Jennifer Nagel, University of Toronto
June 7, 2023

Lecture 4: Distinctively human mindreading

Curiosity accelerates individual learning by motivating animals to seek out surprising experiences. Social cognition accelerates individual learning by enabling animals to harvest value from each other's current perceptions and stored knowledge. Humans combine these learning accelerants.

Just as curiosity-driven spatial exploration leads animals like rats to construct maps of their environments, so also curiosity-driven social interaction enables us to map out what others do and do not know. While other animals make opportunistic use of each other's intelligence, humans make deliberate and strategic use of it. By actively combining our epistemic forces, humans enter states of joint attention and gain common knowledge.

1. Model-free and model-based reinforcement learning

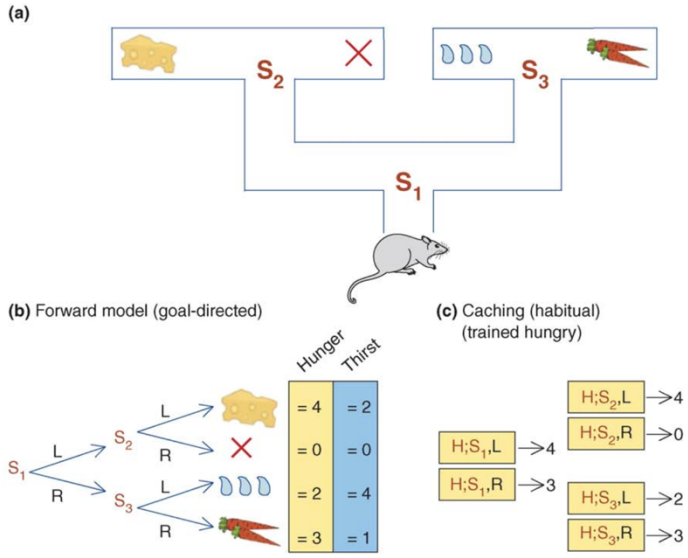

Image source: (Niv, Joel et al. 2006, p.376)

In **model-free RL**, value functions are updated purely on the basis of experienced state-action-reward contingencies, and 'cached' at choice points. Model-free RL governs habitual action (e.g. driving a familiar route on 'mental autopilot').

In **model-based RL**, still on the basis of reward, agents learn relationships between states; they can use the resulting forward model of the task domain to simulate and evaluate novel state-action sequences. Model-based RL enables flexible planning (e.g. figuring out a new way home when you see a roadblock on the horizon).

Many animals, including rats, have both systems for spatial navigation, and trade off between them roughly as needed.

Peter Dayan: model-free control is procedural in character: "it specifies directly the choice of action at each state or location as an imperative command." By contrast, model-based control is declarative: it "provides a set of (semantic) facts about the structure of the environment and the subject in the form of a forward or generative model"; ideally these facts will entail that some particular action is optimal (Dayan 2009: 214).

Successful model-based RL results in states that are robustly accurate. Not everything that the agent incorporates into its model necessarily constitutes knowledge; especially in the early stages of training, the agent can be expected to model various misconceptions about its environment. However, knowledge is the natural endpoint of model-based reinforcement learning: safely accurate representations become basins of attraction stabilized by their propensity to support the capture of reward.

2. Gaze following and shared attention

Two-month-old infants are sensitive to the direction an agent is facing, but eight-month-olds still follow an adult's head turns when her eyes are closed. Between nine and ten months of age there is a sudden improvement in performance ("the nine month revolution"), with infants now following gaze direction only if the adult's eyes are open (Brooks and Meltzoff 2005).

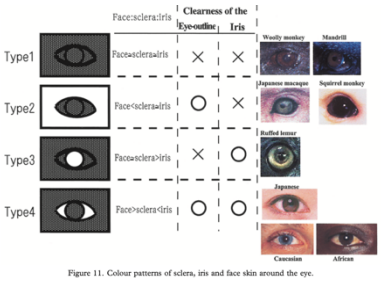

Figure 11. Colour patterns of sclera, iris and face skin around the eye.

Mental state sharing in competition versus cooperation: "Whereas during competition individuals read the minds of their competitors against the competitor's will (when we are competing, I want to conceal my mental states from you), in cooperation and coordination individuals want their partner to read their minds (when we are cooperating and coordinating, I do everything I can to display or advertise my mental states to you to facilitate the process)" (Tomasello 2018, 7).

Image source: (Kobayashi and Kohshima 2001, p.432)

Across cultures, infants begin to use a hand or finger to point objects out to their caregivers between 10 and 14 months of age, often echoing the pointing gestures of their caregivers, and looking back and forth between the object and the caregiver (Liszkowski, Brown et al. 2012). By contrast, chimpanzees do not communicate with each other by pointing (Pika and Liebal 2006).

Left: Stimulus material for (Liszkowski, Brown et al. 2012, image from p.703).

3. Proximal mechanisms of shared attention

Chimpanzee infant-caregiver dyads engage in mutual gaze, and chimpanzees can follow gaze, but they fail to integrate these behaviors: they do not engage in triadic engagement, even when human caregivers try to initiate them (Tomonaga, Tanaka et al. 2004). A chimpanzee will move into another's field of view or provoke another to look at him (Liebal, Call, & Tomasello, 2004), apes use attention-getting moves prior to their deliberate gestures, to ensure that those gestures get seen (Warren, Call et al. 2023), but they will not get others to look at some third object.

Image source: (Mundy and Newell 2007)

Uniquely, humans share our curiosity with each other, vocalizing our surprise overtly. You can double your fun with a surprising item you've seen by drawing it to the attention of another and watching their reaction. A pair of humans who actively engage in surprising each other can make rapid gains, moving towards common knowledge by actively seeking out anything that will produce violation of expectancy in either of them. For us, seeing is a social activity.

We need to have prior expectations in order to violate them, so individual curiosity can drive this social behavior only among agents who have a broad but not unmanageable array of possible reactions. The human communicative repertoire is wide enough to make reactions to novel objects will be hard to predict. Caregivers can elicit emotional reactions in the infant, which may be appropriately hard for caregivers to predict because infants have an inchoate understanding of the world. Infants can elicit appraisals and naming of objects by the caregiver.

Infant pointing gets strong maternal response, with a verbal response for 'proto-declarative pointing' 93% of the time (Olson and Masur 2011). This verbal response usually includes some kind of labelling – Olson and Masur found 62% of responses including 'at least one object, action or internal state label' for pointing. The 'internal state' labels virtually always referenced seeing for pointing gestures, and wanting for reaching gestures. The mother's lexicalization of the state of seeing ("oh, you see the balloon!") may aid the child's understanding of that mental state (Taumoepeau and Ruffman 2006, Taumoepeau and Ruffman 2008). Meanwhile, infant pointing significantly predicts vocabulary size, with maternal responsiveness enhancing this effect (Rowe and Goldin-Meadow 2009, Kirk, Donnelly et al. 2022).

"Utterances containing only single word expressives or fillers (e.g., oh, hey, uh-oh) were not included in the analysis as the maternal utterances analysis focused only on linguistic items with semantic content" (Cameron-Faulkner, Malik et al. 2021, 278). "Oh" is the default expression of surprise in English; it's also very heavily used, showing up about more than once a minute in adult conversation (Reece, Cooney et al. 2023, Table 5).

4. Detecting states of knowledge and belief

Typically developing children pass the Knowledge Access task between ages 3-4 and the False Belief task about a year later (Wellman, Cross et al. 2001, Callaghan, Rochat et al. 2005, Liu, Wellman et al. 2008, Rakoczy 2017, Rakoczy 2022). A wealth of other abilities (hypothetical reasoning, dual naming, etc.) appear at the age of false belief understanding (reviewed in Tomasello 2019).

5. Separating knowledge attribution and belief attribution

In Fabricius's Gettier case version of the unexpected contents task: the experimenter shows the child a familiar brand-name candy package and asks her to say what it contains. The child is then shown that the package contains not the expected candy, but a pencil. The experimenter then removes the pencil, in front of the child, and refills the package with the appropriate brand of candy. The child is then asked the following questions: "What did we put inside the bag [Control Question 1]? What was inside the bag first [Control Question 2]? I have a friend waiting outside the room. It is Elmo. He hasn't seen inside this bag. When he first looks at the bag, before he opens it, will he think there are M&M's or a pencil [a pencil or M&M's, counterbalanced] inside [belief question]? Why will Elmo think that [justification question]?" (Fabricius, Boyer et al. 2010, 1405).

Belief question: "When he first looks at the bag, before he opens it, will he think there are M&M's or a pencil?" Over 80% of the youngest children in the study (roughly 3.5 years old) report that Elmo will think that there are M&Ms in the M&Ms bag. Performance then worsened, so that by 4.5 years of age, children are below chance on this true belief task; only at age 6.5 does performance recover fully.

The 'justification' question: "Why will Elmo think that?"
The youngest children: "There's M&Ms inside"; "He wants them." (reality/goal reasoning)
The oldest children: "It has M&Ms on it"; "It's an M&Ms bag". (belief reasoning)
The middle group: "He didn't see where you put the M&Ms", "He won't know M&Ms are in there" "He doesn't know so he'll have to guess". (Recognition of Gettiered agent as ignorant; no belief reasoning.)

The absence of knowledge was salient in responses to the justification question, across all ages: "Child 1: "He hasn't looked inside; he doesn't know what's inside; he won't know if it's a pencil or not;" Child 2: "He doesn't know what it is;" Child 3: "He doesn't know if it's M&M's or a pencil;" Child 4: "He'll think maybe M&M's, because there could be M&M's and there could not be" (Fabricius, Boyer et al. 2010, 1405).

Alan Leslie: children reason in terms of beliefs from an early age, but initially fail the false belief task because they rely on the "true belief default", according to which, "because people's mundane beliefs are usually true, the best guess about another person's belief is that it is the same as one's own." (Leslie, Friedman et al. 2004).

6. Model-based mindreading, conversation, and belief attribution
Mark Ho, Rebecca Saxe and Fiery Cushman on the view that the core function of theory of mind is to predict action by inferring mental states: "this is like saying that a theory of cooking is useful mostly for guessing what is coming out of the kitchen. Theories are useful not only for passively predicting the future, but also for actively planning to change it. Just as we can plan and cook a meal, we can use Theory of Mind to plan to change people's actions and feelings, by intervening on their perceptions, beliefs, and desires" (Ho, Saxe et al. 2022, p.959).

The new CANDOR corpus of conversation shows about a thousand words of backchannel for every hour of conversation, appearing on about two-thirds of conversational turns that are five words or more. "Oh" is the third-most common backchannel word (7.3%), after "yeah" (39.9%) and "mhm" (15.3%) (Reece, Cooney et al. 2023).

Dialogue 1:	*Dialogue 2:*
Alice: Carlos is having a party Friday night. Bob: Oh. Are you going?	Alice: Carlos is having a party Friday night. Brad: Yeah. Are you going?

Backchannel devices in other languages work similarly (e.g. Koivisto 2016, Wu and Heritage 2017). Conversation may also be driven by reciprocity-related motivation to interest and be interested (Tantucci, Wang et al. 2022). We also converse for instrumental purposes, but the curiosity-dominated model of conversation does something to explain certain otherwise puzzling aspects of human language use, including the volume of talk (Mehl & Pennebaker 2003), and the fact that little speaking time is spent on practical plans (Dunbar, Marriott et al. 1997, Dessalles 2020).

Models can be inverted: having developed a map of a target's epistemic and motivational states, you can predict action, or invert the model to infer the mental states that makes sense of an observed action (Jara-Ettinger 2019). Once factive mental states are detected, you can also invert the model to infer a state of the world when you see an action (if Martha knows where the cake is, watch where she goes when she wants cake).

Models can be used to generate simulated experience: so, we can simulate the experience that would make sense of someone's otherwise puzzling action – because the person who is reaching for the empty container is doing something that would make sense if he knew the reward were located there, we can see him as believing that this is the case. This simulated experience is decoupled content, distinct from our primary representation of reality (Westra and Nagel 2021).

Belief attribution gains value when mindreaders are tracking not just simple agents who act in response to shared physical reality, but also complex agents who can think about decoupled contents, express them, and act strategically. A chimpanzee can waste its own time acting on a misconception: an ignorant human can communicate that misconception, wasting your time, too. Other animals can send mistaken signals, but only within a range fixed by evolutionary pressures to curb the costs of false alarms. Humans can send a vastly wider array of signals, and our decoupled reasoning also enables us to pursue them further, compounding earlier mistakes by drawing unwarranted inferences. We have more material to work with here, and the stakes are higher, both for figuring out whether other agents are knowledgeable, and for figuring out what it is that they believe. This is a project of daunting computational complexity; next time we'll see how it is executed.

References:

Aboody, R., H. Huey and J. Jara-Ettinger (2022). "Preschoolers decide who is knowledgeable, who to inform, and who to trust via a causal understanding of how knowledge relates to action." Cognition **228**: 105212.

Brooks, R. and A. N. Meltzoff (2005). "The development of gaze following and its relation to language." Developmental science **8**(6): 535-543.

Buckner, C. (forthcoming). "A Forward-Looking Theory of Content." Ergo, an Open Access Journal of Philosophy.

Callaghan, T., P. Rochat, A. Lillard, M. L. Claux, H. Odden, S. Itakura, S. Tapanya and S. Singh (2005). "Synchrony in the Onset of Mental State Reasoning." Psychological Science **16**(5): 378-384.

Cameron-Faulkner, T., N. Malik, C. Steele, S. Coretta, L. Serratrice and E. Lieven (2021). "A cross-cultural analysis of early prelinguistic gesture development and its relationship to language development." Child development **92**(1): 273-290.

Dayan, P. (2009). "Goal-directed control and its antipodes." Neural Networks **22**(3): 213-219.

Dessalles, J.-L. (2020). "Language: The missing selection pressure." Theoria et Historia Scientiarum **XVII**: 7-57.

Dretske, F. (1983). Knowledge and the Flow of Information. Cambridge, MIT Press.

Dunbar, R. I. M., A. Marriott and N. D. C. Duncan (1997). "Human conversational behavior." Human Nature **8**(3): 231-246.

Fabricius, W. V., T. W. Boyer, A. A. Weimer and K. Carroll (2010). "True or false: Do 5-year-olds understand belief?" Developmental Psychology **46**(6): 1402.

Ho, M. K., R. Saxe and F. Cushman (2022). "Planning with theory of mind." Trends in Cognitive Sciences **26**(11): 959-971.

Jara-Ettinger, J. (2019). "Theory of mind as inverse reinforcement learning." Current Opinion in Behavioral Sciences **29**: 105-110.

Kirk, E., S. Donnelly, R. Furman, M. Warmington, J. Glanville and A. Eggleston (2022). "The relationship between infant pointing and language development: A meta-analytic review." Developmental Review **64**: 101023.

Kobayashi, H. and S. Kohshima (2001). "Unique morphology of the human eye and its adaptive meaning: comparative studies on external morphology of the primate eye." Journal of human evolution **40**(5): 419-435.

Koivisto, A. (2016). "Receipting information as newsworthy vs. responding to redirection: Finnish news particles aijaa and aha (a)." Journal of Pragmatics **104**: 163-179.

Leslie, A., O. Friedman and T. German (2004). "Core mechanisms in theory of mind." Trends in cognitive sciences **8**(12): 528-533.

Liszkowski, U., P. Brown, T. Callaghan, A. Takada and C. De Vos (2012). "A prelinguistic gestural universal of human communication." Cognitive Science **36**(4): 698-713.

Liu, D., H. M. Wellman, T. Tardif and M. A. Sabbagh (2008). "Theory of mind development in Chinese children: A meta-analysis of false-belief understanding across cultures and languages." Developmental Psychology **44**(2): 523.

Marcos, H. (1991). "How adults contribute to the development of early referential communication?" European Journal of Psychology of Education **6**: 271-282.

Millikan, R. G. (1987). Language, thought, and other biological categories: New foundations for realism, MIT press.

Mundy, P. and L. Newell (2007). "Attention, joint attention, and social cognition." Current directions in psychological science **16**(5): 269-274.

Niv, Y., D. Joel and P. Dayan (2006). "A normative perspective on motivation." Trends in cognitive sciences **10**(8): 375-381.

Olson, J. and E. F. Masur (2011). "Infants' gestures influence mothers' provision of object, action and internal state labels." Journal of Child Language **38**(5): 1028-1054.

Peterson, C. C., H. M. Wellman and D. Liu (2005). "Steps in theory-of-mind development for children with deafness or autism." Child development **76**(2): 502-517.

Pika, S. and K. Liebal (2006). Differences and similarties between the natural gestural communication of the great apes and human children. The evolution of language, World Scientific**:** 267-274.

Rakoczy, H. (2017). "In defense of a developmental dogma: Children acquire propositional attitude folk psychology around age 4." Synthese **194**(3): 689-707.

Rakoczy, H. (2022). "Foundations of theory of mind and its development in early childhood." Nature Reviews Psychology: 1-13.

Reece, A., G. Cooney, P. Bull, C. Chung, B. Dawson, C. Fitzpatrick, T. Glazer, D. Knox, A. Liebscher and S. Marin (2023). "The CANDOR corpus: Insights from a large multimodal dataset of naturalistic conversation." Science Advances **9**(13): eadf3197.

Rowe, M. L. and S. Goldin-Meadow (2009). "Differences in early gesture explain SES disparities in child vocabulary size at school entry." Science **323**(5916): 951-953.

Tantucci, V., A. Wang and J. Culpeper (2022). "Reciprocity and epistemicity: On the (proto) social and cross-cultural 'value'of information transmission." Journal of Pragmatics **194**: 54-70.

Taumoepeau, M. and T. Ruffman (2006). "Mother and infant talk about mental states relates to desire language and emotion understanding." Child development **77**(2): 465-481.

Taumoepeau, M. and T. Ruffman (2008). "Stepping stones to others' minds: Maternal talk relates to child mental state language and emotion understanding at 15, 24, and 33 months." Child development **79**(2): 284-302.

Tomasello, M. (2018). "How children come to understand false beliefs: A shared intentionality account." Proceedings of the National Academy of Sciences **115**(34): 8491-8498.

Tomasello, M. (2019). Becoming human: A theory of ontogeny, Belknap Press.

Tomonaga, M., M. Tanaka, T. Matsuzawa, M. Myowa-Yamakoshi, D. Kosugi, Y. Mizuno, S. Okamoto, M. K. Yamaguchi and K. A. Bard (2004). Development of social cognition in infant chimpanzees (Pan troglodytes): Face recognition, smiling, gaze, and the lack of triadic interactions 1, Wiley Online Library.

Warren, E., J. Call and G. Gergely (2023). "On the murky dissociation between expression and communication." Behavioral and Brain Sciences **46**: 44-45.

Wellman, H., D. Cross and J. Watson (2001). "Meta analysis of theory of mind development: The truth about false belief." Child Development **72**(3): 655-684.

Westra, E. and J. Nagel (2021). "Mindreading in conversation." Cognition **210**: 1-15.

Wu, R.-J. R. and J. Heritage (2017). Particles and epistemics: Convergences and divergences between English and Mandarin. Enabling human conduct, John Benjamins**:** 273-298.