Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Behavioral Sciences

Theory of mind as inverse reinforcement learning

Julian Jara-Ettinger[1,2]

We review the idea that Theory of Mind—our ability to reason about other people's mental states—can be formalized as inverse reinforcement learning. Under this framework, expectations about how mental states produce behavior are captured in a reinforcement learning (RL) model. Predicting other people's actions is achieved by simulating a RL model with the hypothesized beliefs and desires, while mental-state inference is achieved by inverting this model. Although many advances in inverse reinforcement learning (IRL) did not have human Theory of Mind in mind, here we focus on what they reveal when conceptualized as cognitive theories. We discuss landmark successes of IRL, and key challenges in building human-like Theory of Mind.

Addresses
[1] Department of Psychology, Yale University, New Haven, CT, United States
[2] Department of Computer Science, Yale University, New Haven, CT, United States

Current Opinion in Behavioral Sciences 2019, **29**:105–110

This review comes from a themed issue on **Artificial Intelligence**

Edited by **Matt Botvinick** and **Sam Gershman**

https://doi.org/10.1016/j.cobeha.2019.04.010

Human theory of mind

Imagine going to meet a friend for coffee only to find yourself sitting alone. You know your friend is scattered, so you start to suspect that she got distracted on the way. Or maybe she lost track of time, or got the date flat-out wrong. As you're thinking how typical this is of her, you suddenly remember that the coffee shop has a second location right next to your friend's office. Without talking to her, you realize that she probably had the other location in mind; that (just like you) she forgot the coffee shop had two locations; and that, for all you know, she's probably sitting there wondering why you didn't show up.

To make sense of what went wrong, you had to use a mental model of your friend's mind—what she prefers, what she knows, and what she assumes. This capacity, called a *Theory of Mind* [1,2], lets us intuit how people we're familiar with might act in different situations. But, beyond that, it also lets us infer what even strangers might think or want based on how they behave. Research in cognitive science suggests that we infer mental states by thinking of other people as utility maximizers: constantly acting to maximize the rewards they obtain while minimizing the costs that they incur [3–5]. Using this assumption, even children can infer other people's preferences [6,7], knowledge [8,9], and moral standing [10–12].

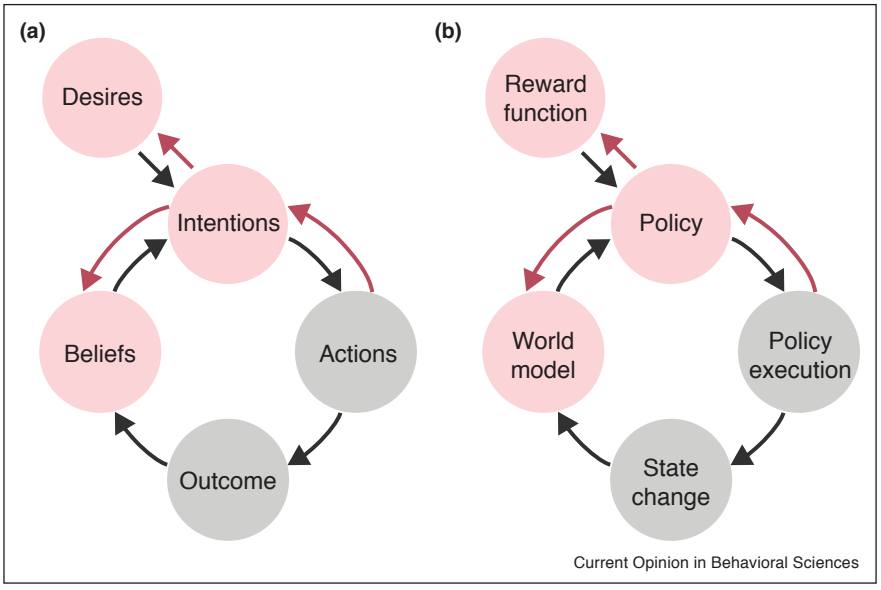
Theory of mind as inverse reinforcement learning

Computationally, our intuitions about how other minds work can be formalized using frameworks developed in a classical area of AI: model-based reinforcement learning (hereafter reinforcement learning or RL)[1] . RL problems focus on how to combine a world model with a reward function to produce a sequence of actions, called a policy, that maximizes agents' rewards while minimizing their costs. Thus, the principles of RL planning resemble the assumptions that we make about other people's behavior [5,3,4]. Taking advantage of this similarity, we can formalize our mental model of other people's minds as being roughly equivalent to a reinforcement learning model (Figure 1). Under this approach, mental-state inference from observable behavior is equivalent to *inverse* reinforcement learning (IRL): inferring agents' unobservable model of the world and reward function, given some observed actions.

Inverse Reinforcement Learning problems face a critical challenge: We can often explain someone's actions by appealing to different combinations of mental states. Returning to the coffee shop example in the introduction, to make sense of what happened, we did not just settle on the first plausible explanation (e.g., maybe your friend lost track of time), but continuously sought more explanations, even if the ones we already had were good enough (because, even if they explained your friend's absence, they could still be wrong). Thus, mental-state inference requires tracking multiple explanations and weighting them by how well they explain the data. Bayesian inference—a general approach that successfully characterizes how people "invert" intuitive theories in many domains of cognition [19]—has been effective in explaining how people do this. In simple two-dimensional displays, IRL through Bayesian inference produces human-like judgments when inferring people's goals [16], beliefs [17], desires [4], and helpfulness [12].

[1] The term reinforcement learning emphasizes the learning component, but the framework also captures how agents act under complete knowledge of the world and the rewards in it.

Figure 1



Simple schematic of how Theory of Mind can be modeled as Inverse Reinforcement Learning. This approach follows a tradition in cognitive science that argues that people make sense of their environment through working mental models [2,13–15]. (a) Core Theory of Mind components. People's beliefs about the world, combined with their desires, determine what they intend to do. People's intentions guide their actions, which produce outcomes that change their beliefs about the world. Pink arrows represent mental-state inference. (b) Core model-based reinforcement learning components. A world model combined with the reward function generate a policy via utility maximization. Executing the policy produces state changes, which, in turn, lead the agent to revise its world model. Pink arrows represent inverse reinforcement learning: recovering the latent world model and reward function, given an observed policy execution. In practice, there is little agreement on how to map elements from RL models onto Theory of Mind. [16], for instance, interpreted reward functions as goals, [17 as desires, and [18 as context-specific intentions.

Inverse reinforcement learning in use

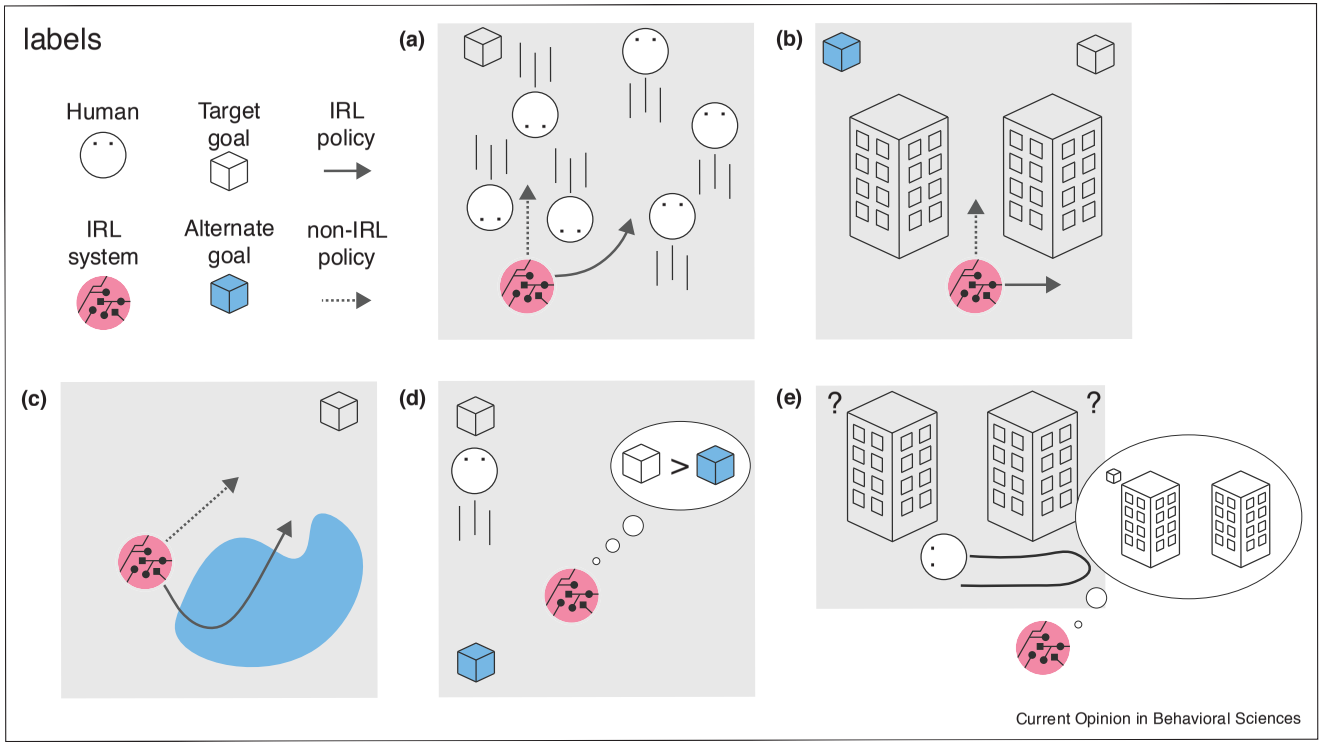
In Cognitive Science, Theory of Mind has been theoretically and empirically posited as central to a broad array of cognitive activities from language understanding [28,29] to moral reasoning [30–32]. Research in IRL suggests the same (Figure 2). In robotics, RL planners that integrate IRL can predict where pedestrians are headed and preemptively adjust their plan to avoid collisions (Figure 2a; [20,21]). Conversely, RL planners can also use IRL on their own actions to ensure that observers will be able to infer the robot's goal as quickly as possible (Figure 2b; [22–24,33]). Using a similar logic, IRL can also be integrated into RL planners to generate pedagogical actions designed to help observers learn about the world (Figure 2c; [25]). IRL has also been fruitful in solving the problem of aligning a system's values with our own. Explicitly encoding reward functions into RL planners is prone to errors and oversights. Systems with IRL can instead infer the reward function from a person's actions, and use it as their own. This allows for easy transfer of rewards across agents (Figure 2d; [26,27]), including rewards that encode moral values [34]. More broadly, IRL can jointly infer other agents' beliefs and desires (including desires to help or hinder others; [12,11]), and even the location of unobservable rewards, by watching other agents navigate the world (Figure 2e; [17,35]).

Finally, cognitively-inspired models of language understanding are not usually conceptualized as IRL because the domain lacks the spatiotemporal properties typical of RL problems. These models, however, share similar key ideas with IRL in that they work by modeling speakers as rational agents that trade off costs with rewards. This approach explains how we determine the meaning behind ambiguous utterances [36]; how we infer speakers' knowledge based on their choice of words (e.g., suspecting that the speaker knows there are two cats if we hear them refer to 'the *big* cat' instead of just 'the cat') [37]; how we make sense of non-literal word meanings [38]; and even how speakers use prosody to ensure listeners will get the meaning they wish to convey [39] (see [40] for review).

Making inverse reinforcement learning useful

Despite the success of IRL, its practical use is limited because inverting reinforcement learning models is computationally expensive. Deep learning—a subclass of AI, historically known for its emphasis on biological, rather than cognitive, plausibility [13,41]—has recently shown a strong advantage in speed over competing approaches, especially in the reinforcement learning domain [42–44]. Recent work has shown that it is also possible to implement IRL in neural networks [45–47], but these implementations face challenges characteristic of deep

Figure 2



Conceptual illustrations of IRL in use. These schematics capture the key ideas behind each advance, but, for the sake of clarity, diverge from the actual experiments in the cited work. The circuit represents a planner with IRL. The gray cube represents a rewarding target and the blue cube represents a non-rewarding potential target. Dotted arrows show valid policies for an RL planner, and the gray arrows show preferred paths after IRL is integrated into the planner. (a) IRL can predict crowd movements and adjust policies accordingly [20,21]. (b) IRL can be used to favor paths that allows observers to quickly infer its goal (moving upwards is equally efficient than moving rightward, but it would make the agent's goal temporarily ambiguous) [22–24]. (c) IRL can be used to design actions that 'teach' about the world, such as detouring to reveal that it is safe to navigate through the blue region [25]. d) IRL can be used to infer and copy another agent's reward function [26,27]. e) IRL infers the location of the gray cube, based on the agent's actions [17].

learning: they require vast amounts of labeled examples for training and they do not generalize well to new tasks or environments [13]. For instance, state-of-the-art IRL through deep learning [47] requires 32 million training examples to perform goal-inference at the capacity of a six-month-old infant [48]. If humans acquired Theory of Mind in a similar way, infants would need to receive almost 175,000 labeled goal-training episodes per day, every day.

These challenges are already being mitigated by networks specifically designed to implement IRL [49,46,45]. And meta-learning—algorithms that, when trained on multiple tasks, learn general properties that reduce the need for data —will likely play a role in years to come [50–52]. Yet, deep IRL with the flexibility of more traditional IRL models [17,53] remains distant [13]. One solution that has proved fruitful in other domains is to marry the two approaches [54–57]. A deep net can be trained to quickly transform observed actions into candidate mental states. After this initial guess, a full-blown symbolic RL model can take over to refine these inferences and use them for a variety of tasks including generating predictions, producing explanations, and making social evaluations. Beyond its practical usefulness, this approach may provide a cognitively-plausible theory that resembles the dichotomy between fast automatic agency detection [58–60] and richer but slower mental-state reasoning in humans [61,53,17].

Inverse reinforcement learning as theory of mind

While Inverse Reinforcement Learning captures core inferences in human action-understanding, the way this framework has been used to represent beliefs and desires fails to capture the more structured mental-state reasoning that people use to make sense of others [61,62].

Belief representations

RL frameworks were historically designed to deal with uncertainty in the broadest sense, including uncertainty about the agent's own position in space (e.g., a noisy sensor may not correctly estimate a robot's distance to a wall). IRL often uses RL models called Partially Observable Markov Decision Processes [63], where beliefs are represented as probability distributions over every possible state of the world (e.g., [17]). This guarantees that the representation is coherent and complete, but it also lacks structure that human Theory of Mind exploits.

When we infer other people's mental states, we often infer small parts of what they know or believe (e.g., inferring that Sally didn't know a coffee cup had leftover wine as we see her take a sip and spit it out) without reasoning about beliefs that are clearly true (e.g., is Sally aware that she is standing on her feet?) or irrelevant (e.g., does Sally know the speed of sound?). Yet, current IRL models can only evaluate the plausibility of beliefs that are complete descriptions of everything an agent believes. Intuitively, this is because the only way to tell whether beliefs about some aspect of the world matter, is by testing if they do.Humans appear to solve this problem by assuming that other people's beliefs are similar to our own in most ways. If so, IRL may become more human-like if it is initialized with an assumption that other people's beliefs in immediate situations are similar to its own representation of the world, and then, proposals about other people's beliefs are not meant to provide a full description of what's in their mind, but rather to capture in what ways their beliefs are critically similar or different from our own.

Desire representations

Current IRL models typically represent desires as a function that assigns a numerical value to each possible state of the world (although note that there is little agreement on how to map components of RL models onto concepts in human Theory of Mind [17,16,53,18]). While useful for predicting agents' immediate actions (namely, keep navigating towards inferred high-reward states), this formalism does not reveal where these rewards come from, and it does not specify how to predict what the rewards may be in a new environment. To achieve this, it is critical to recognize that rewards are often the combination of simpler desires working at different timescales and levels of abstraction. Making sense of even the simplest actions, such as watching someone get coffee, involves considering different sources of rewards (perhaps not only enjoying coffee, but also the company of friends), their tradeoffs (they may have a meeting soon, preventing them from going to the superior coffee shop that is located farther away), and the costs the agent was willing to incur (time, distance, and money). From an observer's standpoint, actions alone do not contain enough information to reveal how many sources of costs and rewards are at play. This suggests that effective IRL needs strong inductive biases that exploit knowledge about the general types of rewards agents have, the types of rewards that are usually at play in different contexts, and the specific rewards that different agents act under.

A bigger challenge to current approaches is that reward functions fail to capture the logical and temporal structure of desires. When we reason about others, we recognize that their desires can depend on other desires (someone might only enjoy coffee after having eaten something), that they can depend on context (drinking coffee may be more appealing for someone in the morning), and that they can be conjunctive (liking coffee with sugar, but neither in isolation) or disjunctive (liking coffee and milk, but not together). A crucial challenge towards human-like Theory of Mind is developing reward representations that support expressing desires which can be fulfilled in

multiple ways, with spatiotemporal constraints, and varying degrees of abstraction. Advances in hierarchical RL may play a critical role towards this goal [64,65]. In addition, recent work suggests that representations originally developed to explain how people build complex concepts by composing simpler ones [14,66] may be useful. Under this approach, desires are represented as propositions built by composing potential sources of rewards, and reward functions are synthesized in each context accordingly. In models like these, mental-state inference corresponds to inferring the agents' unobservable reward function, as well as the proposition that generated it [18], and it produces human-like inferences that capture temporal and logical structures of desires.

Beyond inverse reinforcement learning

Human intuitive theories are often approximations of the phenomena they aim to explain [5,67], allowing us to ignore complexities that are less useful for prediction and explanation, much in the same way that scientific theories gain explanatory power through abstraction and simplification [68,1,69,70]. Theory of Mind in humans may be successful precisely because it only approximates how humans actually make choices. If so, IRL may need to depart from frameworks developed in RL, which focus on the nuances of action production.

Perhaps the greatest challenge in modeling Theory of Mind as Inverse Reinforcement Learning lies in capturing variability in thinking. IRL focuses on recovering the beliefs and desires under the assumption that all agents make choices and take actions in identical ways. Yet, we recognize that two people with the same beliefs and desires may still make different reasonable choices and take different reasonable actions. Theory of Mind in the real world goes beyond mental-state inference and includes learning agent-specific models of how people think. We recognize that people forget and misremember, that they get impatient, they fail to think of solutions that feel obvious in retrospect, and they experience frustration and regret. For IRL as Theory of Mind to succeed, we must build a model that is more human than RL.

Acknowledgments

This work was supported by a Google Faculty Research Award. Thanks to members of Yale's Computation and Cognitive Development lab for feedback on an earlier version of this manuscript.

References

1. Dennett DC: *The intentional stance*. MIT Press; 1989.
2. Gopnik A, Meltzoff AN, Bryant P: *Words, thoughts, and theories*. 1997.
3. 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.
4. Jern A, Lucas CG, Kemp C: **People learn other peoples preferences through inverse decision-making**. *Cognition* 2017, **168**:46-64.
5. Jara-Ettinger J, Gweon H, Schulz LE, Tenenbaum JB: **The naïve utility calculus: Computational principles underlying commonsense psychology**. *Trends Cognit Sci* 2016, **20**:589-604.
6. Jara-Ettinger J, Gweon H, Tenenbaum JB, Schulz LE: **Childrens understanding of the costs and rewards underlying rational action**. *Cognition* 2015, **140**:14-23.
7. 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.
8. Jara-Ettinger J, Floyd S, Tenenbaum JB, Schulz LE: **Children understand that agents maximize expected utilities**. *J Exp Psychol: Gen* 2017, **146**:1574.
9. H. Richardson, C. Baker, J. Tenenbaum, R. Saxe, The development of joint belief-desire inferences, in: Proceedings of the Annual Meeting of the Cognitive Science Society, volume 34.
10. Jara-Ettinger J, Tenenbaum JB, Schulz LE: **Not so innocent: Toddlers inferences about costs and culpability**. *Psychol Sci* 2015, **26**:633-640.
11. Kiley Hamlin J, Ullman T, Tenenbaum J, Goodman N, Baker C: **The mentalistic basis of core social cognition: Experiments in preverbal infants and a computational model**. *Develop Sci* 2013, **16**:209-226.
12. Ullman T, Baker C, Macindoe O, Evans O, Goodman N, Tenenbaum JB: Help or hinder: Bayesian models of social goal inference, in: Advances in neural information processing systems 1874–1882.
13. Lake BM, Ullman TD, Tenenbaum JB, Gershman SJ: **Building machines that learn and think like people**. *Behav Brain Sci* 2017, **40**.
14. Goodman ND, Tenenbaum JB, Feldman J, Griffiths TL: **A rational analysis of rule-based concept learning**. *Cognit Sci* 2008, **32**:108-154.
15. Goodman N, Mansinghka V, Roy DM, Bonawitz K, Tenenbaum JB: Church: a language for generative models, arXiv preprint arXiv:1206.3255 (2012).
16. Baker CL, Saxe R, Tenenbaum JB: **Action understanding as inverse planning**. *Cognition* 2009, **113**:329-349.
17. Baker CL, Jara-Ettinger J, Saxe R, Tenenbaum JB: **Rational quantitative attribution of beliefs, desires and percepts in human mentalizing**. *Nat Hum Behav* 2017, **1**:0064.
18. Velez-Ginorio J, Siegel M, Tenenbaum JB, Jara-Ettinger J: *Interpreting actions by attributing compositional desires*. 2017.
19. Tenenbaum JB, Kemp C, Griffiths TL, Goodman ND: **How to grow a mind: Statistics, structure, and abstraction**. *Science* 2011, **331**:1279-1285.
20. Kim B, Pineau J: **Socially adaptive path planning in human environments using inverse reinforcement learning**. *Int J Soc Robot* 2016, **8**:51-66.
21. Kretzschmar H, Spies M, Sprunk C, Burgard W: **Socially compliant mobile robot navigation via inverse reinforcement learning**. *Int J Robot Res* 2016, **35**:1289-1307.
22. Dragan AD, Lee KC, Srinivasa SS: **Legibility and predictability of robot motion**. *Proceedings of the 8th ACM/IEEE international conference on Human-robot interaction* 2013:301-308.
23. Dragan A, Srinivasa S: *Generating legible motion*. 2013.
24. Dragan A, Srinivasa S: **Integrating human observer inferences into robot motion planning**. *Autonomous Robots* 2014, **37**:351-368.
25. Ho MK, Littman M, MacGlashan J, Cushman F, Austerweil JL: **Showing versus doing: Teaching by demonstration**. *Adv Neural Inform Process Syst* 2016:3027-3035.

26. Hadfield-Menell D, Russell SJ, Abbeel P, Dragan A: **Cooperative inverse reinforcement learning**. *Adv Neural Inform Process Syst* 2016:3909-3917.

27. D. Malik, M. Palaniappan, J. F. Fisac, D. Hadfield-Menell, S. Russell, A. D. Dragan, An efficient, generalized bellman update for cooperative inverse reinforcement learning, arXiv preprint arXiv:1806.03820 (2018).

28. Rubio-Fernández P: **The director task: A test of theory-of-mind use or selective attention?** *Psychonomic Bull Rev* 2017, **24**:1121-1128.

29. R. X. Hawkins, H. Gweon, N. D. Goodman, Speakers account for asymmetries in visual perspective so listeners don't have to, arXiv preprint arXiv:1807.09000 (2018).

30. Young L, Cushman F, Hauser M, Saxe R: **The neural basis of the interaction between theory of mind and moral judgment**. *Proc Natl Acad Sci* 2007, **104**:8235-8240.

31. Young L, Camprodon JA, Hauser M, Pascual-Leone A, Saxe R: **Disruption of the right temporoparietal junction with transcranial magnetic stimulation reduces the role of beliefs in moral judgments**. *Proc Natl Acad Sci* 2010, **107**:6753-6758.

32. Moran JM, Young LL, Saxe R, Lee SM, O'Young D, Mavros PL, Gabrieli JD: **Impaired theory of mind for moral judgment in high-functioning autism**. *Proc Natl Acad Sci* 2011, **108**:2688-2692.

33. D. Strouse, M. Kleiman-Weiner, J. Tenenbaum, M. Botvinick, D. J. Schwab, Learning to share and hide intentions using information regularization, in: Advances in Neural Information Processing Systems 10270-10281.

34. Kleiman-Weiner M, Saxe R, Tenenbaum JB: **Learning a commonsense moral theory**. *Cognition* 2017, **167**:107-123.

35. S. Reddy, A. D. Dragan, S. Levine, Where do you think you're going?: Inferring beliefs about dynamics from behavior, arXiv preprint arXiv:1805.08010 (2018).

36. Frank MC, Goodman ND: **Predicting pragmatic reasoning in language games**. *Science* 2012, **336** 998–998.

37. Rubio-Fernández P, Jara-Ettinger J: *Joint inferences of speakers beliefs and referents based on how they speak*. 2018.

38. Kao JT, Wu JY, Bergen L, Goodman ND: **Nonliteral understanding of number words**. *Proc Natl Acad Sci* 2014, **111**:12002-12007.

39. Bergen L, Goodman ND: **The strategic use of noise in pragmatic reasoning**. *Topics in cognitive science* 2015, **7**:336-350.

40. Goodman ND, Frank MC: **Pragmatic language interpretation as probabilistic inference**. *Trends Cognit Sci* 2016, **20**:818-829.

41. Hassabis D, Kumaran D, Summerfield C, Botvinick M: **Neuroscience-inspired artificial intelligence**. *Neuron* 2017, **95**:245-258.

42. Mnih V, Kavukcuoglu K, Silver D, Rusu AA, Veness J, Bellemare MG, Graves A, Riedmiller M, Fidjeland AK, Ostrovski G *et al.*: **Human-level control through deep reinforcement learning**. *Nature* 2015, **518**:529.

43. Silver D, Huang A, Maddison CJ, Guez A, Sifre L, Van Den Driessche G, Schrittwieser J, Antonoglou I, Panneershelvam V, Lanctot M *et al.*: **Mastering the game of go with deep neural networks and tree search**. *Nature* 2016, **529**:484.

44. LeCun Y, Bengio Y, Hinton G: *Deep learning, nature* 2015, **521**:436.

45. C. Finn, S. Levine, P. Abbeel, Guided cost learning: Deep inverse optimal control via policy optimization, in: International Conference on Machine Learning, 49-58.

46. M. Wulfmeier, P. Ondruska, I. Posner, Deep inverse reinforcement learning, CoRR, abs/1507.04888 (2015).

47. N. C. Rabinowitz, F. Perbet, H. F. Song, C. Zhang, S. Eslami, M. Botvinick, Machine theory of mind, arXiv preprint arXiv:1802.07740 (2018).

48. Woodward AL: **Infants selectively encode the goal object of an actor's reach**. *Cognition* 1998, **69**:1-34.

49. M. Wulfmeier, P. Ondruska, I. Posner, Maximum entropy deep inverse reinforcement learning, arXiv preprint arXiv:1507.04888 (2015).

50. A. Santoro, S. Bartunov, M. Botvinick, D. Wierstra, T. Lillicrap, Meta-learning with memory-augmented neural networks, in: International conference on machine learning, 1842-1850.

51. C. Finn, P. Abbeel, S. Levine, Model-agnostic meta-learning for fast adaptation of deep networks, arXiv preprint arXiv:1703.03400 (2017).

52. K. Xu, E. Ratner, A. Dragan, S. Levine, C. Finn, Learning a prior over intent via meta-inverse reinforcement learning, arXiv preprint arXiv:1805.12573 (2018).

53. J. Jara-Ettinger, L. E. Schulz, J. B. Tenenbaum, A naive utility calculus as the foundation of action understanding (under review).

54. Yildirim I, Freiwald W, Tenenbaum J: **Efficient inverse graphics in biological face processing**. *bioRxiv* 2018:282798.

55. I. Yildirim, T. D. Kulkarni, W. A. Freiwald, J. B. Tenenbaum, Efficient and robust analysis-by-synthesis in vision: A computational framework, behavioral tests, and modeling neuronal representations, in: Annual conference of the cognitive science society, volume 1.

56. J. Wu, I. Yildirim, J. J. Lim, B. Freeman, J. Tenenbaum, Galileo: Perceiving physical object properties by integrating a physics engine with deep learning, in: Advances in neural information processing systems, 127-135.

57. P. Moreno, C. K. Williams, C. Nash, P. Kohli, Overcoming occlusion with inverse graphics, in: European Conference on Computer Vision, Springer, 170-185.

58. Gao T, McCarthy G, Scholl BJ: **The wolfpack effect: Perception of animacy irresistibly influences interactive behavior**. *Psychol Sci* 2010, **21**:1845-1853.

59. van Buren B, Uddenberg S, Scholl BJ: **The automaticity of perceiving animacy: Goal-directed motion in simple shapes influences visuomotor behavior even when task-irrelevant**. *Psychonomic Bull Rev* 2016, **23**:797-802.

60. Scholl BJ, Tremoulet PD: **Perceptual causality and animacy**. *Trends Cognit Sci* 2000, **4**:299-309.

61. Malle BF: *How the mind explains behavior: Folk explanations, meaning, and social interaction*. MIT Press; 2006.

62. Heider F: *The psychology of interpersonal relations*. Psychology Press; 2013.

63. Sutton RS, Barto AG: *Reinforcement learning: An introduction*. MIT Press; 2018.

64. T. D. Kulkarni, K. Narasimhan, A. Saeedi, J. Tenenbaum, Hierarchical deep reinforcement learning: Integrating temporal abstraction and intrinsic motivation, in: Advances in neural information processing systems, 3675-3683.

65. J. Andreas, D. Klein, S. Levine, Modular multitask reinforcement learning with policy sketches, arXiv preprint arXiv:1611.01796 (2016).

66. Piantadosi ST, Tenenbaum JB, Goodman ND: **The logical primitives of thought: Empirical foundations for compositional cognitive models**. *Psychol Rev* 2016, **123**:392.

67. Battaglia PW, Hamrick JB, Tenenbaum JB: **Simulation as an engine of physical scene understanding**. *Proc Natl Acad Sci* 2013:201306572.

68. Pylyshyn ZW: *Computation and cognition*. Cambridge, MA: MIT press; 1984.

69. Wimsatt WC, False models as means to truer theories, Neutral models in biology (1987) 23-55.

70. Forster M, Sober E: **How to tell when simpler, more unified, or less ad hoc theories will provide more accurate predictions**. *Br J Philosophy Sci* 1994, **45**:1-35.