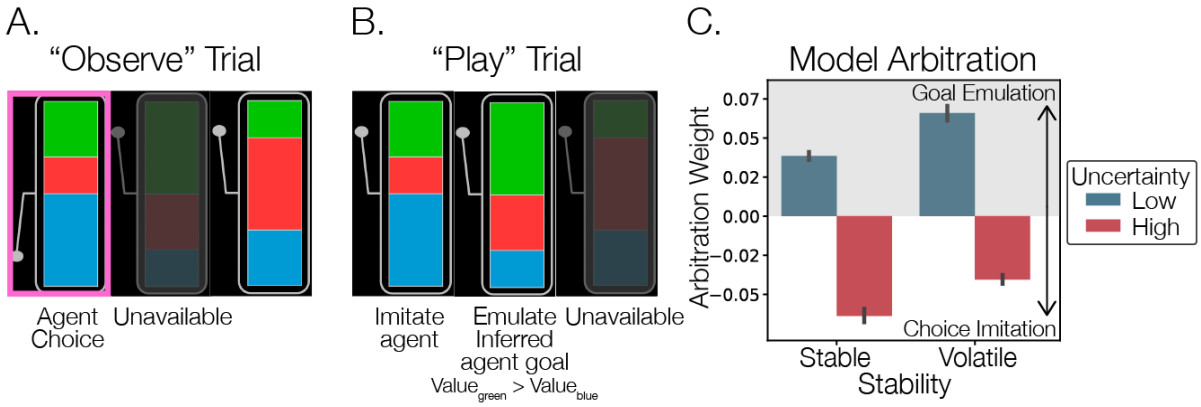

Figure 1. Observational learning task

Suppose the other agent selects the *left* slot machine (Fig 1A). From this example, it is clear that the agent is not interested in the red token, but uncertain if the other agent's goal was to maximize the probability of a blue or green token. During all trials of a given block, the position (i.e., left, middle, right) and probability distributions of each slot machine are fixed, however the unavailable option varies to modulate the difficulty of goal inference. For example, if the rightmost machine in Fig 1A was unavailable to the agent, a left choice would then clearly indicate a goal to obtain the rewarding blue token. In a third of the trials, participants were able to play the game themselves for the potential to earn money (Fig 1B). If the participant selected the middle slot machine after observing Fig1A, then they most likely believed that the other agent's goal was to maximize green rather than blue tokens. Across blocks, the authors manipulated the uncertainty of the outcome probability distributions and also switched which color is associated with a payoff, akin to a hidden reversal.

The authors evaluated support for two different observational learning strategies. The *choice imitation* model simply learned which slot machine (e.g., left, middle, or right) was more frequently chosen by the other agent using reinforcement-learning. The *goal emulation* model, in contrast, attempted to learn the other agent's goal (i.e., which color yielded a payoff) by updating the value of each color via an approximate bayesian updating rule and selecting the machine with the highest overall expected value. In addition, the authors explored models that combined both strategies and incorporated an arbitration control mechanism to determine which strategy should be employed. The proposed arbitration mechanism is similar to previous work comparing nonsocial model-free and model-based reinforcement-learning (Daw et al., 2005). The basic idea is that the arbitration controller uses relative uncertainty to choose which strategy to employ. When a particular strategy can accurately predict the agent's choices, it gets a higher weight, but when the model is "surprised" and starts to become less reliable, the other strategy gets a higher weight. In other words, if the agent's goals become more difficult to infer, then the reliability of a goal emulation model decreases and an imitation model is favored. Conversely if the agent's choices appear to be more stochastic (due to rapidly changing goals), the reliability of an imitation learning model decreases and participants will be more likely to employ a goal emulation model.

Overall, the authors find that computational models employing an arbitration mechanism provided a better account of participant's behavioral data compared to models employing a single strategy. This was supported by directly fitting the models to participant's behavioral data and also simulating the models to demonstrate that both imitation and goal emulation behavior could be generated by the model. This is an important and often overlooked step when attempting to falsify computational models (Palminteri et al., 2017). In addition, the authors found that their 2 x 2 design, crossing certainty in goals with volatility of payoff color, impacted the model's arbitration weight. Low uncertainty conditions, where an agent's choices more clearly indicate its goal, favored the use of emulation models over imitation, while imitation was favored slightly more in stable environments relative to volatile ones. These behavioral results from Study 1, centered within each subject, can be seen in Fig 1C.

The authors explored brain regions that were potentially associated with the arbitration mechanism by correlating signals from their computational model with trial-to-trial fluctuations in BOLD signal. They found that the goal emulation reliability signal significantly correlated with the right anterior insula, while the imitation reliability signal correlated with the medial orbitofrontal cortex. They also calculated the degree of surprise in the goal emulation strategy when participant's observed the outcome of the other agent's actions by calculating the Kullback-Leibler divergence between the prior and posterior values of each color. They found that this surprise signal significantly correlated with regions associated with the salience network including bilateral insula, dorsal anterior cingulate, dorsolateral prefrontal cortex and parietal cortex.

This paper provides a substantial improvement in our understanding of the computations underlying observational learning, particularly in how different types of learning strategies such as choice imitation and goal emulation can be flexibly applied across different learning environments that may vary in uncertainty and volatility. The field of social neuroscience is just

beginning to embrace the use of computational models (Cheong et al., 2017) and this paper provides an important advance in demonstrating how to move beyond basic reinforcement models. Furthermore, this work demonstrates the importance of using carefully controlled experimental designs in developing new models and evaluating their performance across different experimental controls. However, one potential limitation to this approach is whether these models are specific to this particular experiment, or if they will generalize to other observational learning contexts. Looking to the future, we hope the field will begin to embrace the use of naturalistic designs when studying the neurocomputational mechanisms underlying social cognition (Wheatley et al., 2019). Real social interactions reflect non-stationary dynamic processes as people mutually adapt their behavior and may have different types of signals and error structures than will be present in an artificial interaction. Inadequately sampling the psychological phenomena of interest (e.g., goal emulation) with overly constrained experimental designs, will bias researchers to converge on overly simplistic explanatory models, which are unlikely to generalize to real world contexts (Jolly and Chang, 2019).

A notable strength of this paper is the inclusion of an additional replication sample. The authors preregistered their computational models and brain findings based on Study 1 prior to the collection of Study 2. While some of their brain findings replicated, such as the correlations with goal emulation reliability and KL-divergence, unfortunately the correlation with imitation reliability in the OFC did not. It is currently unusual for researchers to include a replication sample in neuroimaging studies due to the large expense in collecting data. This practice is incredibly important for minimizing experimenter bias and overfitting data, particularly in studies that involve lots of experimenter degrees of freedom (e.g., computational models and neuroimaging). However, this replication study also raises new issues. First, how should replication results be reported? Charpentier et al (2020) present so many different analyses (e.g., ROI, whole-brain, different parametric regressors across 10 different computational models) across both study samples, that readers may have a difficult time sorting out what are the key results that they should take away from this work. Should preregistered hypotheses carry more value than ones generated after data collection even if the effects replicate across both studies? What about reviewers' comments on the manuscript? Because these can never be preregistered, should they be demarcated as "exploratory"? Second, should we be more concerned with minimizing false positives or false negatives? Neuroimaging studies are traditionally highly underpowered (Cremers et al., 2017) with meta-analyses estimating an approximate power of 8% (Button et al., 2013). If a single study is underpowered, then filtering results by additional underpowered replications will certainly reduce the likelihood of reporting a spurious finding, but will also dramatically increase the likelihood of missing true effects hidden in the data, which might have emerged if the two samples had been combined to increase the power. Third, replication studies are expensive and funding agencies and early career scientists may choose to prioritize new discoveries rather than confirming old ideas when deciding how to allocate their limited resources. This has the potential to further exacerbate the economic inequality across laboratories where only a limited number of well funded groups can afford to publish cutting edge work that includes independent replications. Though we certainly appreciate the importance of minimizing experimenter bias and false positives, there are many other important issues to consider with the

use of preregistration and replication studies. We look forward to future discussions surrounding these important issues as the field begins to grapple with these quickly changing research norms.

Acknowledgments

The authors would like to thank Emma Templeton for her helpful comments on earlier drafts of this manuscript. L.J.C is funded by the NIH (R01MH116026 & R56 MH080716) and NSF (1848370). The authors declare no competing interests.

References

Button, K.S., Ioannidis, J.P.A., Mokrysz, C., Nosek, B.A., Flint, J., Robinson, E.S.J., and Munafò, M.R. (2013). Power failure: why small sample size undermines the reliability of neuroscience. Nat. Rev. Neurosci. *14*, 365–376.

Charpentier, C.J., Iigaya, K., and O'Doherty, J.P. (2020). A Neuro-computational Account of Arbitration between Choice Imitation and Goal Emulation during Human Observational Learning. Neuron.

Cheong, J.H., Jolly, E., Sul, S., and Chang, L.J. (2017). Computational Models in Social Neuroscience. Computational Models of Brain and Behavior 229–244.

Cremers, H.R., Wager, T.D., and Yarkoni, T. (2017). The relation between statistical power and inference in fMRI. PLoS One *12*, e0184923.

Daw, N.D., Niv, Y., and Dayan, P. (2005). Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. Nat. Neurosci. *8*, 1704–1711.

Gonzalez, B., and Chang, L.J. (In press). Computational models of mentalizing. In Theory of Mind, M. Gilead, and K. Ochsner, eds.

Jolly, E., and Chang, L.J. (2019). The Flatland Fallacy: Moving Beyond Low–Dimensional Thinking. Top. Cogn. Sci.

Olsson, A., Knapska, E., and Lindström, B. (2020). The neural and computational systems of social learning. Nat. Rev. Neurosci. *21*, 197–212.

Palminteri, S., Wyart, V., and Koechlin, E. (2017). The Importance of Falsification in Computational Cognitive Modeling. Trends Cogn. Sci.

Wheatley, T., Boncz, A., Toni, I., and Stolk, A. (2019). Beyond the Isolated Brain: The Promise and Challenge of Interacting Minds. Neuron *103*, 186–188.