From Theory of Mind to Theory of Environment: Counterfactual Simulation of Latent Environmental Dynamics

Ryutaro Uchiyama

Singapore University of Technology and Design
ryutaro_uchiyama@sutd.edu.sg

Abstract

The vertebrate motor system employs dimensionality-reducing strategies to limit the complexity of motor coordination, thereby facilitating efficient motor control. But when environments are dense with hidden action–outcome contingencies, motor complexity can promote behavioral innovation. Humans, perhaps uniquely, infer this density of latent environmental dynamics from social cues, by drawing upon computational mechanisms shared with Theory of Mind. This proposed "Theory of Environment" supports behavioral innovation by expanding the dimensionality of motor exploration.

The problem of behavioral innovation

The flexibility and creativity of human behavior remain an enigma. Theories of cultural evolution explain how the emergence of conformism and imitation enabled behavioral innovations to persist and cumulate across generations, resulting in the ecological success of our species (Henrich 2016). Behavioral innovation itself, however, remains poorly understood, often relying on assumptions of random variation. Here we propose a novel socio-cognitive mechanism, grounded in Theory of Mind computation (Barnby et al. 2024), that helps bridge this explanatory gap.

The human brain controls roughly 600 muscles and 350 joints to generate desirable outcomes in a 3-dimensional space, yielding a highly redundant system in which a given action objective can be realized by a vast number of possible motor configurations. To reduce this sprawling complexity (Bernstein 1967), the vertebrate motor system organizes muscular activation into coordinated "muscle synergies" (Overduin et al. 2008) that impose strategic low-dimensional constraints onto high-dimensional biomechanics. By limiting the variability of movement coordination, such strategic adaptations inhibit open-ended exploration of possible movement-coordination structures, but in turn facilitate efficient whole-body control. Open-ended behavioral exploration is generally a costly investment, as evolution optimizes for multiplicative (i.e., geometric mean) fitness, where a single zero-fitness episode wipes out all prior gains. Assumptions of additive utility in reinforcement learning hence underestimate this vulnerability to exploration risk. The limited behavioral repertoire of non-human primates (Tennie, Call, and Tomasello 2009) represents not a functional deficit, but rather a hard-won solution to the problem of motor complexity.

Recently, leading research groups in human evolutionary biology (Morgan and Feldman 2024) and cognitive science (Chu, Tenenbaum, and Schulz 2024) have independently argued that the species-unique feature of human behavior is its open-ended variability. Such claims suggest that humans may have innovated the means to "unbind" acquired constraints on movement degrees-of-freedom – thus becoming able to not only reduce but also *expand* motor exploration complexity. Recent approaches in the movement sciences illustrate how such increases in the dimensional complexity of motor coordination can facilitate skill acquisition (Dhawale, Smith, and Ölveczky 2017; Seifert et al. 2016).

Real ecological environments typically contain an unbounded number of hidden action–outcome contingencies (environmental dynamics) that can be potentially unlocked by skill acquisition – constituting an open-ended search space. These latent environmental goals (Molinaro et al. 2024) can offset the cost of open-ended behavioral exploration, thus incentivizing learners to unbind their motor degrees-of-freedom, rather than remain locked into a low-dimensional repertoire optimized for known goals. The density of these latent goals ("teleological depth") in a given environment indexes the scope of potential future gains in the *controllability* of environmental outcomes (Ligneul et al. 2022; Mancinelli, Roiser, and Dayan 2021) as a function of skill development. But how does an agent assess the teleological depth of a given environment and adaptively calibrate the complexity of behavioral exploration?

A four-fold typology of social goal inference

Echoing Vygotsky (1980) as well as Blackburne, Frith, and Yon (2025), we argue that the socio-cultural environment facilitates behavioral exploration, and does so by supplying teleological "depth cues". Without these social cues, the density of latent goals in a given environment could only be evaluated by actual open-ended exploration, incurring significant vulnerability to exploration risk. We refer to this social inference of teleological depth as *theory of environment* (ToE), and situate it in a 2×2 typology with other better studied mechanisms of social goal inference (*Figure 1*):

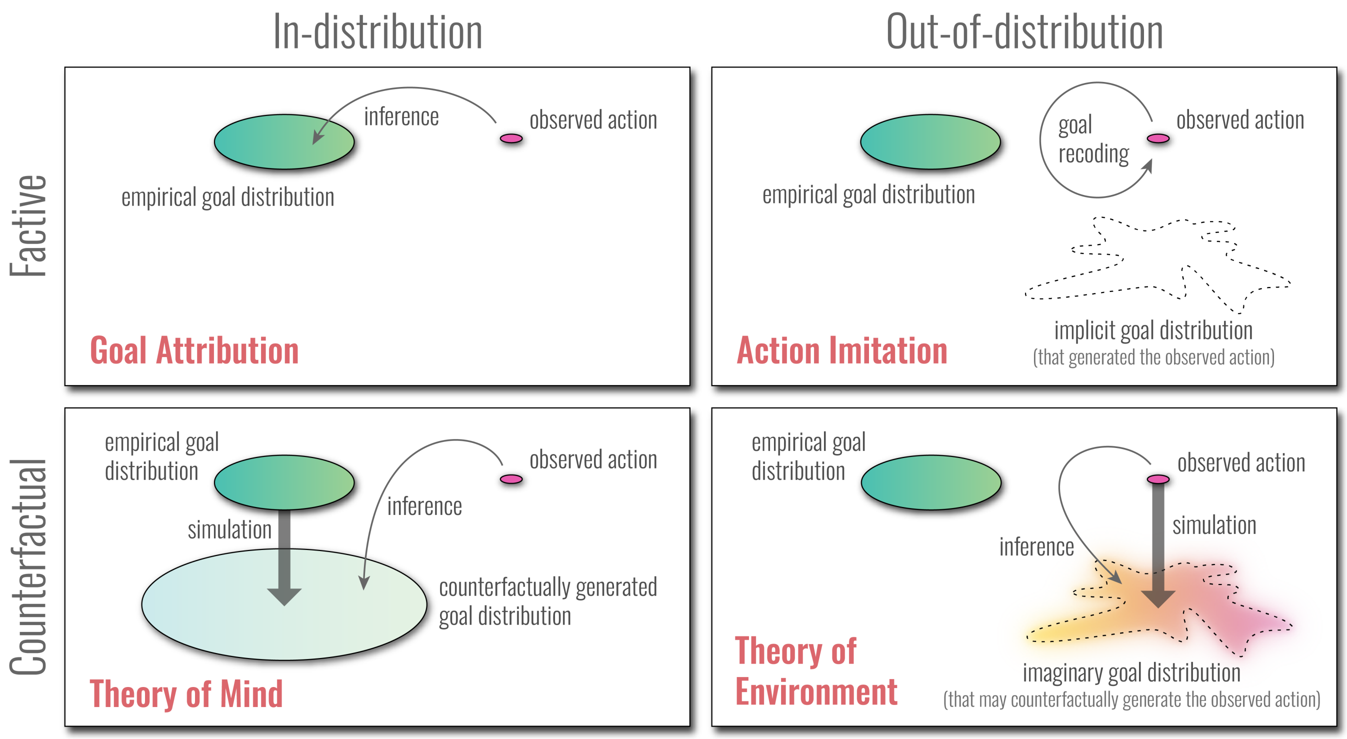

Figure 1: Four modalities of social goal inference, identifying qualitative variation along two representational dimensions: (a) *factive* vs. *counterfactual*, pertaining to the omission/use of counterfactual simulation; and (b) *in-* vs. *out-of-distribution*, pertaining to the scope of goal inference – either bounded or unbounded by the current known hypothesis space.

1. Goal attribution: From the first year of life, human infants expect others' actions to be goal-directed. Infants are prolific in their attribution of goals not only to observed behaviors, but also to artifactual and natural objects, for example when interpreting the agentic purpose of wrenches or clouds (Kelemen 1999).

2. Theory of Mind (ToM): When observing an agent who acts upon a false belief, simple goal attribution is thwarted, instead requiring "meta-representation" of hidden mental states and counterfactual goals, i.e., theory of mind (ToM). Full-fledged ToM appears later in development than goal attribution (Gergely and Csibra 2003), and is is observed reliably only in humans. Some non-human primates make use of a simpler, "factive" ToM that circumvents the cost of counterfactual simulation (Phillips et al. 2021). Due in part to this cost of counterfactuals, hypothesis-generation in ToM is constrained to the well-defined ("in-distribution") domain of known goals. This limitation is shared by *inverse reinforcement learning* (IRL) – a common algorithmic approximation of ToM (Baker et al. 2017). IRL scales poorly in complex environments, and is often restricted to closed-ended task domains. ToM is thus inadequate when observing someone posting mail, if the observer lacks prior knowledge of the environmental dynamics of mail service. A cumulatively cultural species is guaranteed regular encounters with such *causally opaque* behavior (Henrich 2016), suggesting the need for "out-of-distribution" inference mechanisms.

3. Action Imitation: Imitation can be seen as the *recoding* of an observed action into a novel goal unto itself (Lyons, Young, and Keil 2007; Schachner and Carey 2013) – mechanistically consistent with *hindsight relabeling* methods in goal-conditioned RL (Andrychowicz et al. 2017). Imitation circumvents the cost of counterfactual generation, making it a sample-efficient mechanism for out-of-distribution learning, akin to *episodic control* (Lengyel and Dayan 2008). But being tethered to literal observations, imitation lacks the generative flexibility of counterfactual simulation.

4. Theory of Environment (ToE): In our mail-posting example, ToM fails to converge on an adequate resolution, but this computational failure can itself serve as a valuable cue – to switch from postulation of hidden mental states to postulation of hidden environmental dynamics. Both ToM and ToE depend upon counterfactual generation, and likely draw upon a common computational machinery. But whereas ToM resolves ambiguity by exploring a known (in-distribution) hypothesis space, ToE can do so only by actual behavioral exploration – a costlier, more open-ended search domain. ToE generates *out-of-distribution counterfactuals*, similar to imagination. With these representational features, ToE incentivizes the expansion of motor complexity, to facilitate discovery of novel environmental dynamics.

In sum – the computational mechanisms of ToM may have purpose beyond mentalization. ToE deploys counterfactual generation of possible worlds, to support open-ended behavioral exploration. Cultural evolutionary theories commonly model social and asocial learning as exploitation and exploration, respectively. Our proposal decomposes this dichotomy into a new theoretical combination, by postulating a socio-cultural basis for human environmental exploration.

References

Andrychowicz, M.; Crow, D.; Ray, A.; Schneider, J.; Fong, R.; Welinder, P.; McGrew, B.; Tobin, J.; Abbeel, O. P.; and Zaremba, W. 2017. Hindsight Experience Replay. In *Advances in Neural Information Processing Systems*, volume 30.

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

Barnby, J. M.; Alon, N.; Bellucci, G.; Schilbach, L.; Frith, C. D.; and Bell, V. 2024. A Standard Framework for Social Cognition: interoperable algorithms for inference and representation.

Bernstein, N. A. 1967. *The co-ordination and regulation of movements*. Pergamon Press.

Blackburne, G.; Frith, C. D.; and Yon, D. 2025. Communicated priors tune the perception of control. *Cognition*, 254: 105969.

Chu, J.; Tenenbaum, J. B.; and Schulz, L. E. 2024. In praise of folly: flexible goals and human cognition. *Trends in Cognitive Sciences*, 28(7): 628–642.

Dhawale, A. K.; Smith, M. A.; and Ölveczky, B. P. 2017. The Role of Variability in Motor Learning. *Annual Review of Neuroscience*, 40(1): 479–498.

Gergely, G.; and Csibra, G. 2003. Teleological reasoning in infancy: the naıve theory of rational action. *Trends in Cognitive Sciences*, 7(7): 287–292.

Henrich, J. 2016. *The secret of our success: How culture is driving human evolution, domesticating our species, and making us smarter*. Princeton, NJ: Princeton University Press.

Kelemen, D. 1999. Why are rocks pointy? Children's preference for teleological explanations of the natural world. *Developmental Psychology*, 35(6): 1440–1452.

Lengyel, M.; and Dayan, P. 2008. Hippocampal contributions to control: The third way. In Platt, J. C.; Koller, D.; Singer, Y.; and Roweis, S. T., eds., *Advances in neural information processing systems*, volume 20, 889–896. Curran Associates, Inc.

Ligneul, R.; Mainen, Z. F.; Ly, V.; and Cools, R. 2022. Stress-sensitive inference of task controllability. *Nature Human Behaviour*, 6(6): 812–822.

Lyons, D. E.; Young, A. G.; and Keil, F. C. 2007. The hidden structure of overimitation. *Proceedings of the National Academy of Sciences of the United States of America*, 104(50): 19751–6.

Mancinelli, F.; Roiser, J.; and Dayan, P. 2021. Internality and the internalisation of failure: Evidence from a novel task. *PLOS Computational Biology*, 17(7): e1009134.

Molinaro, G.; Colas, C.; Oudeyer, P.-Y.; and Collins, A. 2024. Latent Learning Progress Drives Autonomous Goal Selection in Human Reinforcement Learning.

Morgan, T. J. H.; and Feldman, M. W. 2024. Human culture is uniquely open-ended rather than uniquely cumulative. *Nature Human Behaviour*.

Overduin, S. A.; d'Avella, A.; Roh, J.; and Bizzi, E. 2008. Modulation of Muscle Synergy Recruitment in Primate Grasping. *The Journal of Neuroscience*, 28(4): 880–892.

Phillips, J.; Buckwalter, W.; Cushman, F.; Friedman, O.; Martin, A.; Turri, J.; Santos, L.; and Knobe, J. 2021. Knowledge before belief. *Behavioral and Brain Sciences*, 44: e140.

Schachner, A.; and Carey, S. 2013. Reasoning about 'irrational' actions: When intentional movements cannot be explained, the movements themselves are seen as the goal. *Cognition*, 129(2): 309–327.

Seifert, L.; Komar, J.; Araújo, D.; and Davids, K. 2016. Neurobiological degeneracy: A key property for functional adaptations of perception and action to constraints. *Neuroscience Biobehavioral Reviews*, 69: 159–165.

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

Vygotsky, L. S. 1980. *Mind in Society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.