Published in final edited form as:
Curr Opin Behav Sci. 2021 April ; 38: 110–115. doi:10.1016/j.cobeha.2021.01.006.

Learning from other minds: An optimistic critique of reinforcement learning models of social learning

Natalia Vélez[1], **Hyowon Gweon**[2]

[1]Department of Psychology, Harvard University

[2]Department of Psychology, Stanford University

Abstract

Reinforcement learning models have been productively applied to identify neural correlates of the value of social information. However, by operationalizing social information as a lean, reward-predictive cue, this literature underestimates the richness of human social learning: Humans readily go beyond action-outcome mappings and can draw flexible inferences from a single observation. We argue that computational models of social learning need *minds*, i.e, a generative model of how others' unobservable mental states cause their observable actions. Recent advances in inferential social learning suggest that even young children learn from others by using an intuitive, generative model of other minds. Bridging developmental, Bayesian, and reinforcement learning perspectives can enrich our understanding of the neural bases of distinctively human social learning.

Keywords

social learning; reinforcement learning; Bayesian modeling; cognitive neuroscience

Introduction

Suppose you see a friend open a cabinet by kicking it. If you considered only superficial contingencies between your friend's action (kicking) and its outcome (successfully opening the cabinet), then you might learn that kicking opens the cabinet. When we learn from others, however, we learn far more than meets the eye. For instance, you might infer that the cabinet is stuck, even though you never saw your friend struggle to open the cabinet; if you can explain away her action (e.g., because her hands were loaded with books), you might infer that kicking is not necessary at all [1]. How is this possible? Beneath the surface of observable actions is a rich, causal structure of unobservable mental states, such as the actor's goals, beliefs, and desires. By reasoning about how and why others' actions came to be, learners can draw powerful inferences that go beyond the data.

Address for correspondence: Hyowon Gweon, Department of Psychology, Stanford University, 450 Jane Stanford Way, Building 420, Stanford, CA 94305, Phone: (650) 498-1192, gweon@stanford.edu.

Declaration of interests
The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Recent advances in probabilistic, or Bayesian, models of social cognition have provided the foundation for formal theories of inferential social learning—that is, of how agents update their beliefs by drawing inferences from evidence generated by others ([••2]; see also [3]). However, despite their success in explaining the power and the flexibility of human social learning, the neural mechanisms that support these inferences are vastly underexplored. On the other hand, reinforcement learning models have been applied to study the neural substrates of social learning, but they have yet to incorporate formal theories of how unobservable mental states give rise to observable actions and outcomes. Below we review what reinforcement learning models have revealed about the neural basis of social learning, what they have missed, and what can be gained by bridging insights from inferential social learning.

Social learning as a reinforcement learning problem

Reinforcement learning (RL) models describe how agents optimize their actions through trial and error by learning a mapping between the environment, the actions available to them within the environment, and the outcomes associated with those actions [4]. Parameters of RL models appear to be directly instantiated in evolutionarily conserved neural circuits [5]. Regions including prefrontal cortex, orbitofrontal cortex, and striatum track the expected value of rewards, the magnitude of the actual reward, and the discrepancy between the two, or prediction error [6]. These neural circuits play a causal role in reward-guided learning [7–9]. Thus, RL provides a powerful framework for studying the brain as an information-processing system: It offers a computational specification of the problem to be solved, a suite of algorithms that describe possible solutions, and a link between algorithms and their physical implementation. Its success in characterizing non-social learning across Marr's levels of analysis [10,11] has naturally led to extensions of this approach to understand how humans learn in social contexts [12,••13].

In the past decade, RL models have been productively applied to study the neural basis of social learning [••13–15]. Existing work has largely emphasized continuities between the neural mechanisms that support individual learning through trial and error, and those that support social learning through repeated interactions with social partners. Neural signals that track rewards and prediction errors during individual learning also vicariously track rewards that others have received [16,17]. These signals can also be directly modulated through instruction; for example, informing participants of the reward probabilities associated with each action dampens striatal reward prediction error signals [18,19]. In addition to learning about rewards in the environment, RL models have been applied to address how humans learn about others. Recent studies suggest that separable neural signals track reward-predictive properties of social information, such as how often an advisor is accurate [20] or how generously a donor will share from their endowment [21].

For all the successes of this approach, however, it is also important to consider what RL models have missed about social learning. The studies above largely treat others' actions and advice as a lean, reward-predictive cue—e.g., the model's estimates of how often an advisor is correct fluctuate from trial to trial, and the task for the learner is to figure out the expected value of the cue. But social information is far more than a reward-predictive cue:

It is *curated by other minds*. What is missing from past work is a causal, generative model of how those actions and advice came to be. By inverting these generative models, learners can gain far more from social information than superficial contingencies between actions and outcomes [22–••24]. These limitations present an opportunity to consider a different approach that complements RL: inferential social learning.

Social learning as probabilistic inference over structured representations

Before we define inferential social learning, we will first place the idea in context. Inferential social learning is part of a larger intellectual project that aims to characterize human intelligence as a powerful inference engine—one that performs probabilistic inferences over structured representations of the world, of other minds, and of the intersection between the two [25,26]. Probabilistic cognitive models have been successfully applied to characterize many aspects of human learning and reasoning, including prediction [27–29], causal judgments [30,31], concept acquisition [32], active hypothesis testing [33], and learning and exploration in early childhood [34,35].

The key idea behind **inferential social learning** is that learners recover the meaning underlying others' actions by inverting an intuitive, causal model of how agents think, plan, and act [••24,36,37]. Key signatures of inferential social learning emerge early in life. Even infants can learn the meanings of words [35], discover hidden object properties [34], and infer others' preferences [38,39] from a handful of examples by considering how those examples were selected by a demonstrator. In addition, preschoolers can draw sophisticated inferences by considering the demonstrator's unobservable mental states, such as their knowledge and communicative intent [40,41]. For example, if a knowledgeable teacher pedagogically demonstrates that a toy plays music, preschoolers not only learn that the toy plays music; they also infer that it is the toy's *only* function and constrain their exploration accordingly [40]. This inference is supported by an abstract understanding of how cooperative teachers select evidence; if there had been additional functions, the teacher would have shown them [3,42,43]. Children also penalize teachers who violate this expectation, such as a teacher who demonstrates only one function of a multi-function toy [44,45]. This result cannot be explained by tracking the accuracy of social information, because the teacher did provide the truth; it's just not the "whole truth."

Inferential social learning poses two key challenges for RL accounts of social learning. First, beyond learning reward-predictive properties of social information, such as accuracy, humans can also use repeated observations to infer the latent process that gave rise to others' actions. In one study (Figure 1; [••24]), adults received advice from two advisors who were equally knowledgeable and equally accurate, but who differed in their strategy: One advisor was conservative, the other risk-seeking. If participants only considered how often the advice is accurate, then they should not distinguish between these two advisors at all. Indeed, across both conditions, participants followed advice at similar rates and achieved similar earnings. However, participants differed in *how* they used the advice. Participants flexibly adjusted their use of advice based on the advisor's strategy, and they could explicitly report the shape of the advisor's choice function. Their choice behavior was best described by a probabilistic model that jointly infers the shape of the advisor's choice function and the value of unobservable options that are known to the advisor, compared to a model that tracked the advisor's accuracy.

Second, not all social learning can be easily explained as gradually learning costs and rewards through repeated observations of others. Even preverbal infants expect other agents to act efficiently [46,47], and they can infer the value of goals from a handful of observations, based on the costs that agents are willing to incur to attain them [48]. The **naïve utility calculus** provides a formal model of the computations that underlie these inferences; at its core, it proposes that our inferences about others' behavior are guided by the principle that agents act as utility maximizers, maximizing the rewards of their goal-directed actions while minimizing the costs [22,37]. This principle even extends to how learners interpret information provided by others, and how teachers select what to teach. For instance, 5- and 6-year-olds have an intuitive understanding of how much information is necessary for accurate learning and prefer not to learn from teachers who provide unnecessarily costly, overinformative demonstrations. As teachers, children not only resist being overinformative, but also choose what to teach (and what to let learners discover on their own) in ways that maximize learners' utilities and avoid unnecessary costs [49,50]. Thus, inferential social learning grounds our understanding of how humans learn from and teach others within a broader unifying framework that explains learning as rich, powerful inferences guided by an intuitive model of other minds [••2].

So far, we have critiqued the shortcomings of characterizing social learning using existing reinforcement learning models, and we have emphasized what they lack in light of approaches that characterize social learning as probabilistic inference. However, the two are not incompatible with one another. One of the core ideas in inferential social learning is that learners interpret information provided by others using an intuitive theory of others as utility maximizers [22,37]; this idea mirrors the ways in which reinforcement learning models instantiate scientific theories of how humans plan and make utility-maximizing decisions. Thus, inferential social learning can in principle be recast as an inverse reinforcement learning problem [51]. Recent work has applied inverse reinforcement learning models to study neural signals that track unobservable reward distributions inferred from others' actions [52]; related work has found neural signals that track value inferred from other people's stated confidence [53] and signals that arbitrate between inferring value from others' observable actions and using simpler strategies, such as copying [•54]. Looking farther into the future, the formalisms of reinforcement learning models are flexible enough that there is a gap between what these models can do, and how they have been applied so far to study the neural basis of social learning. This gap provides an opportunity to consider what could be gained by bridging the two frameworks.

New directions in the study of social learning

To close, we bring together insights from probabilistic and reinforcement learning approaches to social learning. We can now imagine the neuroscience of social learning not only as it has been, but as it could be.

First, synthesizing probabilistic and reinforcement learning approaches will allow us to gain purchase into how neural representations of mental states support inferential social learning. RL models have allowed us to identify neural signals that track reward-predictive properties of social information (e.g., accuracy); however, what underlies this social information is a fairly lean generative process. For example, if an advisor is simply right or wrong with some fixed probability [20], then their advice has as much latent structure as a coin flip. Here, the inferential social learning approach shines: It makes explicit commitments about the nature and the contents of the rich, structured representations involved in learning from others, in ways that RL models do not. One key potential substrate for these representations is the Theory of Mind network: Regions in this network are robustly engaged when reasoning about others' mental states [55] and contain representations of abstract mental state features, such as the perceptual source, valence, and evidentiary strength of others' beliefs [56,57]. However, less is known about how value computations and mental state representations work in tandem to support social learning. Here, insights from RL approaches can help guide our predictions. For example, abstract features of an advisor's mental state, such as their perceptual access, may be represented in regions within the Theory of Mind network, while the value of choice options that are inferred through their advice may be represented in reward-guided regions [••24].

Second, some kinds of social information not only provide information about the world, but also *feel good*; for example, a mentor's feedback can both improve a manuscript and provide a flush of pride. In other words, social feedback can be both epistemically valuable and intrinsically rewarding. How is social feedback represented and processed in the brain, and how does it differ from feedback from the physical environment? Past work has claimed that social and non-social rewards are represented using a common neural currency; for example, overlapping regions of striatum track both monetary rewards and social rewards (e.g., smiles and gains to reputation; [58,59]). In these tasks, social feedback is intrinsically rewarding, but it does not provide additional epistemic value—that is, rather than providing information about rewards, the social information itself *is* the reward. Yet positive feedback in social contexts—praise, smiles, and encouragement—is not merely a reward to be maximized. Children and adults alike interpret social feedback as communicative acts [60–•62]; for example, the same praise can be interpreted differently, based on the quality of the thing being praised and the selectivity of the person providing it. Thus, the computations that underlie learning from social feedback likely involve a generative model of the person providing it. Probabilistic models can help formalize how learners compute the epistemic and intrinsic value of social feedback [63]. Integrating both perspectives can provide insights into how the brain represents the epistemic value of social information and whether neural signals that track the epistemic value of social information overlap with or are distinct from those that track non-social value.

Our final question concerns how cognitive neuroscience can help build better scientific theories about our intuitive theories of the world. One key advantage of the naïve utility calculus is that it is *not* a theory of decision-making or action selection. Unlike RL, which aims to describe how people make decisions, the naïve utility calculus describes an intuitive theory—that is, people's lay intuitions of how others make decisions. While intuitive theories provide a "good enough" correspondence to reality, they can also differ from scientific theories in striking ways; for example, lay intuitions about physics are good enough to throw balls and build towers, but differ from classical mechanics [64,65]. Understanding how intuitive theories are neurally instantiated is a particularly exciting direction for future research. Some progress has been made in the domain of intuitive physics [66,•67] and aspects of Theory of Mind [56,57]. Yet it remains to be seen how the brain represents the costs and rewards of others' actions [68], and how these representations relate to neural signatures of action understanding [69,•70] or of one's own costs and rewards [71].

And we have only begun to scratch the surface—by bridging the two perspectives, we may also gain insights into the neural computations that support learning and exploration in social contexts [41,72], evaluations of others' reliability as advisors and teachers [20,60], and even our decisions about what is best to teach [49,73]. Bridging these perspectives can enrich our understanding of how the human brain supports social learning that is powerful, effective, and distinctively human.

References

1. Gergely G, Bekkering H, Király I: Rational imitation in preverbal infants. Nature 2002, 415:755.
2. Gweon H: Cognitive Foundations of Distinctively Human Social Learning. Trends Cogn Sci in press.•• This review paper summarizes recent computational and developmental research on inferential social learning, suggesting that human social learning is best explained as powerful probabilistic inferences over rich, structured representations, guided by an early-developing generative model of other agents.
3. Shafto P, Goodman ND, Griffiths TL: A rational account of pedagogical reasoning: teaching by, and learning from, examples. Cogn Psychol 2014, 71:55–89. [PubMed: 24607849]
4. Sutton RS, Barto AG: Reinforcement Learning, second edition: An Introduction. MIT Press; 2018.
5. Dayan P, Niv Y: Reinforcement learning: the good, the bad and the ugly. Curr Opin Neurobiol 2008, 18:185–196. [PubMed: 18708140]
6. Kable JW, Glimcher PW: The neurobiology of decision: consensus and controversy. Neuron 2009, 63:733–745. [PubMed: 19778504]
7. Howard JD, Reynolds R, Smith DE, Voss JL, Schoenbaum G, Kahnt T: Targeted Stimulation of Human Orbitofrontal Networks Disrupts Outcome-Guided Behavior. Curr Biol 2020, 30:490–498.e4. [PubMed: 31956033]
8. Sharpe MJ, Chang CY, Liu MA, Batchelor HM, Mueller LE, Jones JL, Niv Y, Schoenbaum G: Dopamine transients are sufficient and necessary for acquisition of model-based associations. Nat Neurosci 2017, 20:735–742. [PubMed: 28368385]
9. Steinberg EE, Keiflin R, Boivin JR, Witten IB, Deisseroth K, Janak PH: A causal link between prediction errors, dopamine neurons and learning. Nat Neurosci 2013, 16:966–973. [PubMed: 23708143]
10. Marr D: Vision: A computational investigation into the human representation and processing of visual information. San Francisco: WH Freeman; 1982.
11. Niv Y, Langdon A: Reinforcement learning with Marr. Curr Opin Behav Sci 2016, 11:67–73. [PubMed: 27408906]
12. Behrens TEJ, Hunt LT, Rushworth MFS: The computation of social behavior. Science 2009, 324:1160–1164. [PubMed: 19478175]
13 ••. Olsson A, Knapska E, Lindström B: The neural and computational systems of social learning. Nat Rev Neurosci 2020, 21:197–212. [PubMed: 32221497] This comprehensive review traces decades of progress in uncovering the neural bases of social learning and claims that social and non-social learning follow the same computational principles; by contrast, the present work critiques what this view has missed.

14. Charpentier CJ, O'Doherty JP: The application of computational models to social neuroscience: promises and pitfalls. Soc Neurosci 2018, 13:637–647. [PubMed: 30173633]

15. Lockwood PL, Klein-Flügge MC: Computational modelling of social cognition and behaviour—a reinforcement learning primer. Soc Cogn Affect Neurosci 2020, doi:10.1093/scan/nsaa040.

16. Morelli SA, Sacchet MD, Zaki J: Common and distinct neural correlates of personal and vicarious reward: A quantitative meta-analysis. Neuroimage 2015, 112:244–253. [PubMed: 25554428]

17. Morelli SA, Knutson B, Zaki J: Neural sensitivity to personal and vicarious reward differentially relate to prosociality and well-being. Soc Cogn Affect Neurosci 2018, 13:831–839. [PubMed: 30016481]

18. Atlas LY, Doll BB, Li J, Daw ND, Phelps EA: Instructed knowledge shapes feedback-driven aversive learning in striatum and orbitofrontal cortex, but not the amygdala. Elife 2016, 5.

19. Li J, Delgado MR, Phelps EA: How instructed knowledge modulates the neural systems of reward learning. Proc Natl Acad Sci U S A 2011, 108:55–60. [PubMed: 21173266]

20. Boorman ED, O'Doherty JP, Adolphs R, Rangel A: The behavioral and neural mechanisms underlying the tracking of expertise. Neuron 2013, 80:1558–1571. [PubMed: 24360551]

21. Hackel LM, Doll BB, Amodio DM: Instrumental learning of traits versus rewards: dissociable neural correlates and effects on choice. Nat Neurosci 2015, 18:1233–1235. [PubMed: 26237363]

22. Jara-Ettinger J, Schulz LE, Tenenbaum JB: The naive utility calculus as a unified, quantitative framework for action understanding. Cogn Psychol 2020, 123:101334. [PubMed: 32738590]

23. Siegel JZ, Mathys C, Rutledge RB, Crockett MJ: Beliefs about bad people are volatile. Nat Hum Behav 2018, 2:750–756. [PubMed: 31406285]

24 ••. Vélez N, Gweon H: Integrating Incomplete Information With Imperfect Advice. Top Cogn Sci 2019, 11:299–315. [PubMed: 30414253] This paper finds that adults flexibly adjust their use of advice based on the knowledge, intent, and strategy of the advisor, in ways that are consistent with a probabilistic model that jointly infers how the advisor selects advice and what the advisor can see.

25. Lake BM, Ullman TD, Tenenbaum JB, Gershman SJ: Building machines that learn and think like people. Behav Brain Sci 2017, 40:e253. [PubMed: 27881212]

26. Tenenbaum JB, Kemp C, Griffiths TL, Goodman ND: How to grow a mind: statistics, structure, and abstraction. Science 2011, 331:1279–1285. [PubMed: 21393536]

27. Battaglia PW, Hamrick JB, Tenenbaum JB: Simulation as an engine of physical scene understanding. Proc Natl Acad Sci U S A 2013, 110:18327–18332. [PubMed: 24145417]

28. Griffiths TL, Tenenbaum JB: Optimal predictions in everyday cognition. Psychol Sci 2006, 17:767–773. [PubMed: 16984293]

29. Teglas E, Vul E, Girotto V, Gonzalez M, Tenenbaum JB, Bonatti LL: Pure Reasoning in 12-Month-Old Infants as Probabilistic Inference. Science 2011, 332:1054–1059. [PubMed: 21617069]

30. Goodman ND, Ullman TD, Tenenbaum JB: Learning a theory of causality. Psychol Rev 2011, 118:110–119. [PubMed: 21244189]

31. Griffiths TL, Tenenbaum JB: Theory-based causal induction. Psychol Rev 2009, 116:661–716. [PubMed: 19839681]

32. Lake BM, Salakhutdinov R, Tenenbaum JB: Human-level concept learning through probabilistic program induction. Science 2015, 350:1332–1338. [PubMed: 26659050]

33. Markant DB, Gureckis TM: Is it better to select or to receive? Learning via active and passive hypothesis testing. J Exp Psychol Gen 2014, 143:94–122. [PubMed: 23527948]

34. Gweon H, Tenenbaum JB, Schulz LE: Infants consider both the sample and the sampling process in inductive generalization. Proc Natl Acad Sci U S A 2010, 107:9066–9071. [PubMed: 20435914]

35. Xu F, Tenenbaum JB: Word learning as Bayesian inference. Psychol Rev 2007, 114:245–272. [PubMed: 17500627]

36. Baker CL, Jara-Ettinger J, Saxe R, Tenenbaum JB: Rational quantitative attribution of beliefs, desires and percepts in human mentalizing. Nature Human Behaviour 2017, 1:0064.

37. Jara-Ettinger J, Gweon H, Schulz LE, Tenenbaum JB: The Naïve Utility Calculus: Computational Principles Underlying Commonsense Psychology. Trends Cogn Sci 2016, 20:589–604. [PubMed: 27388875]

38. Kushnir T, Xu F, Wellman HM: Young children use statistical sampling to infer the preferences of other people. Psychol Sci 2010, 21:1134–1140. [PubMed: 20622142]

39. Lucas CG, Griffiths TL, Xu F, Fawcett C, Gopnik A, Kushnir T, Markson L, Hu J: The child as econometrician: a rational model of preference understanding in children. PLoS One 2014, 9:e92160. [PubMed: 24667309]

40. Bonawitz E, Shafto P, Gweon H, Goodman ND, Spelke E, Schulz L: The double-edged sword of pedagogy: Instruction limits spontaneous exploration and discovery. Cognition 2011, 120:322–330. [PubMed: 21216395]

41. Wu Y, Gweon H: Preschool-aged children jointly consider others' emotional expressions and prior knowledge to decide when to explore. 2019, doi: 10.31234/osf.io/ckh6j.

42. Shafto P, Goodman ND, Frank MC: Learning From Others: The Consequences of Psychological Reasoning for Human Learning. Perspect Psychol Sci 2012, 7:341–351. [PubMed: 26168471]

43. Shneidman L, Gweon H, Schulz LE, Woodward AL: Learning from others and spontaneous exploration: A cross-cultural investigation. Child Dev 2016, 87:723–735. [PubMed: 27189400]

44. Gweon H, Pelton H, Konopka JA, Schulz LE: Sins of omission: children selectively explore when teachers are under-informative. Cognition 2014, 132:335–341. [PubMed: 24873737]

45. Gweon H, Asaba M: Order Matters: Children's Evaluation of Underinformative Teachers Depends on Context. Child Dev 2018, 89:e278–e292. [PubMed: 28542838]

46. Gergely G, Csibra G: Teleological reasoning in infancy: the naive theory of rational action. Trends Cogn Sci 2003, 7:287–292. [PubMed: 12860186]

47. Liu S, Spelke ES: Six-month-old infants expect agents to minimize the cost of their actions. Cognition 2017, 160:35–42. [PubMed: 28040551]

48. Liu S, Ullman TD, Tenenbaum JB, Spelke ES: Ten-month-old infants infer the value of goals from the costs of actions. Science 2017, 358:1038–1041. [PubMed: 29170232]

49. Bridgers S, Jara-Ettinger J, Gweon H: Young children consider the expected utility of others' learning to decide what to teach. Nat Hum Behav 2020, 4:144–152. [PubMed: 31611659]

50. Gweon H, Shafto P, Schulz L: Development of children's sensitivity to overinformativeness in learning and teaching. Dev Psychol 2018, 54:2113–2125. [PubMed: 30265027]

51. Jara-Ettinger J: Theory of mind as inverse reinforcement learning. Current Opinion in Behavioral Sciences 2019, 29:105–110.

52. Collette S, Pauli WM, Bossaerts P, O'Doherty J: Neural computations underlying inverse reinforcement learning in the human brain. Elife 2017, 6.

53. Campbell-Meiklejohn D, Simonsen A, Frith CD, Daw ND: Independent Neural Computation of Value from Other People's Confidence. J Neurosci 2017, 37:673–684. [PubMed: 28100748]

54 •. Charpentier CJ, Iigaya K, O'Doherty JP: A Neuro-computational Account of Arbitration between Choice Imitation and Goal Emulation during Human Observational Learning. Neuron 2020, 106:687–699.e7. [PubMed: 32187528] These fMRI studies suggest that humans flexibly arbitrate between inferring the reward distributions underlying others' actions ("emulation") and copying others ("imitation"), and distinct neural signals track the arbitration between emulation and imitation.

55. Saxe R, Kanwisher N: People thinking about thinking people: The role of the temporo-parietal junction in "theory of mind." Neuroimage 2003, 19:1835–1842. [PubMed: 12948738]

56. Koster-Hale J, Bedny M, Saxe R: Thinking about seeing: perceptual sources of knowledge are encoded in the theory of mind brain regions of sighted and blind adults. Cognition 2014, 133:65–78. [PubMed: 24960530]

57. Koster-Hale J, Richardson H, Velez N, Asaba M, Young L, Saxe R: Mentalizing regions represent distributed, continuous, and abstract dimensions of others' beliefs. Neuroimage 2017, 161:9–18. [PubMed: 28807871]

58. Izuma K, Saito DN, Sadato N: Processing of social and monetary rewards in the human striatum. Neuron 2008, 58:284–294. [PubMed: 18439412]

59. Lin A, Adolphs R, Rangel A: Social and monetary reward learning engage overlapping neural substrates. Soc Cogn Affect Neurosci 2012, 7:274–281. [PubMed: 21427193]

60. Asaba M, Hembacher E, Qiu H, Anderson B, Frank MC, Gweon H: Young children use statistical evidence to infer the informativeness of praise. In CogSci. . 2018.

61. Ho MK, MacGlashan J, Littman ML, Cushman F: Social is special: A normative framework for teaching with and learning from evaluative feedback. Cognition 2017, 167:91–106. [PubMed: 28341268]

62 •. Ho MK, Cushman F, Littman ML, Austerweil JL: People teach with rewards and punishments as communication, not reinforcements. J Exp Psychol Gen 2019, 148:520–549. [PubMed: 30802127] These behavioral studies find that humans use social feedback to signal desired behavior to a learner; social feedback is a communicative act, rather than a reward to be maximized.

63. Yoon EJ, Frank MC, Tessler MH, Goodman ND: Polite speech emerges from competing social goals. 2018, doi:10.31234/osf.io/67ne8.

64. McCloskey M, Caramazza A, Green B: Curvilinear motion in the absence of external forces: naive beliefs about the motion of objects. Science 1980, 210:1139–1141. [PubMed: 17831469]

65. Saxe R: Against simulation: the argument from error. Trends Cogn Sci 2005, 9:174–179. [PubMed: 15808499]

66. Fischer J, Mikhael JG, Tenenbaum JB, Kanwisher N: Functional neuroanatomy of intuitive physical inference. Proc Natl Acad Sci U S A 2016, 113:E5072–81. [PubMed: 27503892]

67 •. Schwettmann S, Tenenbaum JB, Kanwisher N: Invariant representations of mass in the human brain. Elife 2019, 8.This fMRI study provides one example of how intuitive theories may be neurally instantiated, in the domain of intuitive physics; in particular, neural representations of object mass in regions implicated in physical reasoning are invariant to scenario, material, friction, and motion.

68. Jara-Ettinger J, Gweon H, Tenenbaum JB, Schulz LE: Children’s understanding of the costs and rewards underlying rational action. Cognition 2015, 140:14–23. [PubMed: 25867996]

69. Anzellotti S, Young LL: The Acquisition of Person Knowledge. Annual Review of Psychology 2020, 71:613–634.

70 •. Deen B, Saxe R, Kanwisher N: Processing communicative facial and vocal cues in the superior temporal sulcus: Faces and voices in the STS. Neuroimage in press.This fMRI study finds that STS represents not just faces and voices, but is sensitive to the communicative nature of these signals.

71. Croxson PL, Walton ME, O’Reilly JX, Behrens TEJ, Rushworth MFS: Effort-Based Cost–Benefit Valuation and the Human Brain. J Neurosci 2009, 29:4531–4541. [PubMed: 19357278]

72. Zhang L, Gläscher J: A brain network supporting social influences in human decision-making. Science Advances 2020, 6:eabb4159. [PubMed: 32875112]

73. Apps MAJ, Lesage E, Ramnani N: Vicarious reinforcement learning signals when instructing others. J Neurosci 2015, 35:2904–2913. [PubMed: 25698730]

Highlights

- RL models have been applied to study the neural underpinnings of social learning.
- Past work has largely found neural correlates of observable, reward-predictive cues.
- But, in learning from others, humans make inferences that go beyond observable data.
- Such inferences rely on a model of how others' mental states give rise to actions.
- Bridging RL and Bayesian approaches can deepen our understanding of social learning.

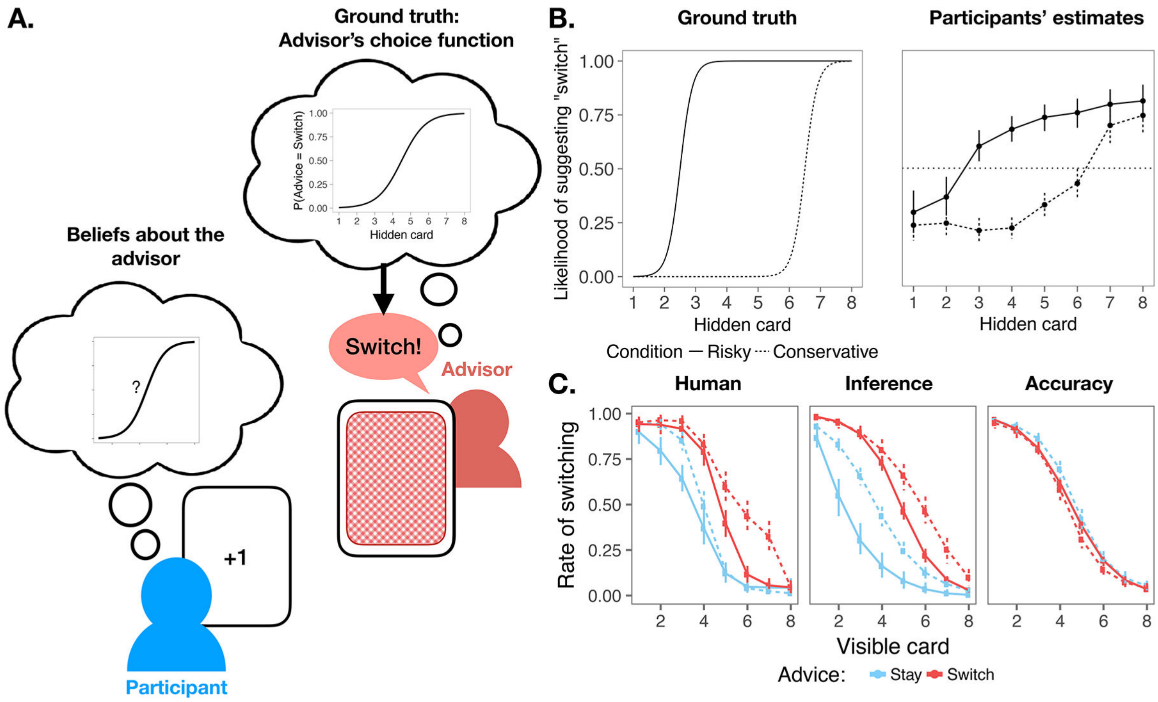

Figure 1: Humans can infer the latent process that gave rise to others' actions.

Schematic of model and task design [••24]: Participants played a card game where they could choose to stay with the points in a visible card, or switch to a hidden card that was only known to the advisor. The probability of the advisor telling participants to "switch" scaled according to a choice function that was unknown to participants. The Inference model worked backwards from the observed evidence to jointly infer the shape of the advisor's choice function and the value of the hidden card. **B**. Ground-truth choice functions of the risky (solid) and conservative (dotted) advisors (left), and participant's explicit reports of the shape of the choice function at post-test (right). **C**. Participants flexibly adjusted their use of advice based on the advisor's strategy *(Human)*; their choice behavior was better described by the *Inference* model than by a model that tracked observed *Accuracy*.