TOWARD A MENTALISTIC ACCOUNT OF EARLY PSYCHOLOGICAL REASONING

Yuyan Luo

University of Missouri

and

Renée Baillargeon

University of Illinois

Address correspondence to Yuyan Luo, Department of Psychological Sciences, University of Missouri, Columbia, MO 65211; email: luoy@missouri.edu.

TOWARD A MENTALISTIC ACCOUNT OF EARLY PSYCHOLOGICAL REASONING

Key words: infant cognition; psychological understanding; mentalistic reasoning

Abstract

Recent investigations of early psychological understanding have revealed three key findings. First, young infants attribute goals and dispositions to any entity they perceive as an agent, whether human or non-human. Second, when interpreting an agent’s actions in a scene, young infants take into account the agent’s representation of the scene, even if this representation is less complete than their own. Third, at least by the second year of life, infants recognize that agents can hold false beliefs about a scene. Together, these findings support a system-based, mentalistic account of early psychological reasoning.

Our ability to make sense of others' intentional actions rests primarily on our ability to understand the mental states that underlie these actions. Thus, when we watch a character in the television show *Lost* gather food in the jungle, we may attribute to the character a whole host of mental states including goals, dispositions, knowledge or ignorance of certain facts about the scene, and even false or pretend beliefs about the scene. Over the past 15 years, considerable progress has been made in uncovering the developmental roots of these attributions. These advances can be roughly organized around three successive questions, with the answer to each question suggesting a broader and richer characterization of early psychological reasoning.

Can young infants reason about the intentional actions of a non-human agent?

One of the first questions raised in investigations of early psychological reasoning was whether young infants can reason about the actions of not only human but also non-human agents. In this research, an agent is defined as an entity that can detect its environment and exert control over its actions (e.g., Leslie, 1995; Luo, Kaufman, & Baillargeon, 2009).

In a seminal task, Woodward (1998) presented 5- and 6-month-olds with two objects on an apparatus floor: object-A on the right and object-B on the left. During the habituation trials, a human agent's arm and hand reached into the apparatus and grasped object-A. During the test trials, the toys' positions were reversed, and the agent either grasped object-A as before (old-object event) or now grasped object-B (new-object event). The infants looked reliably longer at the new- than at the old-object event, suggesting that (1) during the habituation trials, the infants attributed to the agent a particular disposition, a preference for object-A over object-B, and (2) during the test trials, the infants expected the agent to continue acting on this preference and hence to form the goal of reaching for object-A in its new position. In contrast to these positive results, negative results were obtained in additional experiments in which the human agent was replaced with an occluder shaped like an arm and hand, a rod tipped with a sponge, or a mechanical claw. These results suggested that early psychological reasoning is at first restricted to humans and is only gradually extended to non-human agents. As such, these results supported an *experience-based view* of early psychological reasoning. According to this view, as infants become adept at producing intentional actions (e.g., reaching for objects), they become able to understand similar actions by similar agents, in part due to innate pathways for establishing equivalences between themselves and others (e.g., Meltzoff, 2005; Woodward, 2005).

However, the negative results obtained in the occluder, rod, and claw experiments were open to an alternative interpretation: perhaps the infants looked equally at the new- and old-object events, not because they could not attribute dispositions and goals to a non-human agent, but because they were uncertain whether the occluder, rod, and claw were in fact agents. To examine this interpretation, 5-month-olds were tested in a similar experiment except that the agent was a box with autonomous control over its actions (Luo & Baillargeon, 2005; see Figure 1). The infants first received orientation trials in which the box moved back and forth across the apparatus floor. In the familiarization trials, object-A and object-B were introduced, and the box consistently approached and rested against object-A. In the test trials, the objects' positions were reversed, and the box approached either object-A (old-object event) or object-B (new-object event). The infants looked reliably longer at the new- than at the old-object event, suggesting that they viewed the box as an agent, attributed to the box a preference for object-A, and expected the box to continue acting on this preference in the test trials. These results have now been extended to 3-month-old infants (Luo, 2010).

It might be objected that the positive results reported in this section all lend themselves to a simpler, and much less impressive, interpretation: perhaps infants simply form an association between the agent and object-A, and look longer when shown a test event that deviates from this association. However, additional results from single-object conditions rule out this low-level interpretation (see Figure 1): when object-B is absent during the familiarization trials, infants look equally at the new- and old-object events (e.g., Luo, 2010; Luo & Baillargeon, 2005). In these conditions, infants have no information about which object the agent may prefer in test (i.e. it may again prefer object-A, or it may now prefer object-B), and they therefore look equally at the two test events. This is not to say that infants can never attribute preferences to agents in single-object conditions: just as we might attribute a fondness for apples to an individual who went to great lengths to retrieve the last apple from an apple tree, 3-month-olds attribute a preference for object-A to a box agent, even when object-B is absent, if object-A occupies different positions in the familiarization trials and the box consistently adjusts its actions so as to reach object-A (Luo, 2010).

Can young infants recognize when an agent has an incomplete representation of a scene?

The evidence that young infants can reason about the intentional actions of non-human agents supports a *system-based view* of early psychological reasoning. According to this view, infants are born equipped with a psychological-reasoning system that provides them with a skeletal causal framework for interpreting and predicting the intentional actions of any entity they identify as an agent, whether human or non-human (e.g., Gergely & Csibra, 2003; Johnson, 2005; Leslie, 1995). But what is the nature of this causal framework?

According to one proposal (Gergely & Csibra, 2003), this framework consists of a *teleological* system of action interpretation. In contrast to *mentalistic* reasoning, which involves the attribution of mental states, teleological reasoning deals exclusively with physical variables: when watching an agent perform goal-directed actions in a scene, the teleological system considers the agent's actions, the goal-state the agent achieves, and the physical constraints existing in the scene. Teleological interpretations are guided by a principle of *rationality*, which states that agents select goal-directed actions that are both causally appropriate and reasonably efficient. Thus, if agent-A repeatedly jumps over a barrier and reaches agent-B, infants will expect agent-A to arrive at this goal-state again in the future. Furthermore, the principle of rationality will enable infants to predict how agent-A should do so if the barrier is removed (i.e., agent-A should now travel in a straight line). With development, physical variables become incorporated into a mentalistic system, which makes sense of intentional actions in terms of goals and other mental states.

Are young infants capable of mentalistic or only teleological reasoning? One key assumption about teleological reasoning is that it is reality-based: because teleological interpretations deal exclusively with physical variables, infants should not be able to distinguish their representation of a scene from that of an agent—*reality should be as construed by the infants* (Gergely & Csibra, 2003). However, contrary to this assumption, there is now evidence that even young infants recognize that an agent's representation of a scene may be less complete than their own (e.g., Luo & Baillargeon, 2007; Luo & Johnson, 2009; Tomasello & Haberl, 2003). In one experiment, for example, 6-month-olds received familiarization trials in which a female agent sat between two objects, object-A and object-B, and consistently reached for object-A (Luo & Johnson, 2009; see Figure 2). In one (hidden-object) condition, both objects were visible to the infant, but only object-A was visible to the agent: a tall screen hid object-B from her. In two other conditions, both object-A and object-B were visible to the agent: either the screen was shorter so that object-B protruded above it (short-screen condition), or object-B itself was taller so that it protruded above the tall screen (tall-object condition). Following the familiarization trials, the objects' positions were reversed, the screen was removed, and the agent reached for either object-A (old-object event) or object-B (new-object event). In the short-screen and tall-object conditions, as expected, the infants looked reliably longer at the new- than at the old-object event; in the hidden-object condition, in contrast, the infants looked about equally at the two events. The infants thus realized that the agent's repeated actions on object-A during the familiarization trials could not be interpreted as revealing a preference for object-A over object-B if she could not see object-B. Similar results have been obtained with 5-month-olds in experiments with a box agent (Luo, Choi, & Baillargeon, 2010). Likewise, 12.5-month-olds refrained from attributing a preference for object-A over object-B to a female agent if object-B was hidden from her by a screen—but they did attribute such a preference if the agent was aware of object-B's presence behind the screen, because she had placed it there herself in a previous trial (Luo & Baillargeon, 2007).

Together, these results suggest three conclusions. First, infants recognize that an agent's representation of a scene may be incomplete relative to their own (interestingly, Moll and Tomasello (2004) reported evidence that the converse is also true: when an agent looked behind a screen with expressions of excitement, infants moved around the screen to see what the agent could see). Second, these results argue against traditional characterizations of young children as fundamentally egocentric (e.g., Piaget, 1954). Finally, and most importantly for the present discussion, the results reviewed here argue against the notion that early psychological reasoning is purely teleological in nature: by 5 months of age, if not before, infants distinguish between their representation of a scene and that of an agent. Furthermore, infants consider the agent's representation, rather than their own, when interpreting the agent's actions (e.g., Luo & Johnson, 2009) or when deciding how to respond to the agent (e.g., Tomasello & Haberl, 2003).

A recent experiment extended these efforts by asking whether 16-month-olds attend to what *parts* of objects an agent can see (Luo & Beck, 2010; see Figure 3). In the familiarization trials, a female agent faced a pair of objects, one red and the other black or yellow; she consistently pointed to the red object, suggesting that she preferred red over the other colors. Next, two screens were introduced while the agent was absent; the infant was shown both sides of each screen. In one (red-red) condition, one screen was red on both sides and the other screen was green on both sides. In another (red-beige) condition, one screen was red on the front (the infant's side) and beige on the back (the agent's side), and the other screen was green on the front and beige on the back. Finally, in a third (green-red) condition, one screen was green on the front and red on the back, and the other screen had the reverse colors. During the test trials, the agent returned, and the screens were kept upright so that she could see only their backs. The infants expected the agent to continue acting on her color preference, and they took into account what sides of the screens she could see when interpreting her actions. Thus, the infants expected the agent to point to the red screen in the red-red condition, to either screen in the red-beige condition, and to the screen that was green on the front but red on the back in the green-red condition.

Can young infants attribute false beliefs to agents?

Proponents of mentalistic accounts of early psychological reasoning often assume that at least two subsystems, Subsystem1 (SS1) and Subsystem2 (SS2), are involved in infants' attribution of mental states (e.g., Leslie, 1995; Scott & Baillargeon, 2009). SS1 enables infants to attribute to an agent *motivational* states, which specify the agent's motivation in the scene (e.g., goals, dispositions), as well as *reality-congruent informational* states, which specify what accurate information the agent can gather about the scene through perception, memory, or inference (e.g., knowledge, ignorance). When an agent's representation of the scene is incomplete relative to that of the infant (e.g., the agent cannot see an object that the infant sees), a *masking* mechanism blocks the information that is not available to the agent, enabling the infant to interpret or predict the agent's actions in terms of the remaining, shared information. SS2 extends SS1 in that it enables infants to also attribute *reality-incongruent informational* states to agents. When an agent holds a false or a pretend belief about a scene, so that the agent's representation of the scene is incompatible with that of the infant (e.g., the agent believes that a toy is in location-A, but the infant knows it has been moved to location-B), a *decoupling* mechanism enables the infant to hold in mind a separate representation of the scene that incorporates the agent's false or pretend belief but otherwise functions as expected (e.g., Leslie, 1994).

The evidence (reviewed in the last section) that young infants realize that an agent may be ignorant about some aspect of a scene indicates that SS1 is operational in early infancy. As for SS2, there is now evidence that infants aged 13 to 15 months can attribute false or pretend beliefs to agents, suggesting that SS2 is operational by the second year of life (e.g., Onishi & Baillargeon, 2005; Surian et al., 2007; for a review, see Baillargeon, Scott, & He, 2010). Together, these findings leave open two developmental possibilities: (1) SS1 and SS2 are both operational from an early age or (2) SS2 becomes operational some time after SS1. To decide between these possibilities, researchers are now examining whether infants in the first year can attribute false beliefs to agents (e.g., He & Baillargeon, 2009; Kovács, 2009). Preliminary results with infants aged 7 to 8.5 months suggest that, they, too, can attribute false beliefs to agents. Hopefully, future experiments will be able to test whether 3-month-olds (the earliest age at which infants are currently known to demonstrate SS1 abilities) can do so as well. Positive results would indicate that SS1 and SS2 are both operational early in infancy.

A role for experience

The evidence reviewed in this article suggests that young infants' reasoning about agents' actions is consistent with a system-based, mentalistic account of early psychological reasoning. This is not to say, of course, that experience plays little role in infants' reasoning. To the contrary, experience undoubtedly plays a key role in infants' learning about specific goals (e.g., learning why Mommy sometimes holds a small box to her ear), dispositions (e.g., learning that Daddy prefers coffee over milk), or informational states (e.g., learning about blindfolds or dark glasses; see Meltzoff, 2005). There is also substantial evidence that infants' own experiences with certain actions can inform or bias how they perceive others' actions (e.g., Luo, in press; Woodward, 2005).

To illustrate, in a recent experiment, 8-month-olds received familiarization trials in which a female agent stared intently at object-A as opposed to object-B (Luo, in press; see Figure 4). Prior research indicates that 12-month-olds attribute a preference for object-A to the staring agent, but younger infants do not (see Woodward, 2005). In line with these findings, the 8-month-olds in this experiment looked about equally at the new- and old-object events. One interpretation of this negative result was that the infants did not attribute a preference for object-A to the agent because they could not understand why she did not grasp object-A when she could easily do so. In two additional conditions, infants were provided with an explanation for the staring agent's inaction: she either looked at the objects through a small window (small-window condition), or held an object so that her hands were occupied (hands-occupied condition; see Gergely et al., 2002). In both conditions, the infants attributed to the staring agent a preference for object-A. Together, these results suggest that, whereas 8-month-olds eagerly grasp interesting objects within easy reach and interpret others' actions accordingly, 12-month-olds have learned (perhaps through parental admonitions) that one may sometimes look at, but not touch, interesting objects.

Acknowledgments

The preparation of this article was supported by research funds from the University of Missouri to the first author and by a grant from the National Institute of Child Health and Human Development (HD-021104) to the second author.

References

Baillargeon, R., Scott, R. M., & He, Z. (2010). False-belief understanding in infants. *Trends in Cognitive Sciences*, *14*, 110-118.

Gergely, G., Bekkering, H., & Király, I. (2002). Rational imitation in preverbal infants. *Nature, 415*, 6873.

Gergely, G., & Csibra, G. (2003). Teleological reasoning in infancy: The naïve theory of rational action. *Trends in Cognitive Sciences, 7*, 287-292.

He, Z., & Baillargeon, R. (2009, April). *Evidence for False Belief Understanding in 8.5-Month-Olds in a Combined Anticipatory-Looking and Violation-of-Expectation Task*. Paper presented at the biennial meeting of the Society for Research in Child Development, Denver, CO.

Johnson, S. C. (2005). Reasoning about intentionality in preverbal infants. In P. Carruthers, S. Laurence, & S. Stitch (Eds.), *The Innate Mind: Structure and Contents* (pp. 254-271). New York: Oxford University Press.

Kovács, Á. M. (2009, April). *7-Month-Olds Compute the Beliefs of Other Agents*. Paper presented at the biennial meeting of the Society for Research in Child Development, Denver, CO.

Leslie, A. M. (1994). Pretending and believing: Issues in the theory of ToMM. *Cognition, 50*, 211-238.

Leslie, A. M. (1995). A theory of agency. In D. Sperber, D. Premack & A. J. Premack (Eds.), *Causal cognition: A multidisciplinary debate* (pp. 121-141). Oxford, UK: Clarendon Press.

Luo, Y. (2010). *Three-month-old infants attribute goals to a non-human agent*. Manuscript under review.

Luo, Y. (in press). Do 8-month-old infants consider situational constraints when interpreting others’ gaze as goal-directed action? *Infancy*.

Luo, Y., & Baillargeon, R. (2005). Can a self-propelled box have a goal? Psychological reasoning in 5-month-old infants. *Psychological Science, 16*, 601-608.

Luo, Y., & Baillargeon, R. (2007). Do 12.5-month-old infants consider what objects others can see when interpreting their actions? *Cognition, 105*, 489-512.

Luo, Y., & Beck, W. (2010). Do you see what I see? Infants' reasoning about others' incomplete perceptions. *Developmental Science*, *13*, 134-142.

Luo, Y., Choi, Y., & Baillargeon, R. (2010). *Young infants' reasoning about a non-human agent’s motivational and informational states*. Manuscript in preparation.

Luo, Y., & Johnson, S. C. (2009). Recognizing the role of perception in action at 6 months. *Developmental Science, 12*, 142-149.

Luo, Y., Kaufman, L., & Baillargeon, R. (2009). Young infants’ reasoning about events involving inert and self-propelled objects. *Cognitive Psychology*, *58*, 441-486.

Meltzoff, A. M. (2005). Imitation and other minds: The “like me” hypothesis. In S. Hurley & N. Chater (Eds.), *Perspectives on imitation: From neuroscience to social science* (Vol. 2, pp. 55-77). Cambridge, MA: MIT Press.

Moll, H., & Tomasello, M. (2004). 12- and 18-month-old infants follow gaze to spaces behind barriers. *Developmental Science*, *7*, F1-F9.

Onishi, K. H., & Baillargeon, R. (2005). Do 15-month-old infants understand false beliefs? *Science*, *308*, 255-258.

Piaget, J. (1954). *The construction of reality in the child*. New York: Basic.

Scott, R. M., & Baillargeon, R. (2009). Which penguin is this? Attributing false beliefs about object identity at 18 months. *Child Development*, *80*, 1172-1196.

Surian, L., Caldi, S., & Sperber, D. (2007). Attribution of beliefs by 13-month-old infants. *Psychological Science, 18*, 580-586.

Tomasello, M., & Haberl, K. (2003). Understanding attention: 12- and 18-month-olds know what is new for other persons. *Developmental Psychology*, *39*, 906-912.

Woodward, A. L. (1998). Infants selectively encode the goal object of an actor's reach. *Cognition, 69*, 1-34.

Woodward, A. L. (2005). The infant origins of intentional understanding. In R. V. Kail (Ed.), *Advances in child development and behavior* (Vol. 33, pp. 229-262). Amsterdam: Elsevier.

Recommended Readings

Johnson, S. C. (2005). (See References)

Luo, Y., Kaufman, L., & Baillargeon, R. (2009). (See References)

Scott, R., & Baillargeon, R. (2009). (See References)

Wynn, K. (2008). Some innate foundations of social and moral cognition. In P. Carruthers, S. Laurence & S. Stich (Eds.), *The Innate Mind: Foundations and the Future*. Oxford: Oxford University Press.

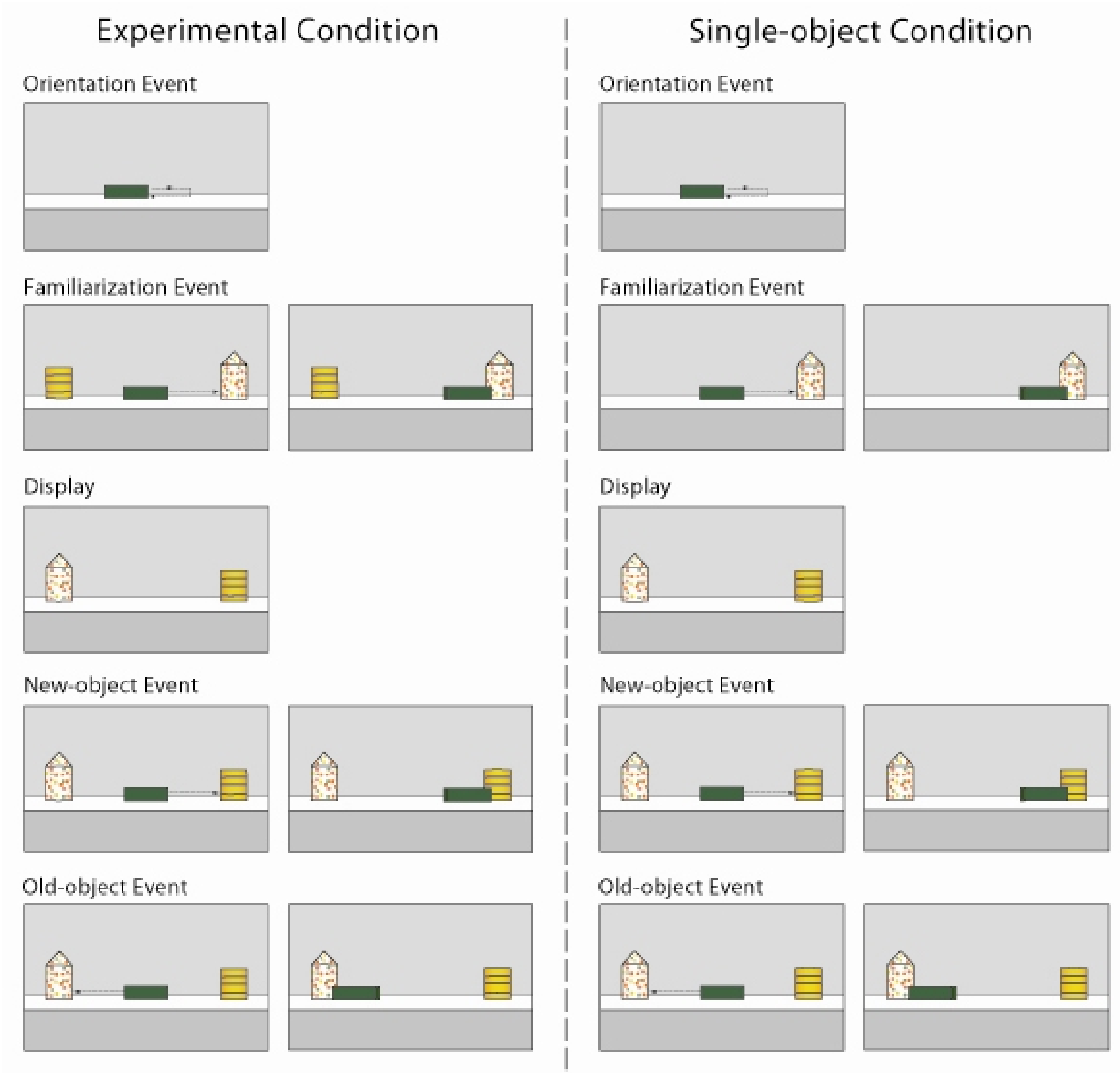
Figure Captions

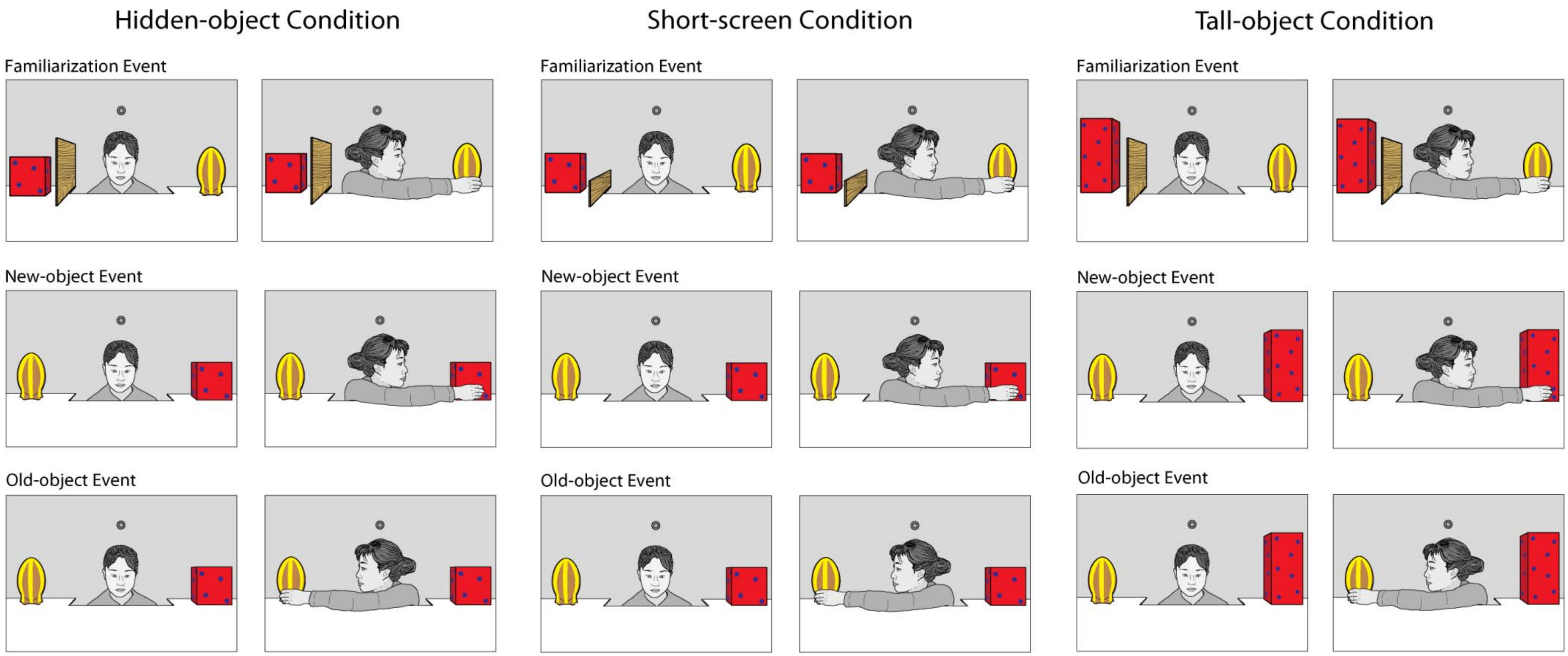
Figure 1. Schematic drawing of the events shown in the experimental and single-object conditions of Luo and Baillargeon (2005).

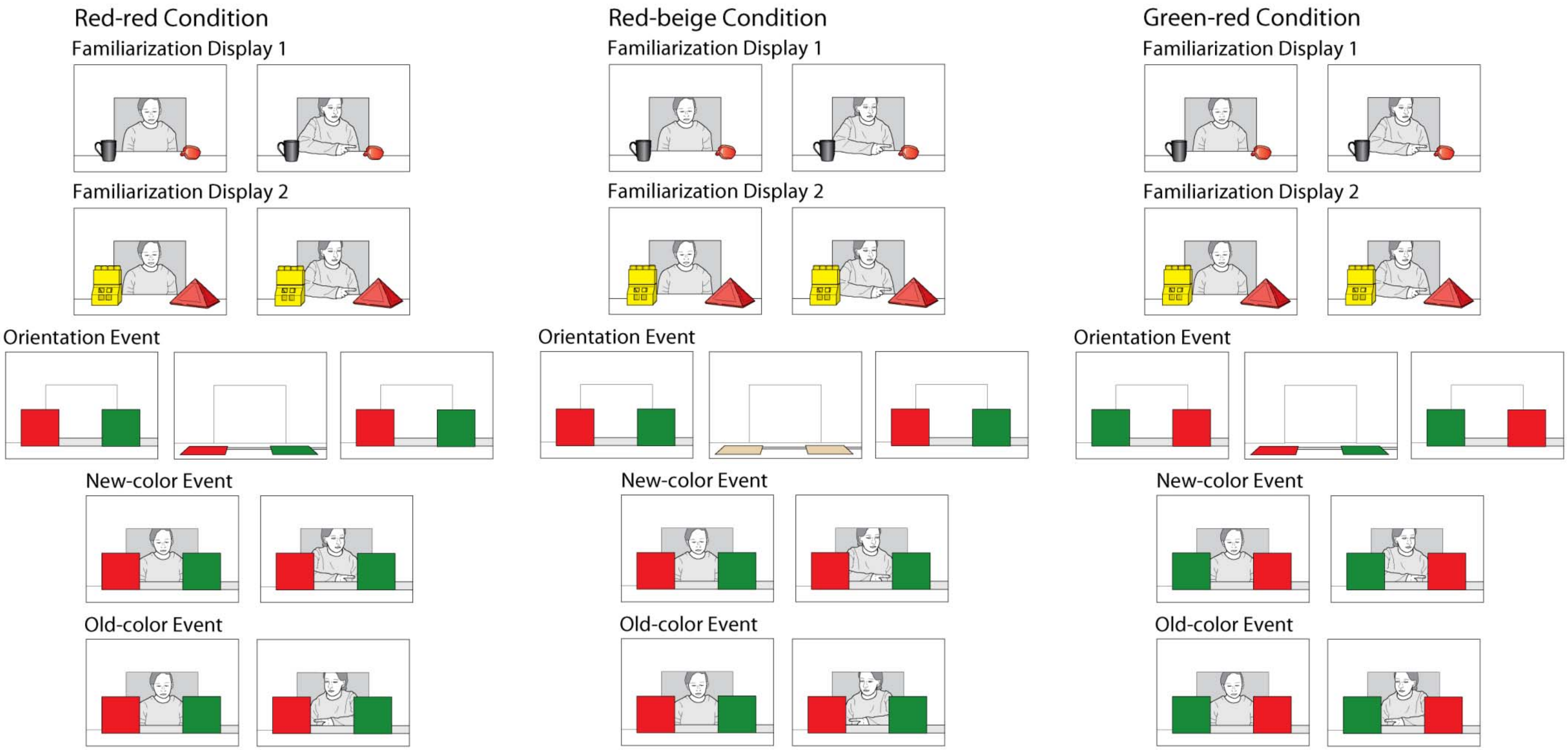
Figure 2. Schematic drawing of the events shown in the hidden-object, short-screen, and tall-object conditions of Luo and Johnson (2009).

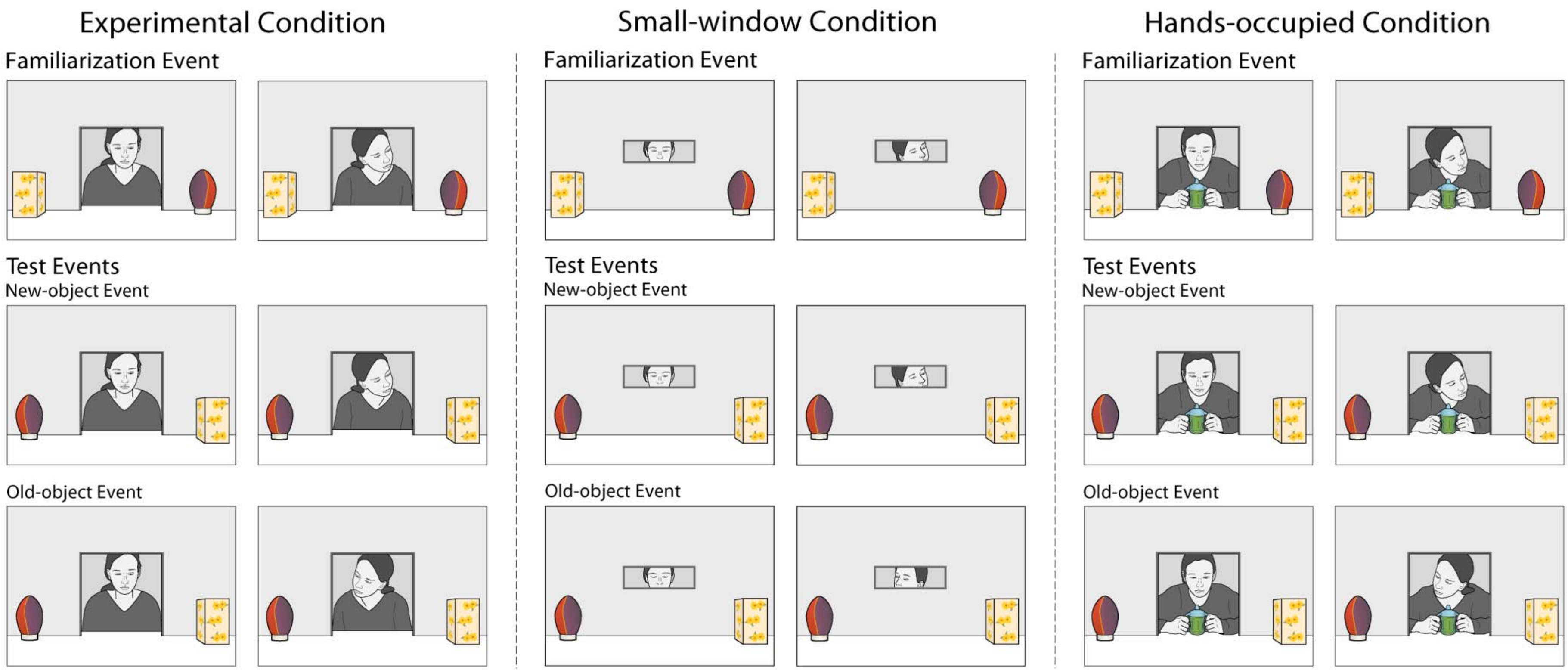
Figure 3. Schematic drawing of the events shown in the red-red, red-beige, and green-red conditions of Luo and Beck (2010).

Figure 4. Schematic drawing of the events shown in the experimental, small-window, and hands-occupied conditions of Luo (in press).


Experimental Condition
Single-object Condition
Orientation Event
Familiarization Event
Display
New-object Event
Old-object Event
Orientation Event
Familiarization Event
Display
New-object Event
Old-object Event


Hidden-object Condition
Short-screen Condition
Tall-object Condition
Familiarization Event
Familiarization Event
Familiarization Event
New-object Event
New-object Event
New-object Event
Old-object Event
Old-object Event
Old-object Event


Red-red Condition
Familiarization Display 1
Familiarization Display 2
Orientation Event
New-color Event
Old-color Event
Red-beige Condition
Familiarization Display 1
Familiarization Display 2
Orientation Event
New-color Event
Old-color Event
Green-red Condition
Familiarization Display 1
Familiarization Display 2
Orientation Event
New-color Event
Old-color Event


Experimental Condition
Familiarization Event
Test Events
New-object Event
Old-object Event
Small-window Condition
Familiarization Event
Test Events
New-object Event
Old-object Event
Hands-occupied Condition
Familiarization Event
Test Events
New-object Event
Old-object Event