Neil Melkot

Professor Callan

Person Perception and Social Cognition

3/17/2025

Infant Social Cognition: Theory-Theory vs. Simulation Theory

The question of how humans come to understand others' minds, even in infancy, has spurred competing theories in developmental psychology. Two prominent accounts are Theory-Theory (TT) and Simulation Theory (ST). TT proposes that humans interpret others using an internal theory of mind framework, akin to a naive scientific theory about how mental states drive behavior (Gopnik and Wellman). ST, by contrast, contends that we understand others by simulating their mental states in ourselves, using our own mind as a model (Harris). The groundbreaking studies by Hamlin, Wynn, and Bloom (2007) and Hamlin (2015), in which even young infants preferentially evaluate "helpful" versus "hindering" individuals, provide a critical test case for these theories. This essay argues that TT offers a superior explanation for these infant social cognition findings than ST. By analyzing the core premises of TT and ST and examining how each would interpret infants' social evaluations, we will see that TT's framework better accommodates the evidence from Hamlin et al. In contrast, ST faces significant developmental and empirical limitations in accounting for the data. Ultimately, the early emergence of infant social evaluation aligns more convincingly with a theory-based understanding of others rather than one grounded in direct simulation, making TT the stronger framework for explaining these results.

Theory‑Theory posits that children are little theorists who construct understandings of other minds in a manner analogous to scientists forming and revising theories (Gopnik and

Wellman). According to this view, championed by scholars like Alison Gopnik and Henry Wellman, children's conceptualizations of people develop by formulating hypotheses about behavior and testing them against experience (Gopnik and Wellman). In other words, from an early age, children gather social information and gradually build an internal "theory of mind" that explains others' actions in terms of mental states (desires, intentions, beliefs) (Gopnik and Wellman). This knowledge is thought to be abstract and systematic, allowing predictions about behavior across situations (Gopnik and Wellman). Crucially, TT holds that these mental-state concepts can exist before complex reasoning fully matures, beginning as rudimentary principles that get refined over time. For instance, toddlers often display an understanding that people act to fulfill visible desires, resembling a causal rule that "if X wants something, and X can see it, X will try to get it" (Gopnik and Wellman). Over development, children update their theory with new evidence, leading to more sophisticated concepts (e.g., grasping false beliefs around age 4) (Gopnik and Wellman). In infancy, TT suggests foundational social-cognitive distinctions—like recognizing a "friend" from a "foe"—may emerge swiftly as part of an innate or early-acquired framework. Proponents of TT contend that certain social evaluations reflect a "biological adaptation" for identifying helpful versus harmful agents (Hamlin et al. 557–560). Under this view, an infant does not need extensive life experience to prefer a helper; a simple internal theory that "agents who help achieve goals are likable" could suffice, allowing infants to interpret abstract social scenarios even before they can speak.

Simulation Theory offers a different account of how we understand others, one grounded in empathy and imagination rather than explicit theorizing. Advocates such as Alvin Goldman and Paul Harris argue that we possess a "working model" of mind in our own introspective experience, and we use it to simulate others' mental processes (Harris). Children (and adults)

essentially project themselves into another's shoes by asking, "What would I do or feel in that situation?" and assuming the other person's mind works similarly (Harris). This process is thought to rely on pretend play and imagination. Studies indicate that by around age 2, children begin to engage in pretend play, and by age 3 they can pretend to have mental states they do not actually possess (Harris). According to Harris (1992), this growing capacity for imagination underpins the child's increasing skill at simulation. Notably, children typically pass false-belief tasks around age 4, aligning with an increased ability to simulate. ST thus suggests that mind-reading improves once children can fully deploy these imaginative skills. How infants understand someone else's mind is less clear under ST, which generally assumes less complex simulation in infancy, possibly relying on innate mirror mechanisms or emotional resonance. Yet proponents acknowledge that ST may not easily explain certain developmental findings, and critics like Rebecca Saxe note that mere projection often fails to account for systematic errors in interpreting others (Saxe). This points to a need for conceptual understanding, as TT proposes. Before exploring ST's limitations, we should review what infants actually do in social evaluation tasks.

In a seminal study, Hamlin, Wynn, and Bloom (2007) presented 6 and 10-month-old infants with puppet-like characters in a simple scenario (Hamlin et al. 2007). One character (the "Climber") struggled to climb a hill, failing repeatedly. On alternating trials, an assisting "Helper" pushed the Climber up, while a "Hinderer" pushed it down. Infants observed these events multiple times as uninvolved onlookers (Hamlin et al. 2007). After habituation, they were prompted to choose between the Helper and the Hinderer. The results were striking: even 6-month-olds robustly preferred the Helper (Hamlin et al. 2007). In a related looking-time test, they also seemed to expect the Climber to approach the Helper, indicating that in the infants'

judgments, helpers are more appealing (Hamlin et al. 2007). Further experiments showed infants preferred a Helper over a neutral character, and a neutral character over a Hinderer, suggesting a hierarchy of 'helping > neutral > hindering' (Hamlin et al. 2007). Such results imply that preverbal infants assess agents based on behavior toward others, which "may serve as the foundation for moral thought and action" (Hamlin et al. 2007). Even before language or complex reasoning, infants distinguish "friend" from "foe," suggesting that social evaluation emerges very early in development.

Subsequent research has extended these findings. By 2015, Hamlin noted a growing body of work showing that even 3-month-olds display preferences for prosocial over antisocial actors (Hamlin 2015). Between 6 and 10 months, infants demonstrate these preferences across multiple scenarios, not just hill-climbing, and they can factor in context (Hamlin 2015). Moreover, they seem to care about intentions, not just outcomes. Hamlin (2015) reports that by the end of the first year, infants "evaluate others based on their prosocial and antisocial mental states rather than the positive and negative outcomes they cause." For instance, a failed helper (good intent, bad outcome) may still be viewed favorably, while a successful hinderer (bad intent, completed outcome) is viewed negatively. This suggests a rudimentary theory of mind in which infants attribute goals and intentions to others and evaluate them accordingly (Hamlin 2015; Hamlin and Wynn, 2011). By the second year, social evaluations grow more sophisticated: toddlers not only prefer helpers, they will actively help those who have helped before, or act antisocially toward former hinderers (Hamlin 2015). These findings indicate that the capacity to distinguish cooperative from obstructive individuals is present quite early, supporting the view that moral cognition has deep roots in infancy.

TT provides a compelling explanation for these infant behaviors. If infants come equipped with an initial "theory" about agents, even a simple one, then their robust preference for helpers follows naturally. The infant sees the Climber's goal (going up the hill) and notices who assists or impedes that goal. Under TT, this is interpreted through a conceptual lens like "helping is positive, hindering is negative." The 2015 experiments emphasize the Climber's gaze, cueing infants to infer a desire to ascend (Hamlin 2015). Recognizing an intention and assessing how others align with that intention is the sort of rule-based process TT predicts. Infants' strong preference for the helper implies a conceptual evaluation, not just a superficial perceptual bias. Indeed, Hamlin's follow-ups ruled out lower-level explanations; the infants were responding to the social dimension of the scene, not just movements (Hamlin 2015). TT further explains how infants prioritize intent over outcome, since only a framework with unseen mental states (like intentions) clarifies why a failed helper might still be liked more than a successful hinderer. TT thus suggests infants interpret others via a rudimentary theory of mind. Hamlin et al. propose that "social evaluation is a biological adaptation," implying an inherent capacity for distinguishing prosocial from antisocial agents (Hamlin et al. 557–560). This aligns with TT's premise that early social reasoning may be guided by an evolved, initial theory about how agents should behave.

In contrast, Simulation Theory struggles to account for many of these data points. According to ST, the infant would solve the task by mentally projecting themselves into the climber's place, then deciding who feels helpful or harmful. But can a 6-month-old truly run such a simulation? Full-fledged simulation presupposes abilities like pretend play and self-other differentiation that typically develop closer to age 3 or 4 (Harris). Although some basic resonance with others' states might exist, ST alone does not clarify how infants systematically

identify a goal (the top of the hill) or infer that helping is good. Nor does it cleanly explain why infants weigh intentions over outcomes. If the infant were merely simulating emotional states, the actual success or failure of climbing might overshadow the actor's intentions. Additionally, scholars like Saxe note that if we relied on simulation alone, we would not see the consistent errors that occur when we try to interpret others—a fact suggesting a need for conceptual principles (Saxe). Hence, ST's account falters when confronted with findings that infants already differentiate prosocial intentions from mere results.

The balance of evidence from Hamlin et al. (2007) and Hamlin (2015) strongly favors Theory‑Theory over Simulation Theory as a framework for infant social cognition. TT explains how even young infants can evaluate others' actions by invoking an embryonic theory of mind that highlights agents, goals, and intentions. Infants judge a helper as "good" and a hinderer as "bad" without needing to place themselves in each role. This early capacity for social judgment aligns with an evolutionary basis for moral reasoning. Simulation Theory, while useful for understanding older children's empathy and role-taking, demands cognitive abilities that infants appear not to possess. In these hill-climbing studies, the infants do more than mirror distress: they make evaluative judgments that hinge on an agent's intent. TT provides a coherent, developmentally plausible account, whereas ST strains to explain the data. Thus, Theory‑Theory emerges as the stronger framework, offering insight into how infants, from the earliest months of life, can distinguish friend from foe in the social world.

Annotated Bibliography

Hamlin, J. K., Wynn, K., & Bloom, P. *Social Evaluation by Preverbal Infants*, Nature, vol. 450, no. 7169, 2007, pp. 557–560.

In this influential study, Hamlin et al. demonstrate that infants as young as 6 months can evaluate individuals based on their social actions. The authors devised a helper-versus-hinderer paradigm and found that infants consistently preferred a helpful character over a hindering one. This paper provides evidence that infants use others' behaviors toward third parties as criteria for preference, suggesting an early capacity for social evaluation. The authors argue that this ability may be an innate foundation for moral cognition, describing it as possibly a "biological adaptation." This source is central to the essay's argument, as it supplies the primary empirical findings that Theory-Theory aims to explain better than Simulation Theory.

Hamlin, J. K. *The Case for Social Evaluation in Preverbal Infants: Gazing Toward One's Goal Drives Infants' Preferences for Helpers over Hinderers in the Hill Paradigm*, Frontiers in Psychology, vol. 5, 2015, Article 1563.

This follow-up research by Hamlin addresses alternative interpretations of the 2007 findings. Specifically, it tests whether infants' helper preference is truly due to social evaluation (understanding the helper's intention relative to a goal) rather than low-level perceptual features. Hamlin reviews a body of evidence that infants, even by 3 to 10 months old, prefer prosocial actors in various scenarios. New experiments in this paper show that infants' preferences depend on the social cue of the climber's goal-directed

gaze, supporting a genuine social-cognitive basis for their evaluations. The paper strengthens the case that infants are sensitive to others' intentions, not just perceptual cues, and thus bolsters the Theory‑Theory interpretation that infants use an intuitive understanding of goals and intentions.

Gopnik, A., and H. M. Wellman. *Why the Child's Theory of Mind Really Is a Theory,* Mind & Language, vol. 7, no. 1–2, 1992, pp. 145–171.

Gopnik and Wellman argue that children's understanding of mind is aptly described as a "theory", an organized, abstract, and causal explanatory system that the child progressively refines. They present developmental evidence to show that conceptual change occurs in children's thinking about minds. This paper underpins the Theory‑Theory perspective by providing the theoretical foundation that the essay applies to infants. Although infants are younger than the ages discussed in this paper, the general concept that even early understandings of agents can be theory-like supports the argument that infants might possess an incipient theory of mind.

Harris, *P. L. Simulation Theory*, 1992.

Paul Harris's work on Simulation Theory posits that children use their capacity for pretend play and introspection to simulate others' mental states. This source offers a basis for understanding how Simulation Theory might explain social cognition. However, its developmental timeline and reliance on mature imaginative processes pose challenges for explaining the sophisticated social evaluations observed in infants. Harris's ideas are used

in the essay to contrast the strengths of Theory‑Theory with the limitations of Simulation Theory.

Saxe, R. *Against Simulation: The Argument from Error*, Trends in Cognitive Sciences, vol. 9, no. 4, 2005, pp. 174–179.

In this article, Rebecca Saxe critiques Simulation Theory by noting that people often make systematic errors in predicting others' mental states, which a pure simulation approach struggles to explain. Saxe argues that if simulation were the sole mechanism, fewer consistent errors would be observed. This critique supports the essay's argument by highlighting a key limitation of Simulation Theory, thereby reinforcing the case for Theory‑Theory as a more robust framework for understanding social cognition.