AMERICAN PSYCHOLOGICAL ASSOCIATION

Journal of Experimental Psychology: General

ISSN: 0096-3445

2025, Vol. 154, No. 2, 497–521
https://doi.org/10.1037/xge0001656

An Associative-Learning Account of How Infants Learn About Causal Action in Animates and Inanimates: A Critical Reexamination of Four Classic Studies

Deon T. Benton
Department of Psychology and Human Development, Peabody College, Vanderbilt University

Considerable research shows that causal perception emerges between 6 and 10 months of age. Yet, because this research tends to use artificial stimuli, it is unanswered how or through what mechanisms of change human infants learn about the causal properties of real-world categories such as animate entities and inanimate objects. One answer to this question is that this knowledge is innate (i.e., unlearned, evolutionarily ancient, and possibly present at birth) and underpinned by core knowledge and core cognition. An alternative perspective that is tested here through computer simulations is that infants acquire this knowledge via domain-general associative learning. This article demonstrates that associative learning alone—as instantiated in an artificial neural network—is sufficient to explain the data presented in four classic infancy studies: Spelke et al. (1995), Saxe et al. (2005), Saxe et al. (2007), and Markson and Spelke (2006). This work not only advances theoretical perspectives within developmental psychology but also has implications for the design of artificial intelligence systems inspired by human cognitive development.

Public Significance Statement
A novel theoretical perspective is offered in the present article that can inform existing theories about how infants learn about the causal properties of objects and entities in the real world.

Keywords: causal learning, mechanisms of change, computational modeling

Supplemental materials: https://doi.org/10.1037/xge0001656.supp

Causal perception—which can be defined broadly as the capacity to "see" or apprehend real-world causal relations between objects and entities—is a fundamental cognitive ability. This capacity not only is what enables people to understand how the world works around them and may underlie later causal reasoning but is an ability that may underlie many early competencies such infants' ostensible knowledge that unsupported objects fall (e.g., Needham & Baillargeon, 1993) or that objects continue to exist when hidden (Baillargeon, 1987; Needham & Baillargeon, 1993). There is now considerable evidence that the ability to perceive causal relations emerges between 6 and 10 months of age (e.g., Bélanger & Desrochers, 2001; Oakes, 1994; Oakes & Cohen, 1990; cf. Mascalzoni et al., 2013; Rakison & Krogh, 2012). For example, in one of the first studies on this topic, Leslie and Keeble (1987) habituated 7½-month-olds either to a direct-launching sequence or to a delayed-launching sequence. In the direct-launching sequence, a first object ostensibly caused a second object to move immediately through direct, physical contact. The delayed-launching sequence was like the direct-launching sequence except that the second object began to move only after a short delay following contact from the first object. Infants were then tested with the reversal of their respective habituation sequences. Leslie and Keeble (1987) found that infants habituated to the direct-launching sequence showed greater dishabituation to the reversal of that sequence than infants habituated to and tested on the reversal of the delayed-launching sequence. Leslie and Keeble (1987) interpreted these findings to mean that the ability to perceive cause-and-effect relations in simple launching sequences emerges by 7½ months of age.

Despite considerable research on infant causal perception of simple launching sequences, relatively little is known about infants' developing knowledge about the causal properties of people and inanimate objects. For instance, it remains unresolved

This article was published Online First September 23, 2024.
Joseph Toscano served as action editor.
The modeling scripts and code have been made publicly available through GitHub. The modeling code is available at https://github.com/dtbenton/benton2024_JEP-G. The data presented here were not presented elsewhere.
The authors have no conflicts of interest to disclose. This work was supported by a grant from the National Institutes of Health (Grant R03HD112586-01) and the Eunice Kennedy Shriver National Institute of Child Health and Human Development (Grant 1R03HD112586-01) awarded to Deon T. Benton.
Deon T. Benton played a lead role in conceptualization, data curation, formal analysis, funding acquisition, investigation, methodology, project administration, resources, visualization, writing–original draft, and writing–review and editing.
Correspondence concerning this article should be addressed to Deon T. Benton, Department of Psychology and Human Development, Peabody College, Vanderbilt University, 230 Appleton Place #552, Nashville, TN 37235, United States. Email: deon.benton@vanderbilt.edu

when and importantly how—or through what cognitive mechanism of change—infants learn that people and objects possess distinct causal properties. One way that people differ from inanimate objects is that they can cause other people to act at a distance, in the absence of physical contact. For example, if person A notices that person B is in the path of oncoming traffic, person A can tell person B to "watch out" to avoid getting hit by the vehicles; person A need not physically move person B for person B to act, although person A could have caused person B to act on contact. In contrast, inanimate objects cannot cause action at a distance in other objects. Instead, they require physical contact from other objects and entities to act and move. This appreciation for the difference between the causal properties of people and inanimate objects is important because it not only underlies our knowledge of how the world works but it may support our knowledge about how best to interact with people and inanimate objects.

One prominent theory that attempted to address how infants learn about the causal properties of people and objects is the core knowledge perspective (e.g., Spelke, 2022; Spelke & Kinzler, 2007, 2009; see Carey, 2009, for a related core cognition account). The crux of this account is that infants are born with a small number of separable "core" systems. Two of these systems—the core system for agents and the core system for objects—enable infants to know how objects and people behave in the real world. For example, the core knowledge system for agents enables infants to know that agents are goal-directed, self-propelled, can move on nonlinear trajectories, and can cause action at a distance in other agents. In contrast, the core system for objects allows infants to know that objects do not move in the absence of physical contact from other objects or agents, are bounded and cohesive, and tend to move on linear trajectories.

Early support for these systems was ostensibly garnered by Spelke et al. (1995). This study investigated whether 7-month-old infants understand that people, but not inanimate objects, can cause action at a distance in other people. Infants were habituated to one of two live events that involved real, three-dimensional objects or people. In the People condition, a person entered the stage from the left and traveled a short distance until it disappeared behind an opaque screen that was located in the middle of the stage. A second, initially half-covered person then emerged from behind the screen and exited the stage to the right. The Inanimate Object condition was identical to the first condition except that it used two inanimate objects (i.e., a red box with a jagged top edge and a blue cylinder). Infants then saw two test events without the screen three times in alternation. In the Contact test event, the first person (or object) moved toward and contacted the other person (or object) at the center of the stage. In the No Contact test event, the first person or object moved toward but ultimately stopped short of the other person or object before the second object or person began to move. Spelke et al. (1995) found that infants in the Inanimate Object condition looked longer at the No Collision test event than at the Collision test event. In contrast, infants in the People condition looked equally long at both test events. These and other findings were interpreted to mean that infants possess core knowledge systems for people and objects. On this account, infants "know" (hence, core *knowledge*) that people, but not inanimate objects, can cause other people to move at a distance.

Although this theoretical account can explain infants' looking patterns in Spelke et al. (1995), it is limited in two notable ways. First, although the core knowledge perspective assumes that infants possess innate—that is, evolutionarily ancient, unlearned, and possibly present from birth (e.g., Carey, 2009; Spelke, 2022)—causal knowledge about people and objects, Spelke et al. (1995) tested 7-month-olds. This is problematic because infants' pattern of looking could have been based on extensive, real-world experience with people and objects rather than on innate causal knowledge about people and objects. Second, it is not necessary to assume that infants are born with innate knowledge about people and objects to explain the looking behavior of infants in Spelke et al. (1995). Such knowledge may instead have derived from an associative-learning mechanism that links salient perceptual, surface features (e.g., eyes or legs) with perceptually distinct kinds of causal events such as events in which things with certain perceptual or surface features (e.g., eyes, legs, arms, heads, all features that people possess) move in the absence or presence of physical contact and events in which things with certain other features (e.g., things without legs, something that defines most objects) move only in the presence of contact.

Here, I extend this second proposal. In particular, I present—and then test through a series of computer simulations—an alternative proposal for the cognitive mechanism through which infants learn about the causal properties of people and objects. This mechanism begins when infants first notice that a link exists between certain perceptual features that are available in the perceptual array and the different kinds of low-level, perceptually based causal events mentioned above. To avoid any confusions about how I am assuming infants represent these different kinds of events, a critical note is worth making here. Unlike the core knowledge account—which assumes that infants go beyond the low-level, perceptually based descriptions of the causal events to represent them abstractly as two *conceptually* distinct kinds of causal events (e.g., the concept *action-at-a-distance causality* vs. the concept *contact causality*)—the present account makes no such assumption. Instead, my view is that in the same way that infants with normal vision can perceive, see, or detect objects and entities in the world without necessary recourse to conceptually rich interpretations of those things (e.g., infants might notice that there is a cylindrical-shaped object in the corner of the room without knowing that the object is from the conceptual category *sphere*), they see people moving at a distance and people or objects moving following physical contact in terms of low-level, perceptually based, kinematic descriptions of those events. Stated plainly, the present account does not assume that infants use abstract, conceptually rich knowledge to interpret low-level causal events—infants simply perceive the events as they are without recourse to conceptually rich, abstract inferences.

Infants may come to notice and then encode links between different surface features and different low-level, kinematic descriptions of causal action upon noticing that some perceptual features tend to co-occur with different kinds of causal action across time and space. For example, infants may learn that perceptual features whose configuration is in the shape of a canonical human leg (although it need not be legs, and the present series of simulations go to great pains to be agnostic about the exact causally relevant feature or features that participate in these links) "go with" causal action at a distance as well as motion on contact based on seeing events in which people cause other people to move both at a distance and on contact. Infants may be attuned to these relations in the first place—that is, they may notice that some feature or small number of features "go with" two kinds of causal action—based on an innate or early emerging orienting bias to attend to movement over nonmovement (e.g., Slater, 1989), among other potentially relevant attention biases. This act of noticing that action following contact and motion at a distance tends to include some causally relevant, low-level perceptual feature of animate entities may serve to establish a nascent link between that feature and the two kinds of perceptually defined causal action, which becomes ever strengthened as the components of the relation are repeatedly experienced together. Although again I remain agnostic about the precise feature or features that infants come to associate with different kinds of causal action (the feature may be legs, but it may be some other feature), there is good reason to think that the feature may well be legs (though this issue should be examined more closely in future research). There are two reasons for this. First, at birth infants' visual acuity is 20/600 and does not reach adult levels until they are approximately 6 months of age (e.g., Ayzenberg & Behrmann, 2024). This means that this level of visual acuity may be sufficient for detecting and encoding large visual features whose overall shape is relatively simple such as the configuration of a person's arms or legs—especially when those features draw infants' attention by moving (e.g., Rakison & Poulin-Dubois, 2001, 2002)—but insufficient for encoding small, detailed visual features such as a person's eyes or their individual fingers or toes. Second and relatedly, when a person walks from one place to the next, not only does their entire body move (i.e., there is *global* movement of the entire entity), but there is also concurrent and synchronized *local* movement of their legs (e.g., flexing, extending, swinging, etc.). Such "double motion"—that is, movement of the entire body with movement of the legs—perhaps combined with poor visual acuity may heighten the salience of legs (or arms) relative to other finer and more detailed body parts. In turn, this may allow infants to form links between legs and different kinds of causal action rather than between some other body part and those same actions.

Whatever the feature is that participates in the relevant causal relation, the consequence of this mechanism is that the presence of one of the features alone will come to trigger an expectation for the other feature (e.g., seeing causal action at a distance), even if the other feature is not physically present. This means that if one (e.g., entities with legs) but not both (e.g., entities with legs *and* causal action at a distance) of the features is physically present, this may cause infants to show heightened or increased looking, as if to expect the second feature (i.e., things with leglike structures) given the first feature (i.e., causal action at a distance). This account may well explain why and crucially *how* infants in Spelke et al. (1995) came to look longer at the No Collision test event relative to the Collision test event in the Inanimate Object condition but not in the People condition; the No Collision event (i.e., causal action at a distance) in the Inanimate Object condition triggered an expectation for objects with certain low-level features that ultimately went unmet.

A Computational Instantiation of the Present Associative-Learning Account: Four Case Studies

The goals of the present simulation studies were twofold. The first goal was to implement the present associative-learning account in an artificial neural network to determine whether it was sufficient to capture infants' looking behavior in Spelke et al. (1995). The second goal was to examine the explanatory breadth of the present associative-learning account. Specifically, I examined whether this account could explain infants' looking behaviors in three other classic studies in the infancy literature. The second study that I focused on in Simulation 2 was Saxe et al. (2005). These authors examined whether infants understand that agents, but not inanimate objects, can cause ballistic motion in other objects. Ten- and 12-month-old infants were habituated to an event in which a beanbag was thrown, from either the left or right side of the stage, over a short wall—the trajectory that the beanbag took as it flew over the barrier is what is meant by ballistic motion. At test a live hand (or a toy truck) appeared from the same side from which the beanbag emerged (Same Side test trial) or from the opposite side from which the beanbag emerged (Different Side test trial), and the experimenter measured the amount of time that infants looked at both events. The results indicated that 10- (Experiment 3) and 12-month-old (Experiment 1) infants looked longer when a hand emerged from the opposite side from which the beanbag emerged than when the hand emerged from the same side from which the beanbag emerged. In contrast, 12-month-old infants looked equally long at the Same Side and Different Side test trials when a toy truck replaced the human hand.

The third study I focused on was Saxe et al. (2007). This study examined whether 7- to 10-month-old infants understood that animate entities, but not inanimate objects, could cause ballistic motion in other things. Infants were habituated to events in which one beanbag was thrown from behind a screen located on the right side of the stage and another beanbag was thrown from behind a screen located on the left side of the stage. At test the two screens were lowered to reveal what was behind them. Behind one screen (e.g., the right screen) was a human hand (Experiment 1) or a puppet (Experiment 2), and behind the other screen (e.g., the left screen) was a toy truck. The screens then rotated back up to occlude the human hand (or puppet) and the toy truck, and a beanbag was once again thrown from behind the right screen on half of the test trials and from behind the left screen on the remaining half of the trials. Saxe et al. (2007) found that infants looked longer when the beanbag emerged from the side on which the train was located (i.e., the unexpected test trials) than when the beanbag emerged from the side on which the human hand or novel puppet was located.

The fourth and final study I focused on was Markson and Spelke (2006). This study examined whether 7-month-olds could quickly learn about the self-propelled motion of objects. Infants were habituated to events in which a wind-up toy (i.e., a first toy animal) moved on its own across a stage as well as events in which a second wind-up toy (i.e., a second toy animal) was made to move across the stage by a human hand. Infants were then shown a test event in which the self-propelled object and the hand-generated object, now motionless, were positioned next to each other on the stage. Markson and Spelke (2006) reasoned that if infants could quickly learn that a novel toy is self-propelled, then infants should look longer at the previously self-propelled object than at the previously hand-generated object, as if to expect it to continue moving. Across six experiments, Markson and Spelke (2006) not only found that infants looked longer at the previously self-propelled object than at the previously hand-generated toy, but they found that this looking pattern persisted over a brief delay (Experiment 2). In addition, infants were at chance in their looking to both sides of the stage when the toy animals were either replaced with amorphous blobs (Experiment 3) or toy trucks (Experiment 5).

Finally, Markson and Spelke (2006) showed that these latter findings were not due to infants' inability to distinguish between the amorphous blobs (Experiment 4) or the toy trucks (Experiment 6).

The rationale for focusing on these studies was that data from them have been interpreted as providing support for core knowledge systems (e.g., Spelke, 2022; Spelke & Kinzler, 2007, 2009) and core cognition (e.g., Carey, 2009). Here I will show that the present associative-learning account as implemented in an artificial neural network can capture the data in these studies. It is worth noting, however, that although the present simulations do demonstrate that associative learning is sufficient to explain how infants learn about the causal action of people and inanimate objects and that core knowledge or cognition is not necessary (for a similar associative-learning account of language learning, see McMurray et al., 2012), they do not (and indeed cannot) show that infants do not possess sophisticated, innate concepts or specialized learning processes. The answer to the question about what mechanism infants and children actually use to learn about the causal properties of people and objects ultimately can only be determined with behavioral experiments, which, ideally, would be designed to test the predictions of competing theoretical accounts.

Before proceeding, a brief note is worth making. Although the current simulations focus on a subset of studies by prominent developmental scientists, the simulations were designed to address a much broader point: Domain-general associative learning may be sufficient to explain the emergence of many phenomena, including the emergence of causal learning in infants and young children (the topic of this article), without recourse to innate, content-rich knowledge or specialized learning mechanisms, skeletal systems, or Fodorian-like modules.

Simulation 1a: Spelke et al. (1995)

Simulation 1a had two aims. First, it examined whether associative learning alone could account for 7-month-old infants' looking behavior in the People and Object conditions in Spelke et al. (1995). Second, it was designed to determine what testable predictions the network makes when "younger networks" are tested.

Method

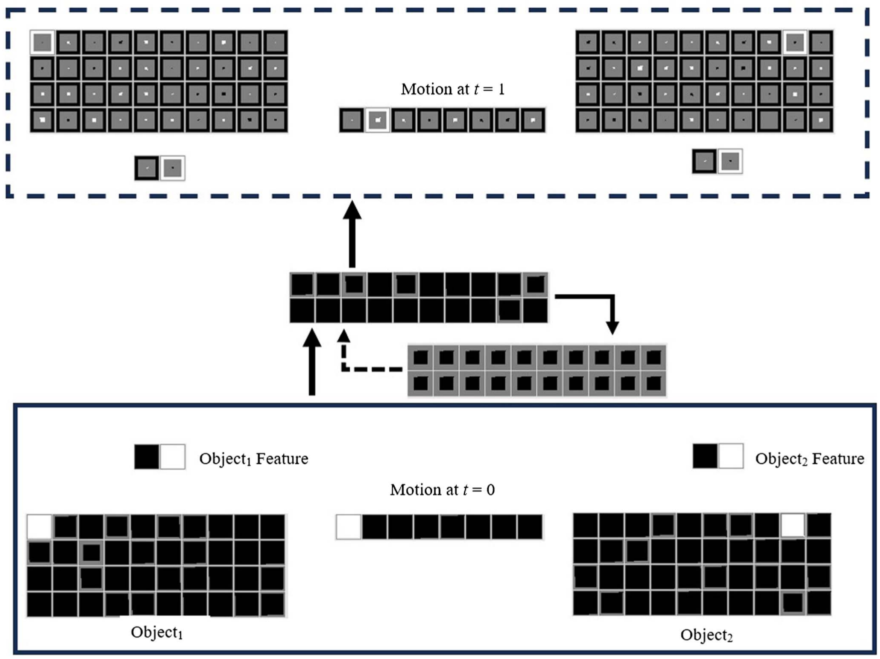
Network Architecture

I used a three-layer, feedforward, autoassociative simple-recurrent network across all simulations (e.g., Elman, 1990; Mareschal et al., 2000). The model was trained using backpropagation in cross-entropy error (e.g., Rumelhart et al., 1986). The activations of the output units were set according to a sigmoid activation function, whereas the activations of the hidden units were set according to a rectified linear unit activation function to prevent the gradients from vanishing as error was backpropagated across layers. The learning rate, momentum, weight decay, and number of hidden units were set to 0.06, 0.9, 0.0001, and 20, respectively, for the older networks, whereas they were set to 0.02, 0.9, 0.001, and 10, respectively, for the younger networks. The values to which these parameters were set is consistent with past developmental connectionist modeling research (e.g., Benton et al., 2021; Mareschal & French, 2000; Rakison & Benton, 2019; Westermann & Mareschal, 2004). Finally, Gaussian noise was added to the output activations of the hidden and output units to reflect the fact that infants' learning and processing tends to be a noisy function of the input they receive from the environment. The value of this parameter was set to 0.1 ($M = 0$, $SD = 0.1$) for the older networks and 1 ($M = 0$, $SD = 1$) for the younger networks to reflect the fact that learning and processing in older infants is presumably less noisy than in younger infants. Together, these parameters implemented a very simple model of age and development that is consistent with previous connectionist simulation studies (e.g., Benton et al., 2021; Benton & Lapan, 2022; Rakison & Lupyan, 2008). Weights were initialized to small random values (sampled uniformly between ±0.1) for all networks regardless of developmental age.

The model consisted of three layers (Figure 1). The input and output layer consisted of five "banks" of units, and the hidden layer consisted of a single group of "hidden" units. Two of the input banks of units—each of which consisted of 40 units—were used to represent the animate entities and inanimate objects, with one animate entity or inanimate object being presented on the left side of the network and the other animate entity or inanimate object being presented on the right side of the network. People and objects were represented in the model as orthogonal patterns of activity; that is, a single unit was used to represent a particular person or object. This ensured that the network's responses at test were based on learned associations between particular features and different types of causal action rather than on the particular features of a given person or object. Given that the similarity between any two people, two objects, or an object and a person was minimized in this (as well as in the following) simulations (due to the use of orthogonal patterns of activation), the network could not rely on the particular representation of an object or person to encode the relevant relations. Crucially, people and objects were presented in both banks of units to simulate the fact that two people (or objects) were present at the same time in the study by Spelke et al. (1995) and the fact that causing action through contact or action at a distance requires that two entities or objects be present.

In addition to these banks of units, two other banks of units represented whether a given person or object possessed animate features (i.e., the Objects Feature group in Figure 1). Note again that I am agnostic about which particular feature infants associate with different kinds of action and am using the term "animate feature" merely as a stand-in for whatever animate or inanimate feature is linked to different kinds of causal action in the real world. Each bank consisted of two units. If the first unit in this bank was set to "on" (i.e., its value clamped to 1) and the second unit in this bank was set to "off" (i.e., its value clamped to 0), this indicated that the person or object to which this bank of units corresponded possessed inanimate object features. However, if the first unit in this bank was set to "off" and the second unit set to "on," this indicated that the person or object to which this bank of units corresponded possessed animate object features. These four banks of units instantiated the autoassociative component of the model. As is true for all autoassociative models, the network's task was to recreate the pattern of activity presented to the input groups along the corresponding output groups through an intermediate group of hidden units. Given that the number of hidden units in the hidden layer was necessarily smaller than that in either the input or output layers, this forced the model to develop a more compact representation of the input that was sufficiently reliable such that when expanded the network reproduced the pattern of activation that was presented as

Figure 1
Schematic of an Untrained Model Used to Simulate Older Networks



Note. The model architecture for younger networks was similar to that for the older models except that the number of units in the hidden layer was reduced to 10. The colored borders that encircle each output unit correspond to the target for that unit. Black borders correspond to targets of "0" (i.e., the unit should be "off"). White borders correspond to targets of "1" (i.e., the unit should be "on"). The size of the colored region within the borders corresponds to the actual activation value of that unit. Larger inner regions correspond to greater activations. Darker colored inner regions for the output units correspond to activation values ≤0.5, whereas lighter colored inner regions correspond to activation values >0.5. For the hidden units, darker colored inner regions correspond to activation values <1, whereas lighter colored inner regions correspond to activation values ≥1. This difference stems from the fact that the activities of the output units are set by a logistic function, whereas the activities of the hidden units are set by a rectified linear unit activation function. See the online article for the color version of this figure.

input to the model along the corresponding groups in the output layer (e.g., Mareschal et al., 2000).

The final bank of units represented the "motion path" (i.e., the movement and the location in network space) of a given object. Although two groups of units were used to represent two different animate entities or inanimate objects on the network's left and right sides, a single motion path was used. This meant that on each time step, a single bit in the motion layer was active according to the current position of the first or second object. The networks' task was to predict the location of a given object (represented by a single unit) at the next time step. Activity presented before the midpoint of this path (i.e., before the fifth bit of the motion path) represented the motion of the animate entity or inanimate object on the left side of the network; activity presented at and after this midpoint encoded the motion of the animate or inanimate on the right side.

Finally, the hidden layer was connected to a corresponding group of "context" units, and these units in turn were connected to the hidden layer. These context units encoded the pattern of activity that was presented along the hidden layer at the immediately preceding previous time step and, functionally, served as another input group. This "recurrent" connectivity between the hidden and context layers provided the model with a rudimentary form of memory such that it was not only able to remember information from the past, but it was able to use that information (via its connections with the hidden layer) to help the network better predict information at the current moment in time.

Training

Pretraining

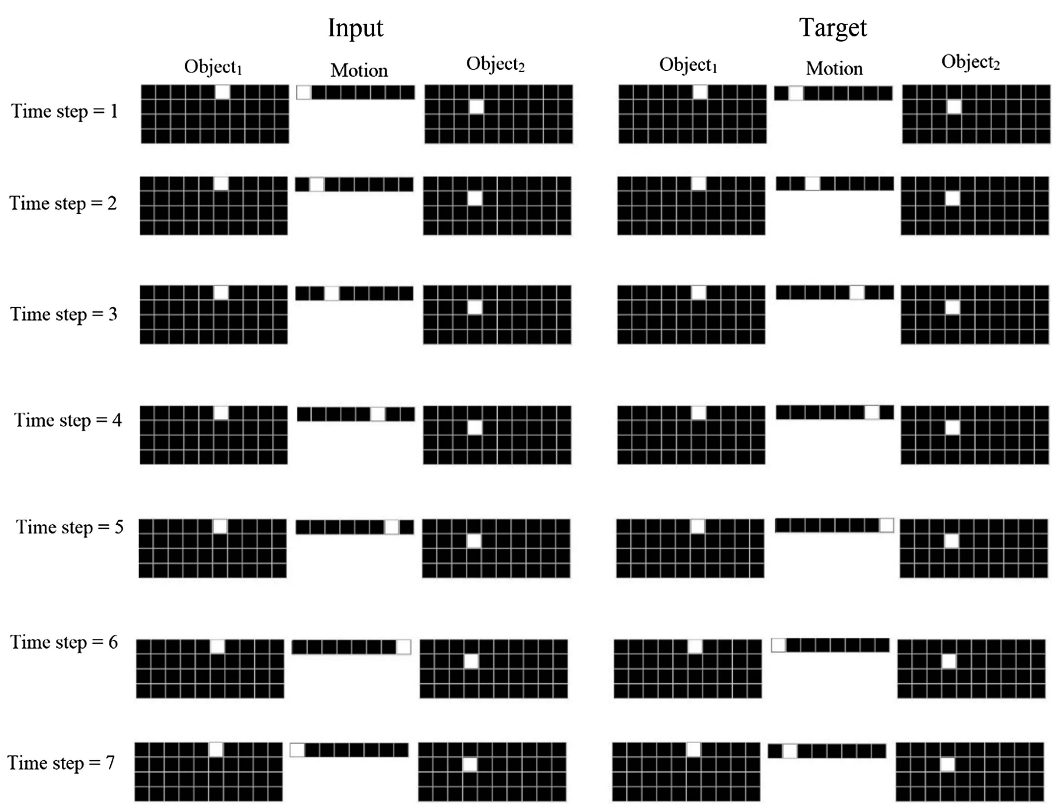
Older networks received 1,500 epochs of pretraining experience, whereas younger networks received 500 epochs of pretraining experience. The pretraining phase corresponded to the "real-world" experience with which infants presumably entered Spelke et al.'s (1995) study. The difference in the amount of pretraining experience between the older and younger networks reflected the fact that older infants necessarily enter experiments with more experience than younger infants. Crucially, there were three kinds of events that networks experienced during this phase. In one event ($N = 8$), objects with visible animate surface features caused other objects with visible animate surface features to move in a way that an adult would describe as objects causing other objects to move at a distance. Such action was instantiated in the following way (though see Figure 2 for a visual depiction of this event as it unfolded over time).

At time $t = 1$, the activation value of the first (or leftmost) unit in the motion group in the input layer was set to 1 (the activation values of the remaining units were set to 0), and the network's task was to activate the second unit (from the left) at time $t = 2$ in the corresponding motion group in the output layer. At time $t = 2$, the activation value of the second input unit in the motion group was set to 1 (all other values were set to 0), and the network's task was to turn on the third output unit (from the left) in the corresponding output group. At time $t = 3$, the third unit (from the left) was turned on, but this time the network had to predict that the second object would begin to move by turning on the sixth unit in the corresponding motion output group. That there was no intermediate movement between the third and sixth units simulated a situation in which a second object begins to move in the absence of contact from a first object. At time $t = 4$, the activation of the sixth input unit in the motion group was set to 1, and the network had to predict that the second object would continue moving by turning on the seventh unit in the motion group at the output layer. At time $t = 5$, the activation value of the seventh motion input unit was set to 1, and the network had to activate the eighth output unit in the motion output group. At this point in training, at time $t = 6$, the second object's motion "wrapped around" the motion vector such that the eighth motion input unit was activated, and the network had to reactivate the first motion input unit. Finally, at time $t = 7$, the first motion input unit was set to 1, and the network's job was to reactivate the second motion unit in the model's output layer.

In another event ($N = 8$), one object with animate features caused another object with animate features to move in a way that an adult would describe spatiotemporally and kinematically as movement following direct, physical contact. This event was similar to the first event except that the first animate entity contacted the second animate entity (i.e., the fourth and fifth units of the motion layer were sequentially activated in a manner similar to the activation of the other units). The final set of events ($N = 8$) were identical to the contact event that included objects with animate features except that the objects with animate features were replaced with objects without animate features or objects with inanimate features. Crucially, this pretraining phase instantiated the notion that objects with animate features can engage in multiple forms of spatiotemporally and kinematically defined causality, whereas objects with inanimate features tend to engage in a single form of spatiotemporally defined movement (i.e., contact causality).

Figure 2

An Example of the Input-Target Patterns Presented to the Model for the "Action-at-a-Distance" Events That Involved Objects With Animate Surface Features



Note. Each row corresponds to a particular point in time.

One potential criticism of the simulation as I have described it to this point is that the relations that networks were trained on during the pretraining phase were considerably "purer" than those that infants might experience in the real world and thus "easier" to learn. For example, in the model networks learned that things with "animate features" engaged in two different kinds of causal action, whereas things with "inanimate features" always engaged in a single action. Such experience is likely at odds with infants' real-world experiences with objects. For example, although generally it is the case that inanimate objects are caused to move by other inanimate objects (as well as by animate entities), a subset of infants' real-world experience with objects arguably also includes unaided object movement. Examples of this are unaided movement by vehicles, toys, and dolls, among other things. What this means is that compared to the current model in which animate and inanimate features are uniquely diagnostic of animacy and inanimacy, respectively, in the real-world animate entities and inanimate objects cannot be distinguished so easily. A model that is faithful to the real-world experiences of infants would need to determine what role, if any, different amounts of experience with atypical object action (e.g., objects moving unaided) have on the networks' ability to learn about the causal properties of animates and inanimates. To address this issue and to increase the ecological validity of the current series of simulations, in this (and in most of the following) simulation, the frequency with which things with "inanimate object features" cause other inanimate objects to move at a distance is varied. The frequencies ranged from 50% to 100% in increments of 10. To illustrate what a given frequency means, consider 60%. This frequency means that 60% of the time inanimate objects were caused to move; 40% of the time those same objects moved unaided. In contrast, a frequency of 70% means that 70% of the time inanimate objects were caused to move, whereas the remaining 30% of the time they moved unaided.

Habituation

Similar to Spelke et al. (1995), all networks—regardless of "age"—were randomly assigned either to the Objects ($N_{\text{older}} = 20$; $N_{\text{younger}} = 20$) condition or to the People condition ($N_{\text{older}} = 20$; $N_{\text{younger}} = 20$). The habituation events were identical to the pretraining events except that motion was absent during them and a new set of unrelated objects with animate features or objects without animate features (depending on the condition to which a given network was assigned) were used. I chose not to model motion during this phase to simulate the fact that infants could not determine whether a contact or no contact event was being shown during habituation in Spelke et al. (1995). This was because in the original study, a central screen obscured the movement of the objects. Given that motion was absent during this phase, networks simply had to reproduce the pattern of activity in each of the input groups along the corresponding output groups. The habituation phase lasted four epochs; this value was used across all simulations.

Testing

Following habituation, networks were tested on two Contact and two No Contact test events, and networks' average response to the test events was assessed. I used as a measure of "looking time" network cross-entropy error (e.g., Sirois & Mareschal, 2002). Error was computed over all the output groups. Larger errors indicate a larger discrepancy between what the network observes (the pattern of activity across the output layer) and what it expects (the target information across the output layer). The contact events were identical to the pretraining contact events, and the no contact events were identical to the pretraining action-at-a-distance events.

Results

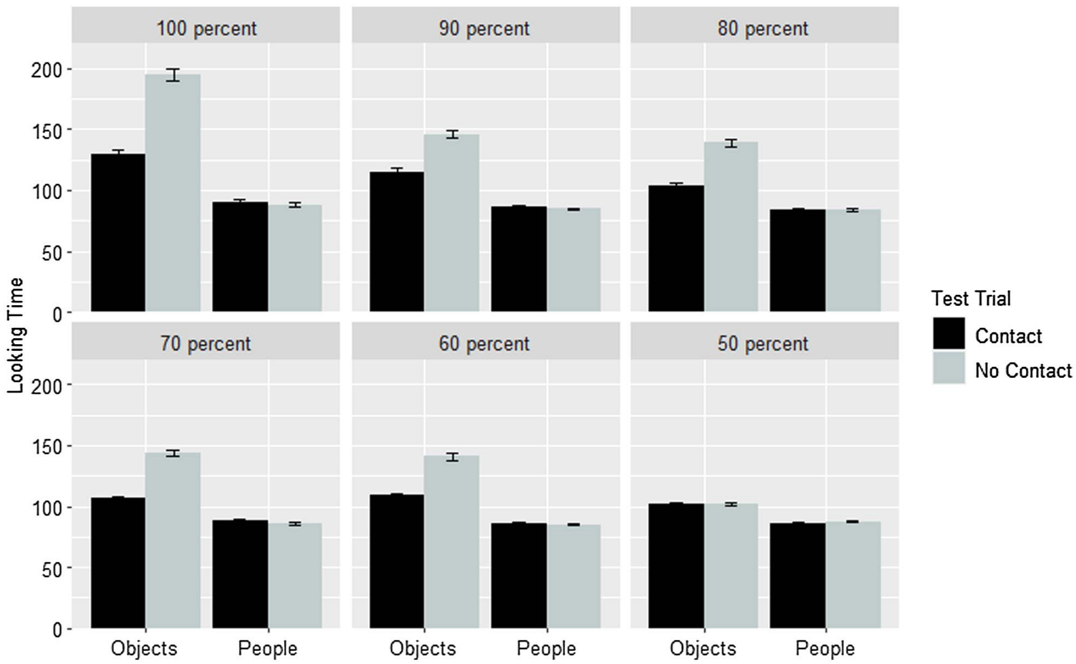
Figure 3 shows networks' mean "looking times" to the Contact and No Contact test events for older networks assigned to the Objects and People condition. Figure 4 shows the corresponding data for the younger models. Before discussing the key findings, it is worth mentioning that formal inferential statistics were not used in any of the simulations reported in this article (although they were included in the Supplemental Materials for completeness). The main reason for this is that the simulations reported here used a real-valued probability measure to operationalize looking time. Such a measure inherently limits the range of possible "looking times" to values that fall between 0 and 1 (due to the use of cross-entropy error in the simulations). The consequence of this constrained measurement range is that network looking time often exhibits reduced variability compared to what is observed in studies with human infants. Given that the same restrictions presumably do not apply to infants, infant looking time should exhibit greater variability than that of models, thereby enabling smaller differences to be reliable here but not in the behavioral studies on which the current simulations were based. The presence of a reliable difference in the current context, in contrast to the absence of one in the corresponding behavioral studies, should not be interpreted to mean that the models failed to capture key behavioral findings. Instead, such a difference should be considered an artifact of the error measure used. The following "analyses" eschewed this issue altogether by focusing instead on whether the qualitative pattern of results obtained by the model matched that observed in infants (for a similar analysis approach, see Mareschal et al., 2000; Mareschal & Shultz, 1999; McClelland & Thompson, 2007; Plaut & Vande Velde, 2017).

As can be seen in Figure 3, older networks assigned to the Object condition looked longer at No Contact test event than at the Contact test event. In contrast, networks assigned to the People condition looked about equally at the Contact and No Contact test events. Both results held for all frequencies except for the 50% frequency, in which 50% of the time inanimate objects moved following contact and the remaining 50% of the time they moved unaided. Crucially, these results replicated those found by Spelke et al. (1995). An interesting facet of these data is that the difference in looking time to the No Collision and Collision test events for networks assigned either to the Object condition or to the People condition decreases as the network's experience with atypical object action (i.e., things with inanimate features causing other things with inanimate features to move at a distance) increases. The basis for this is straightforward: As networks' experience with atypical object action increases, their *relative* experience with typical people action (i.e., things with "animate features" causing both action at a distance and on contact) decreases. It turns out that this general pattern (as well as the explanation for it) applies to remaining simulations as well.

A different pattern of results emerged for the younger models, as can be seen in Figure 4: These models looked about equally at both test events across conditions and frequencies, although there were

Figure 3

"Older" Networks' Mean "Looking Time" (i.e., Cross-Entropy Error) to the Contact and No Contact Test Events Across Conditions



Note. See the online article for the color version of this figure.

places where the younger models showed a slight preference for the No Collision test events compared to the Collision test events. Crucially, this latter result was not specific to the Object condition; it applied consistently to at least three of the four conditions.

Discussion

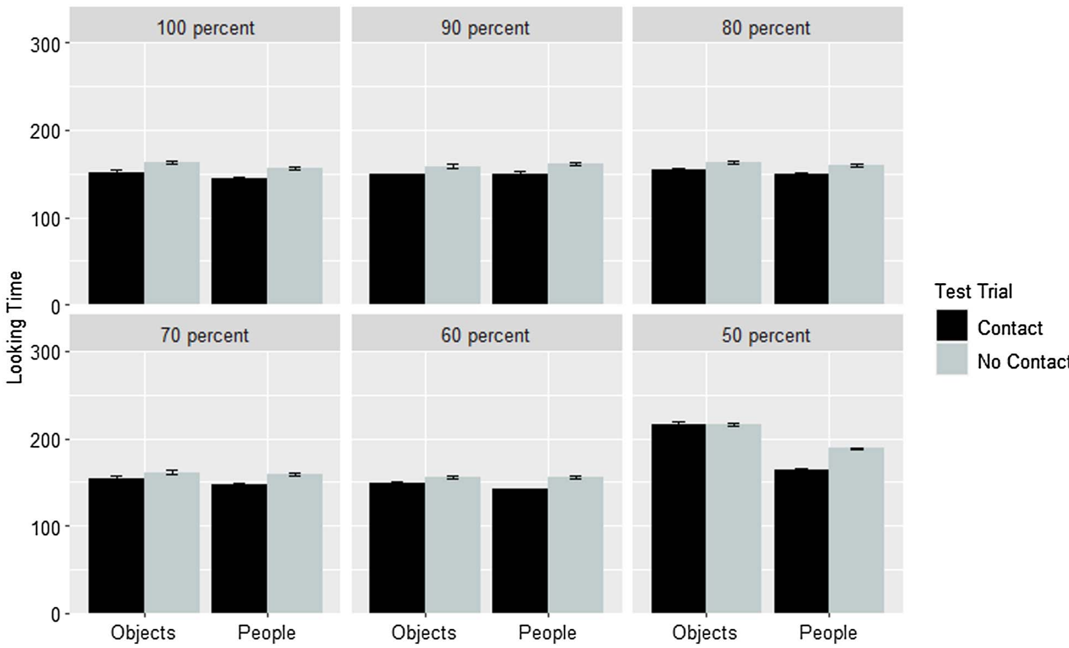
Simulation 1a replicated infants' looking responses in Spelke et al. (1995). Networks looked longer when an object that possessed inanimate features moved in the absence of contact than when it moved following contact, whereas they looked about equally long when an entity that possessed animate features moved in the presence and absence of physical contact from another entity with animate features. Crucially, this result largely did not depend on how frequently inanimate objects engaged in atypical action (e.g., causing other inanimate objects to move at a distance). This result is significant because it suggests that the conclusion that infants' causal knowledge about people and inanimate objects is unlearned may be premature. The current modeling results demonstrate that domain-general associative learning is sufficient to explain how infants learn about the causal properties of people and objects. It is worth noting that there was nothing intrinsic to the training examples used here that distinguished objects with animate features from the objects with inanimate features. As is true of simulations like this one (e.g., Rogers & McClelland, 2004), an object's "meaning" is determined by the associative relations into which it enters rather than by something abstract about the object. In terms of the present simulation, the networks' responses at test were due to learned associations between particular visible, low-level, surface features (available to any low-level parsing routine) and particular low-level, perceptual-based, kinematic depictions of different kinds of causal action.

The present simulation also extended Spelke et al. (1995) by making a series of testable predictions. The first is that infants younger than those tested in Spelke et al. (1995)—although I make no claims about *how* much younger—should be at chance in their looking to the Contact and No Contact test events regardless of the condition to which they are assigned. A second prediction that the network makes, which should also be tested in future research, is that there should be a developmental progression in infants' knowledge about the causal actions of people and objects: Older but not younger infants should show the same pattern of looking as infants in Spelke et al. (1995). Crucially, this is not a prediction that the core knowledge perspective would make. This is because infants' causal knowledge about people and objects is not learned and, as such, should not undergo a developmental progression. Together, this simulation shows that domain-general associative learning can explain infants' developing knowledge about people and object causal action.

Before presenting the next series of simulations that assess the explanatory breadth of the present associative-learning account, it was important to address two potential objections to the current simulations. A first potential objection is that the learning task faced by the current model differs in important ways from that faced by infants in the real world. For example, in the real world, infants must discover which of several co-occurring features is the one that is causally linked to different kinds of actions before they can link that feature (or a small number of features), perhaps via associative learning, with those different kinds of actions. In contrast, in the

Figure 4

"Younger" Networks' Mean "Looking Time" (i.e., Cross-Entropy Error) to the Contact and No Contact Test Events for Networks Assigned to the Objects and People Conditions



Note. See the online article for the color version of this figure.

current simulation, the model was not required to discover these features before associating them with different actions. Instead, these features were merely given to the model. This is presumably an easier task for the model than that encountered in the real world by infants, which could impact the validity of the current model. A complete account of how infants learn about the causal properties of people and object must therefore explain how infants discover those properties and features in the first place. One may argue that the former of these tasks—namely, discovering the relevant causal properties—requires a distinct, domain-specific learning mechanism. However, as Simulation 1b will demonstrate, these features can be discovered within an associative-learning mechanism if one assumes that the to-be-extracted features are the perceptually more salient ones; a separate mechanism is not required. As was discussed in the Introduction, this view is supported by substantial research demonstrating that infants possess at birth perceptual or attention biases that direct their attention to some objects in the perceptual array and away from other aspects in that same array (e.g., Rakison & Lupyan, 2008; Scott & Arcaro, 2023). It is generally assumed that what governs this orienting response is the salience of the various objects in the array; infants' attention will be "pulled" toward the more salient aspects of the array presumably because these features "pop out" to infants. This also means that if the salience of some aspect of the world is increased, then the salience of other, unattended aspects of the world will be decreased. The goal of Simulation 1b was to demonstrate that when some features are more salient than others, an associative-learning mechanism has little difficulty "discovering" the more salient of those features and linking them with different kinds of action. To simplify the simulations and increase their interpretability, this idea was instantiated with just two features; one of these features was more salient than the other. However, there is no reason to think that the same principle that enables the network to discover the causally relevant feature does not also play out in the real world to allow infants to discover the same features. Given that the larger set of simulations presented here were not designed to address this issue—that is, how infants discover the causally relevant features—but rather to instantiate the notion that infants learn about the causal properties of people and objects via associative learning (once the critical features have been extracted), this issue will only be explored in Simulation 1b.

A second potential objection is that because "people" were defined in the current simulations as entities with animate features (i.e., the animate "bit" was turned on in the Animate Features group) and "objects" were defined as things with inanimate features (i.e., the inanimate bit was turned on in the same group), it was neither possible to know how the model would treat anomalous or "hybrid" objects such as objects with animate features or people with inanimate features, nor was it possible to know how the network would respond to unmodified people or objects following "real-world" (i.e., pretraining) experience with hybrid or modified objects. Both of these are issues that stem from the fact that a single, two-unit group, which was used to represent items with animate features or items with inanimate features, was used in Simulation 1a. Yet, assessing how networks respond to these modified stimuli as well as how training on them impacts responding to unmodified stimuli can help further to tease apart core knowledge from associative-learning accounts. If infants' causal knowledge about people and objects is based on rich, abstract knowledge about people rather than on the perceptual features of objects as the core knowledge account assumes, then they should treat a person with inanimate features as a person and an object with animate features as an object. However, if infants' causal knowledge about people and objects is based on learned associations between low-level, perceptual-based features and different kinds of causal action as the present associative-learning account assumes, then their responses to hybrid objects presumably should be affected by the features that those objects possess.

Simulation 1c was designed to determine how a domain-general associative-learning mechanism responds when presented with unmodified and modified stimuli following training to both kinds of stimuli. These responses and behavior of the model can then be used as the basis for future behavioral research. This simulation was identical to Simulation 1b with two key differences. A first difference is that networks' pretraining experience not only included experience with unmodified people (i.e., entities with salient and less salient animate features) and unmodified objects (i.e., objects with salient and less salient inanimate features) but experience with modified objects (i.e., objects with salient inanimate features and less salient animate features). Crucially, both kinds of objects (i.e., unmodified and modified objects) only caused other objects of the same kind to act through contact; it was not the case that modified objects caused other modified objects to act at a distance. Events in which objects caused other objects to act at a distance were not included here for two reasons. First, these events were included in Simulation 1a. Second and most importantly, it was important to determine whether experience with modified objects affected networks' subsequent processing of unmodified and modified objects and people independent of experience with events in which objects cause action in other objects in the absence of physical contact.

The rationale for presenting the network with modified objects but not modified people during the pretraining phase was that infants presumably do not encounter people with object parts appended to them in the real world. However, their real-world encounters with objects probably do involve some amount of experience with modified objects (i.e., objects with animate features in the real world; e.g., a Mr. Potato Head doll). Given that it was unclear how much real-world experience infants have with modified objects, as in Simulation 1 the frequency with which networks experienced them was varied from 50% to 100% in increments of 10. Here, a frequency of, say, 80% means that 80% of the time the network experienced unmodified objects or objects with salient inanimate features; 20% of the time networks experience modified objects or objects that possessed salient inanimate features and less salient animate features. Because networks in this simulation could encounter modified objects during training and modified objects and people during test, it is worth being clear on what determines an item's true or "ontological" status. Here, an item was considered an object if it possessed a salient inanimate feature, whereas an item was considered a person if it possessed a salient animate feature. Thus, a modified object is an object with salient inanimate features and less salient animate features. In contrast, a modified person is an entity with salient animate features and less salient inanimate features. It is worth noting here that there is an implicit assumption that the animacy features of modified objects are less salient than those of unmodified people. As discussed in the Introduction, this difference may stem from the fact that the animacy features on objects tend to be smaller and finer compared to those on people, and thus may capture infants' attention less due to their initial poor visual acuity. Nonetheless, this issue should be explored further in future simulation research.

A second difference is that unlike Simulations 1a and 1b, networks in Simulation 1c were assigned to one of four conditions following pretraining experience: People, Objects, People with Inanimate Features, and Objects with Animate Features. By assigning networks to one of the four conditions, it was possible to determine to what extent network's "real-world" experience with unmodified people and objects as well as modified objects influenced their subsequent responses to these different kinds of things at test.

Simulation 1b: Discovering Causally Relevant Features

This simulation examined whether an associative-learning mechanism could discover which of two features is causally relevant when one of the features is more salient than the other.

Method, Training, and Testing

Network Architecture

Simulation 1b was identical to Simulation 1a except that two additional banks of "feature" units were used to implement the fact that people and objects possessed salient and less or nonsalient features. The first bank of units corresponded to salient features; the second bank of units corresponded to nonsalient features. Both banks of units used the same pattern of activation. Thus, if the first unit in this bank was set to "on" and the second unit set to "off," not only did this indicate the presence of inanimate object features, but this pattern of activation was used for the bank of units that coded for salient features and for the bank of units that coded for nonsalient features. In contrast, if the first unit in this bank was set to "off" and the second unit to "on," this indicated the presence of animate object features, and this pattern of activation was used for the salient and nonsalient banks of feature units. Crucially, to decrease the salience of the nonsalient features relative to the salient ones, thereby modeling attention biases, the learning rate and weight decay of weights from the nonsalient features to the hidden layer and the weights from the hidden layer to the corresponding output group of nonsalient features were set, respectively, to 1×10^{-6} and 0.375. Used in this way, the learning rate can be conceptualized as the speed at which the network learns that two (or more) features are correlated; the weight decay can be conceptualized as a kind of "pressure" that keeps those learned associations from growing too big, thereby outcompeting the salient animate features and the "pull" that salient features might exert on attention. Although these values were chosen somewhat arbitrarily, this approach to modeling attention biases as well as the interpretation of the learning rate and weight decay is well motivated by and consistent with past simulation studies (e.g., Lupyan & Rakison, 2006; Rakison & Lupyan, 2008; Rescorla & Wagner, 1972).

Training and Testing

All aspects of training and test were identical to Simulation 1a except that for half of the networks ($N = 20$), only the salient animate features were used, whereas for the remaining half of the networks ($N = 20$), only the nonsalient animate features were used. The rationale for this division was that it was possible to determine which of the two correlated features (e.g., salient animate features vs. nonsalient animate features) networks associated with different kinds of action. Networks were expected to associate only the salient features with their respective causal actions. Given that the salient and nonsalient features were equally correlated with different actions during the pretraining phase, demonstrating that the pattern of performance observed in Simulation 1a only emerges for networks tested with the salient features and not for networks tested with the nonsalient features would confirm that the network has successfully "discovered" the causally relevant (in this case, the salient) features.

One final point is worth making: The learning parameters used to simulate the older infants were adopted here. This is because the goal was to show how a learner with sufficient information processing abilities might discover causally relevant features. Adopting the learning parameters for the younger models could result in uninterpretable findings: If networks looked equally long at the Contact and No Contact test events in the People and Object conditions, it would not be possible to know whether this was due to a lack of sensitivity to salience or to insufficient learning.

Result

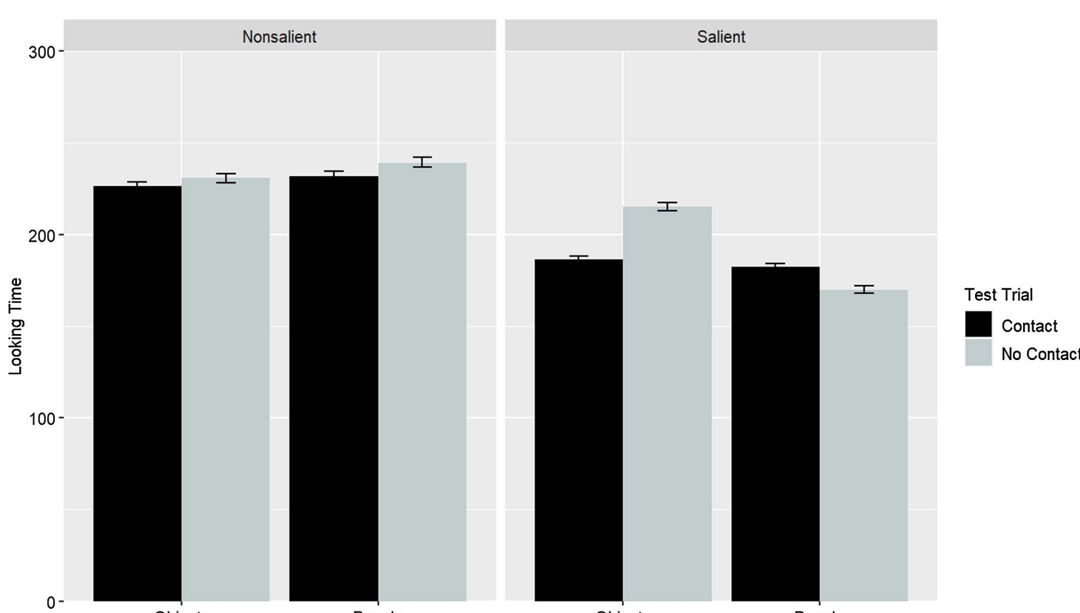
Figure 5 shows networks' mean "looking times" to the Contact and No Contact test events for networks shown nonsalient animate features at test and assigned to the Objects and People conditions and for networks shown salient animate features at test and assigned to the same two conditions. This figure reveals two notable takeaways. First, networks shown the nonsalient features showed longer looking overall at test compared to networks shown the salient features. The basis for this was the lower learning rate and greater weight decay for networks shown nonsalient features compared to networks shown salient features. These lower values prevented the associations between the nonsalient features and the different kinds of causal actions from growing as quickly as those between the salient features and same causal actions. Recall that it is these associations that enable the network to make correct predictions; the stronger the associations, the better one of the components of the association (e.g., a feature) can predict the other component (e.g., a specific causal action). A corollary of this is that network error (or conversely, longer looking) will necessarily be greater for networks shown the nonsalient features than for networks shown the salient features; the nonsalient features do a lesser job than the salient ones at predicting the correct causal actions. Second, only the salient features produced results that replicated those of the older infants in Simulation 1a; the nonsalient features produced equivalent looking across the test trials and conditions.

Discussion

Simulation 1b was designed to demonstrate that an associative-learning mechanism can discover which of two features is causally relevant when both features are equally correlated with different kinds of causal action. The results indicated that the discovery of the causally relevant feature is a natural and emergent consequence of some features being more salient than others. This result addresses the first objection and indicates that an associative-learning mechanism alone can "extract" causally relevant features and that a separate mechanism is not needed to extract such features. The next simulation was carried out to address the second potential objection: It was not possible in Simulation 1a to assess how

Figure 5

Networks' Mean "Looking Time" in the Nonsalient and Salient Conditions Across Contact and No Contact Test Trials for Objects and People



Note. Mean "looking time" (i.e., cross-entropy error) to the Contact and No Contact test events for networks first assigned either to the Nonsalient condition or Salient condition and then either to the People condition or to the Objects condition. Error bars represent standard errors. See the online article for the color version of this figure.

a network would treat unmodified and modified people and objects following experience with unmodified and modified people and objects.

Simulation 1c: How an Associative Learner Treats Hybrid Objects

This simulation assessed how an associative-learning model processed unmodified and modified people and objects following exposure to unmodified people and objects as well as to different “amounts” of exposure to modified objects.

Method, Training, and Testing

Network Architecture, Training, and Test

Simulation 1c was identical to Simulation 1b with two key exceptions. First, in addition to experience with unmodified people and objects during pretraining, networks also experienced modified objects. These are objects with a salient inanimate feature (i.e., the first unit in the feature bank was set to “on” and the second unit set to “off” for the salient group of features) and nonsalient animate feature (i.e., the second unit in the feature bank was set to “on” and the first unit set to “off” for the nonsalient group of features). Crucially, like the unmodified objects, modified objects only caused other modified objects to act through contact. Second, following the pretraining phase, networks were assigned to one of four conditions: Objects, People, Object with Animate Features, and People with Inanimate Features. The modified objects shown during habituation and test were constructed in the same manner as the objects used during pretraining. In particular, a modified person was an entity with a salient animate feature and a nonsalient inanimate feature; a modified object was an object with a salient inanimate feature and a nonsalient animate feature. All other aspects of habituation and test were identical to Simulations 1a and 1b.

Result

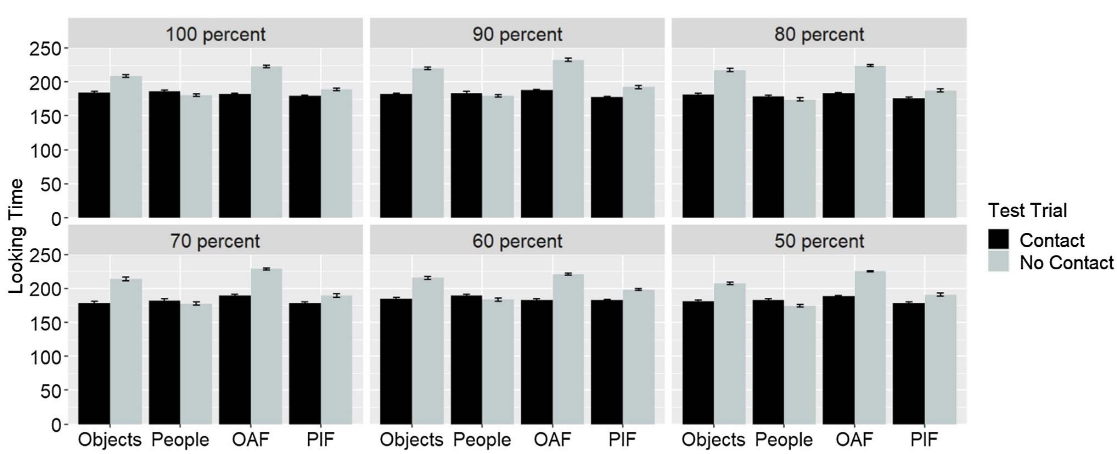
Figure 6 shows the mean “looking times” to the Contact and No Contact test events for networks assigned to one of the four conditions. Consistent with Simulation 1a, networks assigned to the People condition looked about equally at the Contact and No Contact test events across all frequencies, whereas networks assigned to the Objects condition looked longer at the No Contact test event than at the Contact test event. Interestingly, this same pattern was mirrored in networks assigned to the Object with Animate Features condition and those assigned to the People with Inanimate Features condition. Specifically, networks assigned to the Object with Animate Features condition looked longer at the No Contact test event than at the Contact test event. In contrast, the difference in looking to the two test events was close to chance for networks in the People with Inanimate Features condition.

Discussion

Simulation 1c assessed how an associative-learning mechanism treats hybrid objects such as people with inanimate features or objects with animate features following training with standard people and objects as well as hybrid objects. The results demonstrated that networks showed increased looking when unmodified objects caused other unmodified objects to act at distance compared to through contact. In contrast, networks showed equivalent looking to the same two events when they involved unmodified people. Similarly, networks looked longer when modified objects caused other modified objects to act at a distance compared to through contact, but their looking was close to chance when the events used modified people. These results are important for two reasons. First, they support the conclusion made in Simulation 1b that an associative-learning mechanism can not only associate some feature with different kinds of action but can discover the causally relevant feature. Second, they indicate that networks’ looking behavior is heavily influenced by the nature of the features that objects and

Figure 6

Networks’ Mean “Looking Time” (i.e., Cross-Entropy Error) to the Contact and No Contact Test Events for Networks Assigned to Each of the Four Conditions



Note. OAF = objects with animate features; PIF = people with inanimate features. See the online article for the color version of this figure.

entities possess; networks' looking behavior to an object or entity not only will be based on whether that object or entity possesses a salient feature but on the nature of that feature. If the salient feature is an animate feature, networks will treat as equivalent an event in which objects with the salient animate feature causes other objects with the same feature to act through contact and an event in which objects with the salient animate features causes other objects of the same kind to act at a distance. This makes a testable prediction that should be explored in future research: Infants should treat as anomalous events in which people with salient inanimate object features cause action in other people at a distance but not events in which objects with salient animate features cause action on contact in other objects. Crucially, this is not a prediction that the core knowledge account makes. The reason for this is that proponents of this perspective assume that infants possess conceptually rich and abstract knowledge about people and objects that does not depend on the particular perceptual features of the entities or objects.

The goal of the next series of simulations is to assess the explanatory breadth of the present associative-learning account; that is, it was important to show that this account could account for the behavioral findings in other prominent studies that assessed infants' causal knowledge about animate entities and inanimate objects. The model presented in Simulation 2 below was designed to determine whether the present account could explain the findings in Saxe et al. (2005). As was discussed in the Introduction, in this study 10- and 12-month-old infants looked longer when a hand (Experiment 1) emerged from the opposite side of the stage from which a beanbag emerged than when the hand emerged from the same side of the stage from which the beanbag emerged. However, if the event used a toy truck (Experiment 1) or a puppet (Experiment 2), infants looked equally long at both test events. Although these findings have been interpreted to support the existence of core knowledge and core cognition, Simulation 3 tested whether these results could be explained by an associative-learning mechanism. Simulation 2 was designed specifically to implement the idea that in the real world, objects with animate features, but not objects lacking such features, can cause ballistic and nonballistic motion in other objects. Moreover, whenever ballistic motion does occur in the real world, it tends to occur on the same side of physical space as objects with animate features.

Simulation 2: Saxe et al. (2005)

Simulation 2 examined whether associative learning was sufficient to explain infants' looking behaviors in Saxe et al. (2005).

Method

Network Architecture

The network architecture used in Simulation 2 differed from that used in Simulation 1a in two major ways. The first difference was that two motion vectors or groups were used as opposed to one motion group. The rationale for this decision was that ballistic motion occurred on the left and right sides of space in Saxe et al. (2005). Thus, one of the motion groups was located on the left side of "network space"; the other motion group was located on the right side of network space. The second difference concerned the nature of that motion. In Simulation 1a, networks were exposed to "linear" motion. In this simulation, networks were exposed to "ballistic" motion; that is, networks were shown events in which a one- (i.e., an inanimate object) or four-unit (i.e., an animate entity) object either did or did not undergo an inverted U-shaped motion pattern, which is akin to simulated ballistic motion (see Figures 6 and 7). All other aspects of the network architecture used in Simulation 2 were identical to Simulation 1a.

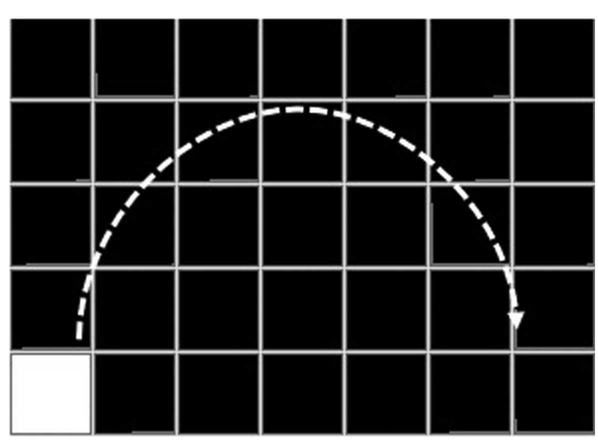
Training

Pretraining

As was the case in Simulation 1a, networks received 1,500 epochs of pretraining or "real-world" experience. In addition, the values of the training parameters (e.g., learning rate, weight decay, noise, and momentum) were identical to those used in Simulation 1a. During pretraining, networks learned the following key relations. First, they learned that object with animate features caused an object (this is the single active unit in Figure 7) to engage in an action that an adult would define as ballistic motion on the left ($N = 8$) and right ($N = 8$) sides of network space. For these events, at time $t = 1$, the activation value of the input unit located in Row 5 Column 1 (this is what is depicted in Figure 1) of the motion group was set to 1 (the activation values of the remaining units were set to 0), and the network's task was to activate the output unit located in Row 4 Column 2 at time $t = 2$ in the corresponding motion group. At time $t = 2$, the activation value of the input unit located in Row 4 Column 2 of the motion group was set to 1, and the network's task was to activate the output unit located in Row 3 Column 3 in the corresponding motion group. The "apex" of the object's trajectory occurred in Row 2 Column 4. This continued until the object reached Row 5 Column 7 of the motion group or the right edge of the motion group. Second, networks learned that animate entities could also *fail* to cause ballistic motion in the abovementioned object on the left ($N = 8$) and right ($N = 8$) sides of network space. The real-world equivalent to this would be situations in which a person finds that an object is too heavy to displace. To instantiate the absence of ballistic motion, the object remained in Row 5 Column 1 across all time steps. Third, the network learned that objects neither caused ballistic motion in other

Figure 7

The Motion Trajectory of an Object in Simulation 2

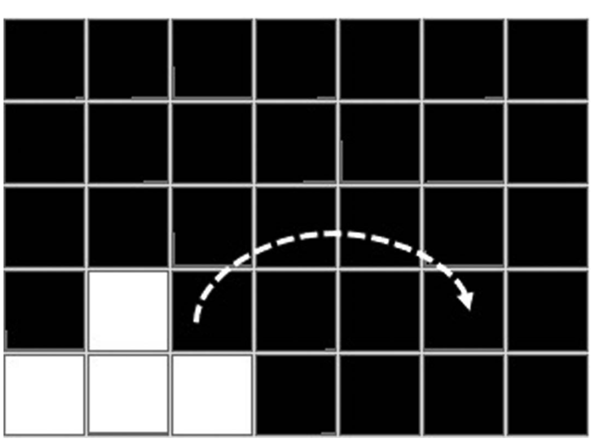


Note. The white dashed arrow indicates the motion path of that object.

objects from the left (N = 16) nor from the right (N = 16) sides of network space. In addition, networks learned that animates themselves could engage in ballistic motion—this would be akin to a person jumping in a ballistic motion trajectory (e.g., an Olympic-like "long jump")—irrespective of whether another animate was located on the same side (N = 16) or different side of network space (N = 16). To simplify these latter events, only a single animate entity—represented by four units (Figure 8)—was used in the motion group. To equate the number of time steps that it took for an object or animate entity to reach the right edge of the motion group—given that the object (represented by a single unit) in the motion group was physically smaller than the animate entity in the motion group and thus should take longer to reach the right edge than the animate entity—the animate remained at the left edge of the motion group for two time steps before beginning to move and then remained at the right edge of the motion group for two time steps after it had completed its (ballistic motion) trajectory. Finally, as in Simulations 1a and 1c, networks also experienced different amounts (i.e., the six different frequencies) of atypical object action. In the current simulation, this means events in which objects with inanimate features cause ballistic motion in other objects.

Despite differences in their particulars, Simulations 1 and 2 instantiated the idea that infants' looking behavior in Spelke et al. (1995) and Saxe et al. (2005) was based on learned associations between some attribute A and some number of other attributes X and Y and that between some attribute ¬A and one of the previous attributes X. In terms of the real world, infants may have learned that things with features that are characteristic of animate entities (attribute A) not only can engage in ballistic motion themselves but can either cause adult-defined ballistic motion in inanimate objects (attribute X) or fail to cause ballistic motion (attribute Y) in those objects as well. In contrast, infants might learn that things with features that are characteristic of inanimate objects (attribute ¬A) cannot produce ballistic motion (attribute Y) in other things. Instantiating the same mechanistic idea across simulations was paramount given that the overarching goal of this article was to demonstrate that the same hypothesized mechanism of change could explain the *emergence* of infants' knowledge about the causal properties of animate entities and inanimate objects without recourse to core knowledge.

Figure 8
The Animate Used in Simulation 2



Note. This object also underwent ballistic motion (represented by the inverted U-shaped arrow).

Habituation

Twenty networks were assigned to the Hand, Train, and Puppet conditions (total N = 60). These corresponded, respectively, to the experimental conditions to which infants were assigned in Experiments 1 and 2 of Saxe et al. (2005). Given that infants were habituated to events in which a beanbag or puppet was thrown over a wall from both sides of the stage, networks were habituated to ballistic motion on both sides of the network's visual field. Crucially, the objects and entities used during habituation were not shown during pretraining. Finally, the length of the habituation phase was four epochs.

Testing

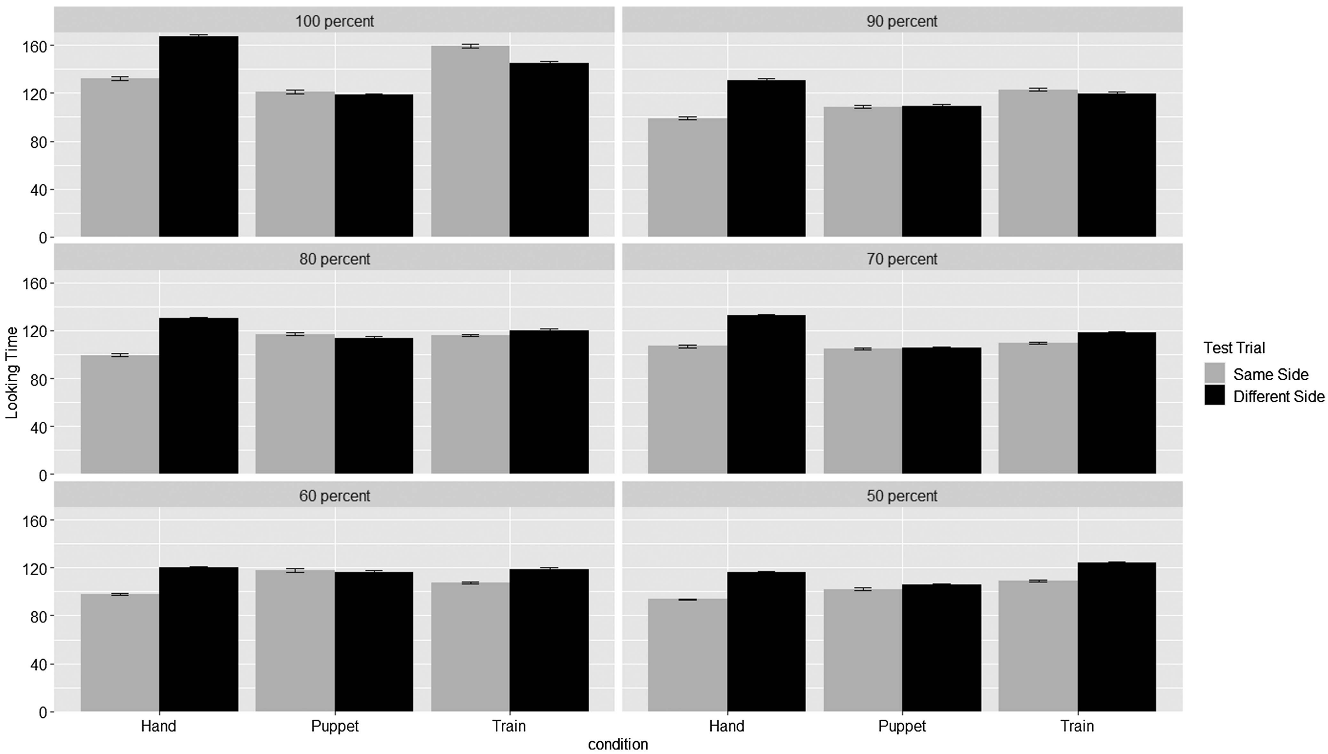
At test, networks received eight Same Side (four Same Side Left; four Same Side Right) and eight Different Side (four Different Side Left; four Different Side Right) test trials. On the Same Side trials, after the inanimate object (akin to the beanbag in Saxe et al., 2005; for networks in the Hand or Train condition) or animate entity (akin to the puppet in Saxe et al., 2005; for networks in the Puppet condition) was thrown, either another animate entity (akin to the human hand) or inanimate object (akin to the train) was revealed on the side of the stage from which the first inanimate object (beanbag) or animate entity (puppet) emerged. On the Different Side trials, either a second animate entity (i.e., a hand) or inanimate object (i.e., a train) was revealed on the opposite side of the stage from which the first inanimate object (beanbag) or animate entity (puppet) emerged. Network cross-entropy error—averaged over the Same and Different Side test trials—produced by both test events was used as a proxy for looking time. Note that just as two novel objects were shown during the test phase in Simulation 1a, two novel objects were shown here. Two objects were shown during the Same and Different Side Right events; the other objects were shown during the Same and Different Side Left events. This explains how the number of test events can go from four in Simulation 1a to eight in Simulation 2, even though both simulations used exactly the same number of objects at test.

Result

Figure 9 shows the networks' mean "looking times" to the Same Side and Different Side test trials across the Hand, Train, and Puppet conditions for each of the six frequencies. As was the case in Saxe et al. (2005), networks assigned to the Hand condition "looked" longer at the Different Side test events than at the Same Side test events. In contrast and like infants in Saxe et al. (2005), networks assigned to the Puppet or Train conditions looked about equally long at the two test events. Crucially, both patterns of results—that is, longer looking to the Different than the Same Side test events for networks assigned to the Hand condition and about equal looking to the same events for networks assigned either to the Puppet or Train conditions—were obtained across all six frequencies. Thus, even when surface features are not perfectly diagnostic of an object or entity's agency status as in the 100% case, networks had no difficulty learning that agents cannot cause objects to undergo ballistic motion

Figure 9

Networks' Mean "Looking Time" (i.e., Cross-Entropy Error) to the Same Side and Different Side Test Trials Across the Hand, Train, and Puppet Conditions Across Six Frequencies



if both occupy opposite sides of space. Additionally, networks, like infants, learned that objects with animate features can, themselves, engage in ballistic motion irrespective of another object's location in space.

One potential objection to the current set of results is that networks assigned to the Train condition behaved differently than infants in the same condition in Saxe et al. (2005). This is because the model's behavior following exposure to the 100% frequency differed from that following exposure to the 50% frequency: Networks trained on the former frequency looked longer at the Same Side test trial than at the Different Side test trial. In contrast, networks trained on the latter frequency showed the opposite pattern. Here, it is important to bear in mind that across all frequencies the networks' pattern of looking to the Same and Different Side events in the Train condition varies. For example, although it is true that for the 100%, 90%, and 50% frequencies the network looked slightly longer at the Same Side test event than at the Different Side test event, the pattern reverses for the 80%, 70%, and 60% frequencies. Thus, the pattern of looking to these events varied unsystematically across frequencies (unlike the systematic behavior across frequencies for networks in the Hand condition) and likely resulted from the random initialization of the weights in the network.

Discussion

Simulation 2 extended Simulation 1a to examine whether domain-general associative learning was sufficient to explain infants' looking behavior in Saxe et al. (2005). These authors found that infants assigned to the Hand condition looked significantly longer at the Different Side test event than at the Same Side test event, whereas infants assigned either to the Train condition or to the Puppet condition looked about equally long at both events. The present simulation results provided a qualitative match to those of Saxe et al. (2005). Together, these results indicate that an associative-learning mechanism—instantiated in an artificial neural network—is sufficient to explain not only how infants acquire causal knowledge about people and objects (Simulation 1a) but how they might acquire causal knowledge about animate entities and inanimate objects broadly (Simulation 2). It should be noted here that I chose not to simulate Experiment 3 in Saxe et al. (2005). In this experiment, infants were not allowed to handle the beanbag before habituation. Given that there was no mechanism for physically handling objects in the present simulation, a simulation of Experiment 3 would have produced identical results to the simulation of Experiment 1. Thus, it seemed unnecessary to model Experiment 3.

The next two simulations were designed to examine further the explanatory breadth of the present associative-learning account. Simulation 3 was carried out to show that associative learning is sufficient to account for infants' looking behavior in a follow-up study to Saxe et al. (2005) by Saxe et al. (2007). This study examined whether 7- and 10-month-olds understood that animate entities, but not inanimate objects, can cause ballistic motion in other objects, rather than the question posed by Saxe et al. (2005), which was whether infants understand that ballistic motion between an object and an agent must occur on the same side of physical space. Saxe et al. (2007) found that 7- and 10-month-olds looked longer when the beanbag emerged from the side on which the train was located than when it emerged from the side on which the human hand (Experiment 1) or novel puppet (Experiment 2) was located. In addition, Experiment 3 revealed that infants looked longer when a hand emerged from the opposite side from which the beanbag was thrown than when it emerged from the same side from which the beanbag thrown. Given that Simulation 2 demonstrated that a neural network will also look longer at Different Side test trials than at Same Side test trials when it involved a human hand, the simulation of Experiment 3 here is essentially a replication of Simulation 2. In addition, given that the only thing that distinguished animate entities from inanimate objects in the present series of simulations was activation along the feature bank of units—that is, I did not model specific kinds of objects (e.g., vehicles, dolls, toys, etc.)— I chose not separately to model Experiments 1 (which involved human hands) and 2 (which involved a puppet object with hair and eyes) in Saxe et al. (2007). One of the motivations for not modeling specific body parts is because, as I mentioned at the outset, it remains unanswered whether infants differentially attend to the surface features of animate entities. Given that this question remains largely open, here it was crucial not to make assumptions about the features to which infants do and do not attend. Thus, Simulation 3 reported two simulations. The first simulation modeled broadly the results from Experiments 1 and 2 in Saxe et al. (2007). Although it was the case in Saxe et al. (2007) that a human hand was used in Experiment 1 and a puppet in Experiment 2, respectively, the behavioral results of both experiments were equivalent. The second simulation modeled Experiment 3 in Saxe et al. (2007), which again is a replication of the Hand condition in Simulation 2.

Simulation 3: Saxe et al. (2007)

Simulation 3 tested the explanatory breadth of the present associative-learning account to determine whether it could explain infants' looking-time data across each of the three experiments in Saxe et al. (2007).

Method, Training, and Testing

The network architecture and the values of the learning parameters used in Simulation 3 were identical to that in Simulation 2.

Pretraining

The pretraining phase—including the length of this phase—was identical to that used in Simulation 2 except that these networks in the current simulation did not additionally learn that things with animate surface features could themselves engage in ballistic motion. This is because this fact was not tested in Saxe et al. (2007).

Habituation

Following pretraining, 20 networks were used to simulate Experiments 1 and 2 in Saxe et al. (2007), and 20 networks were used to simulate Experiment 3 in the same study. The habituation phase used here was identical in all respects to that used in Simulation 2.

Testing

Following habituation, networks in the simulation of Experiments 1 and 2 of Saxe et al. (2007) received eight "Hand" trials—in which four distinct objects with animate features caused ballistic motion in an object—and eight "Train" trials, in which four distinct objects with inanimate object surface features caused ballistic motion in an object. Crucially, all eight test stimuli (i.e., the four animate entities and the four inanimate objects) were not experienced during the pretraining phase. As before, network cross-entropy error served as a proxy for infant looking time.

Results

Figure 10 shows the mean "looking times" to the Hand and Train test trials across all six frequencies for networks used to simulate Experiments 1 and 2 in Saxe et al. (2007). Figure 11 shows the mean looking times to the Same Side and Different Side test trials for networks used to simulate Experiment 3 in the same study across the same six frequencies. As shown in Figure 10, regardless of how frequently objects engaged in atypical action during pretraining (i.e., objects causing other objects to undergo ballistic motion), networks looked longer when the beanbag emerged from the side on which the toy train was located than when it emerged from the side on which the human hand was located. Additionally, as is shown in Figure 11, networks looked longer when the beanbag emerged from the side on which the human hand was not located than when it emerged from the side on which the human hand was located. This latter result replicated that from Simulation 2.

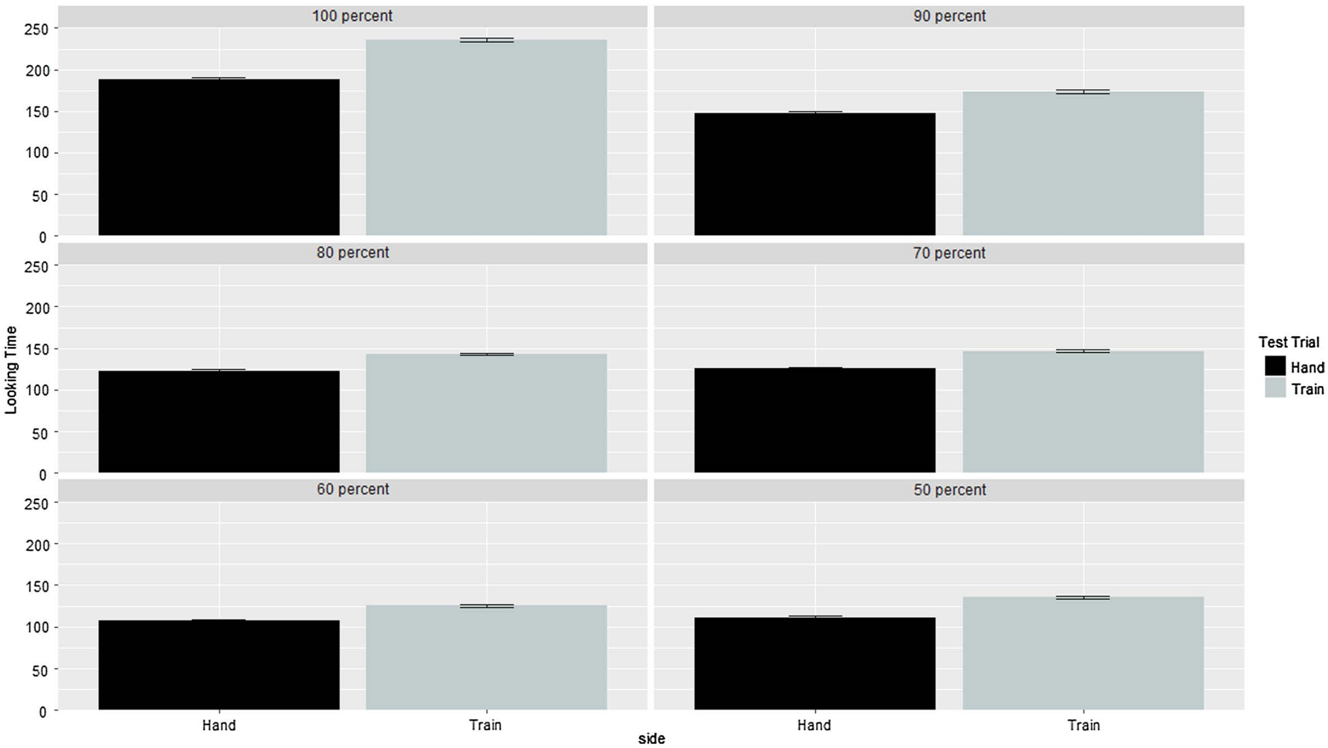
Discussion

The results of this simulation qualitatively matched the behavioral data from Saxe et al. (2007): Networks looked longer at the Train trials than at the Hand trials (Figure 10). Likewise, networks looked longer at the Different Side than at the Same Side test trials (Figure 11). Importantly, as has been the case to this point, these results largely did not depend on how frequently objects engaged in atypical action during pretraining. Combined with the results from Simulations 1 and 2, these results demonstrate that the present associative-learning account is sufficient to account for the behavioral data across the separate experiments in Saxe et al. (2007). In addition, the results from Simulations 2 and 3 are important because they demonstrate the explanatory breadth of the associative-learning account discussed at the outset of this article; this account could explain infants' looking behaviors in Spelke et al. (1995) but could also account for the studies by Saxe et al. (2005) and Saxe et al. (2007).

The goal for the final simulation was to push the model further to determine whether it (and by extension, the present associative-learning account) could capture infants' behavior in yet another classic study, by Markson and Spelke (2006). This study demonstrated that 7-month-olds expected objects shown to be self-propelled in the past, but not objects whose past movements were generated by a human hand, to begin moving spontaneously and without aid in the future, but only when the objects possessed animate-like features (e.g., eyes, hands). Although these findings have been interpreted to be the result of inborn core knowledge principles (e.g., Shutts et al., 2009), the following simulation tested

Figure 10

Networks' Mean "Looking Time" (i.e., Cross-Entropy Error) to Hand and Train Test Trials Across Six Frequencies



Note. See the online article for the color version of this figure.

whether the findings can be explained by the operation of a domain-general associative-learning mechanism.

Simulation 4: Markson and Spelke (2006)

The goal of the final simulation was twofold. First, it was designed to determine whether the present associative-learning account could be extended still further to account for the data in each of the six experiments in Markson and Spelke (2006). Second, it was designed to establish the domain generality of the present account. Given that Markson and Spelke (2006) were interested in whether infants could quickly learn about the motion properties of objects and entities rather than in infants' causal-learning abilities, demonstrating that the present associative-learning account can capture the pattern of results in Markson and Spelke (2006) can speak to its ability to capture findings across domains, and hence to the domain-generality of the current account.

Method

The model architecture used in this simulation was identical to that used in Simulation 1a with three exceptions. First, only one bank of 40 units was used to represent a single animate entity rather than two separate banks of 40 units. Second, the size of the input and corresponding output motion layers was increased from 7 to 35 to accommodate distinct representations for self-propelled and hand-generated causal action (see Figure 12). Third, the input layer included a two-unit "unrelated" group. This made it possible to simulate the unrelated activity that participants engaged in between the familiarization and test phases in Experiment 2 in Markson and Spelke (2006). In addition to these differences, the present simulation differed from Markson and Spelke (2006) in the following way: In their study, infants experienced two blocks of trials, each consisting of a set of familiarization and test trials. Preliminary simulations indicated that the results did not depend on whether networks experienced one or two blocks. Thus, to simplify the present simulations, only one block of familiarization and test trials were simulated.

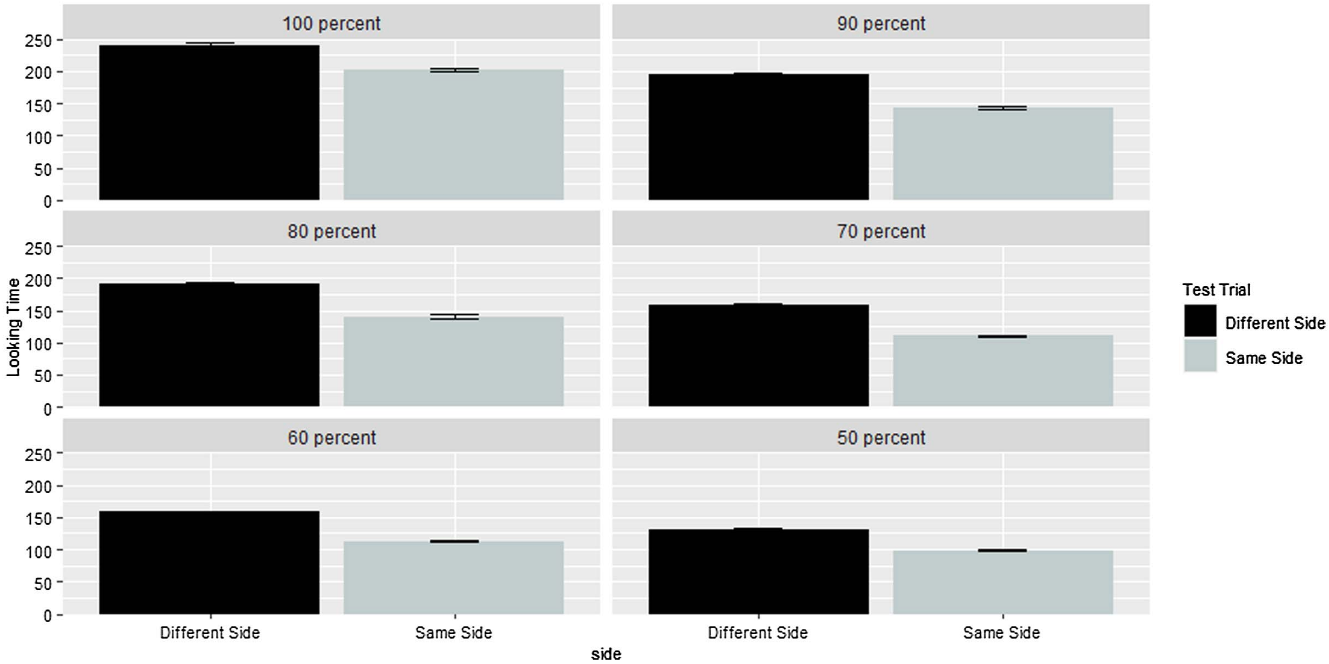
Training and Testing

Pretraining

Networks experienced four kinds of events during the pretraining phase. In one event ($N = 16$), networks learned that objects with animate features could engage in self-propelled motion. The way motion was represented in this simulation is shown in Figure 12 (right side). In another event ($N = 16$), networks learned that animate entities could be caused to move. The representation for this event is also shown in Figure 12 (left side). In a third event ($N = 16$), networks learned inanimate objects could only be caused to move. In a final event, networks experienced different amounts of atypical object action such that objects with inanimate features engaged in self-propelled motion. The length of the hand-generated and self-propelled motion events was three times steps.

Figure 11

Networks' Mean "Looking Time" (i.e., Cross-Entropy Error) to Different Side and Same Side Test Trials Across Six Frequencies



Note. See the online article for the color version of this figure.

Habituation

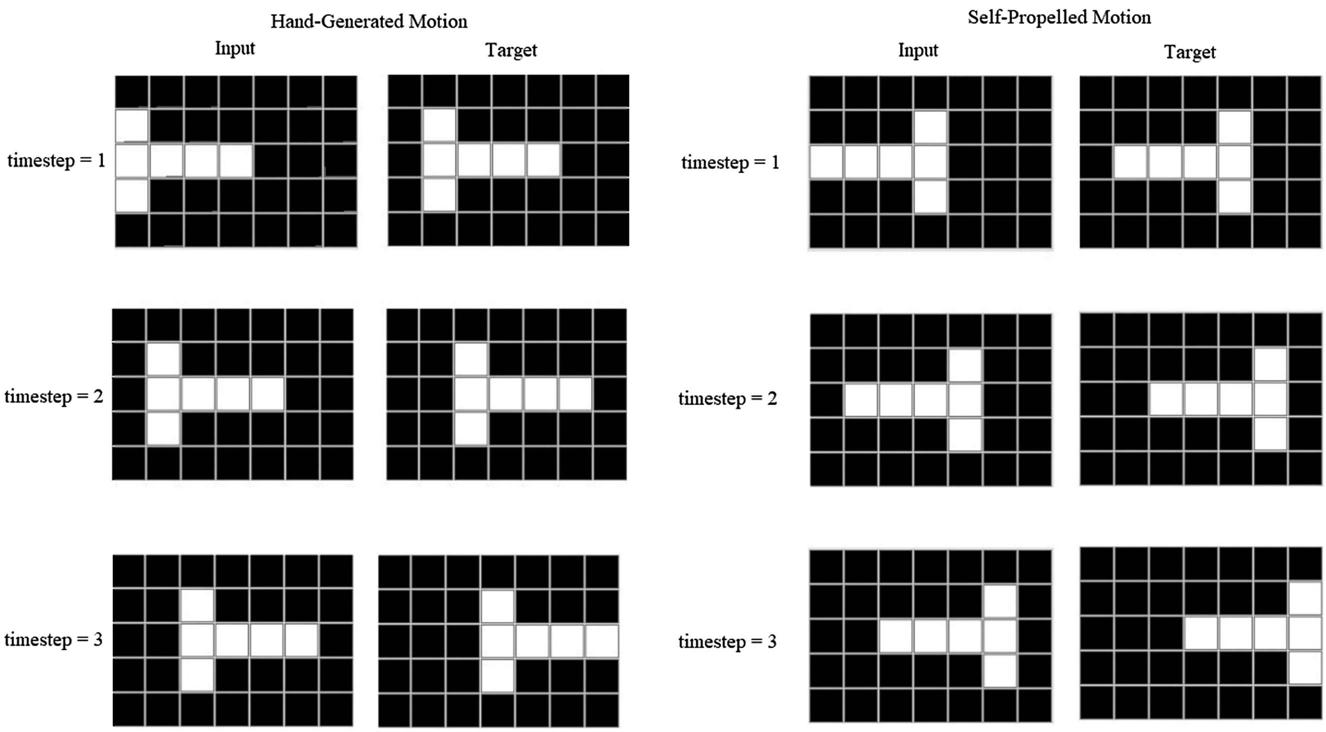
Following pretraining, 20 networks were used to simulate Experiments 1–6 (total $N = 120$). Networks used to simulate Experiments 1, 2, 3, and 5 were familiarized to two self-propelled and two hand-generated events. The simulation of Experiments 1 and 2 used animate entities; the simulation of Experiments 3 and 5 used inanimate objects. An important point to mention is that networks used to simulate Experiment 2 received an intervening "play break" between the familiarization and test phase. During this break, the network simply had to copy the pattern of activation presented as input to the input group labeled "unrelated" to and along the corresponding output group. In the current simulation, the first unit in this group was turned on, and the second unit was turned off—the networks' job was simply to copy this pattern of activation at the output layer. Given that Markson and Spelke (2006) did not specify the activities used during the intervening play break, some liberties were taken here in choosing the activity for the model at that time. Although the particular activity that the network completed during this break was of little importance, it was critical that the task be unrelated to what networks were habituated to and tested on (which presumably was the case in Markson & Spelke, 2006).

Networks used to simulate Experiments 4 and 6 were familiarized to the object to which networks used to simulate Experiments 3 and 5 were familiarized to ensure that their equivalent looking at the test events in the simulations of Experiments 3 and 5—like infants' equivalent looking in these experiments in Markson and Spelke (2006)—were not due to their inability to discriminate them. These networks were then shown this object at test along with an object to which they were not familiarized.

Before proceeding, it is important to distinguish between the dependent measure used in the study by Markson and Spelke (2006) and that used in the current simulation. In their study, Markson and Spelke (2006) posited that if infants were habituated to an event in which object A is self-propelled and an event in which object B is moved by a human hand, then infants should subsequently pay more attention to object A when both are presented motionless on opposite sides of a stage. The rationale is that infants should expect object A to continue moving on its own (as it did during habituation), whereas object B should remain motionless (because a human hand was not present to move it).

In contrast to Markson and Spelke's (2006) testing procedure, in the current simulation, networks were shown two test events. In one event, the self-propelled object from familiarization continued to engage in self-propelled motion at test. In the other event, the hand-generated object from familiarization now engaged in self-propelled motion at test. Networks were expected to "look longer" when the previously hand-generated object engaged in self-propelled motion at test than when the previously self-propelled object engaged in this same motion. Additionally, given that Markson and Spelke (2006) found that this effect was moderated by an object's animacy status, a similar moderation effect was anticipated here. Although this dependent measure differed from that used in Markson and Spelke (2006), the presumed basis for infants' looking in Markson and Spelke (2006) and the model's looking here was the same: Events in which hand-generated objects engage in self-propelled motion should be

Figure 12
An Example of the Input-Target Pairs Used in Simulation 4



Note. The left side of the figure shows the representation of hand-generated motion; the right side of the figure shows the representation of self-propelled motion.

treated as more anomalous than events in which self-propelled objects engage in the same motion.

Results

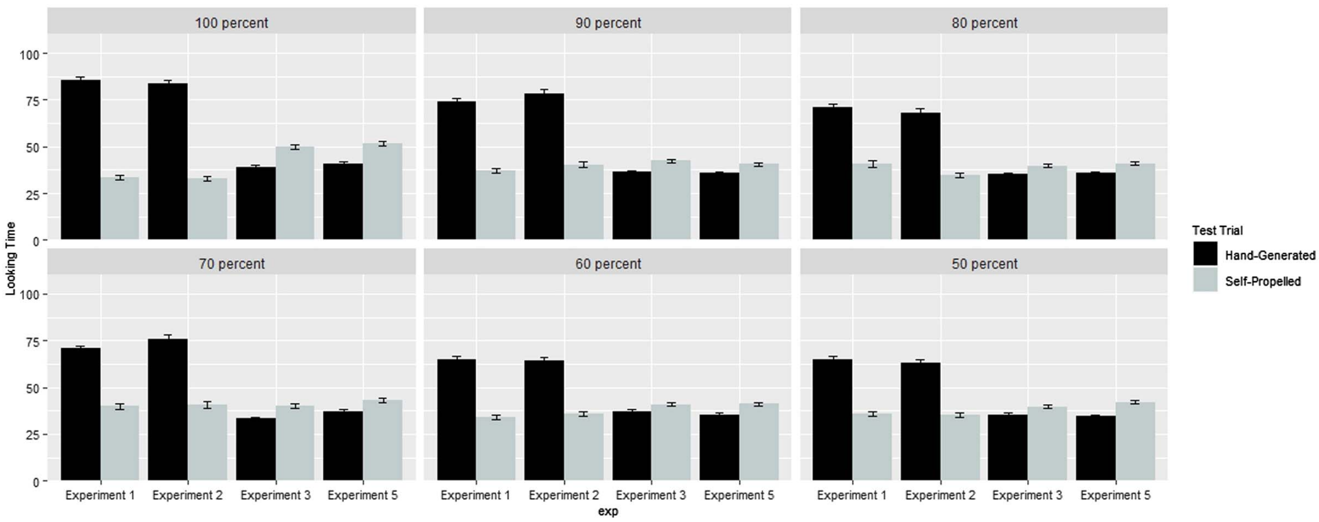
Figure 13 shows the networks' mean looking times to the Hand-Generated and Self-Propelled test events. As can be seen, across all six frequencies networks in Experiments 1 and 2—which used objects with animate features—looked longer when the hand-generated object from familiarization engaged in self-propelled motion than when the self-propelled object from the same phase engaged in self-propelled motion. In contrast, networks in Experiments 3 and 5—which used objects with inanimate features—showed slightly longer (but nearly equivalent) looking to the self-propelled test event than to the hand-generated one. Although this latter result was technically reliable (recall that it was not reliable in Markson & Spelke, 2006), three notes are worth making. First, the direction of networks' responses to the two test events was in the opposite direction as that in Experiments 3 and 5. This suggests that the networks were not simply biased to attend longer to the hand-generated test event than to the self-propelled test event; the events were treated differently based on whether they used objects with animate features (Experiments 1 and 2) or objects with inanimate features (Experiments 3 and 5). Second, it should be clear from Figure 12 that the magnitude of the difference in looking time to the two test events is considerably larger for Experiments 1 and 2 than it was for Experiments 3 and 5. Although Markson and Spelke (2006) did not examine whether this was true in their study, it is safe to assume that these authors would have found the difference in looking to the two test events to be larger in Experiments 1 and 2 compared to Experiments 3 and 5. Third, and as was mentioned in the Results section of Simulation 1a, the reason there was a difference in the amount of time that networks looked at the two test events in Experiments 3 and 5 is because the range of possible "looking" responses was presumably much narrower than that in infants. Unlike the models, infants' looking responses could vary over a much greater range.

Crucially, as can be seen above in Figure 14, networks looked longer at the new test stimulus than at the old test stimulus, and this was true across all six frequencies. This result indicates that, like infants in Markson and Spelke (2006), networks' nearly equivalent treatment of the two test events during Experiments 3 and 5 was not due to their inability to discriminate the familiarization stimuli used in them.

Discussion

The goal of Simulation 4 was to model all six experiments in Markson and Spelke (2006). This study examined whether infants expected objects that were shown to be self-propelled in the past to continue to be self-propelled in the future and objects shown to be caused to move by a human hand in the past to continue to be caused to move in the future. Markson and Spelke (2006) found this to be the case. Similarly, the current data

Figure 13
Networks' Mean "Looking Time" (i.e., Cross-Entropy Error) to Hand-Generated and Self-Propelled Test Events (Left Side) and Their Looking Time to the New and Old Test Trials (Right Side)



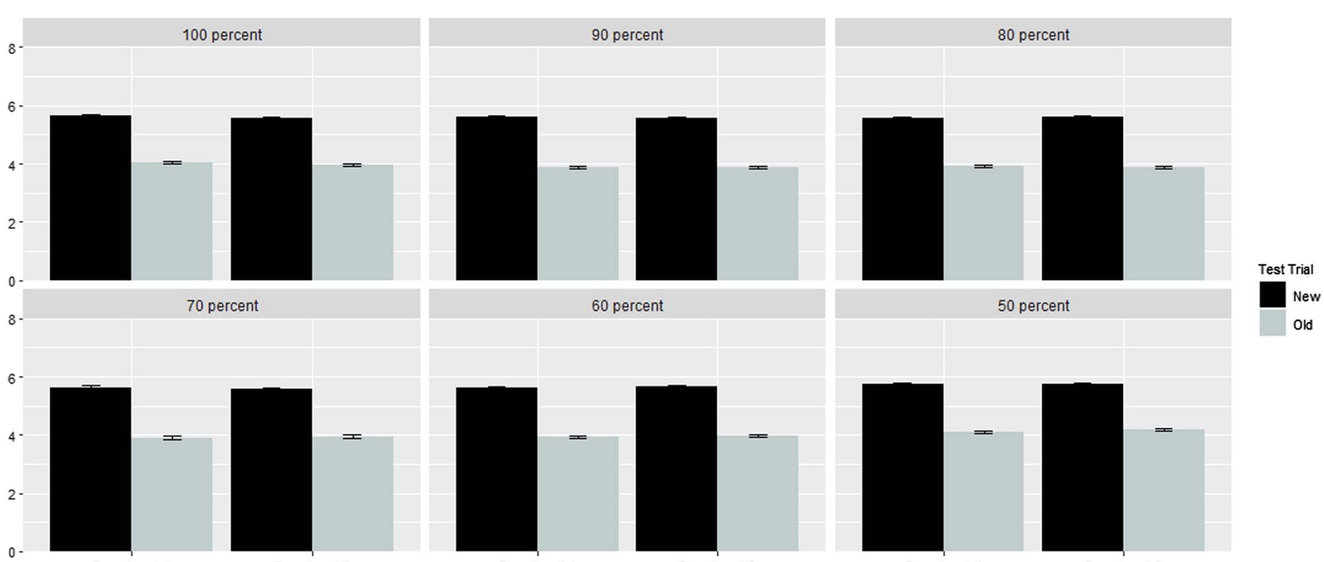
Note. exp = experiment. See the online article for the color version of this figure.

indicated that networks could also learn these causal relations. Perhaps most importantly, Simulation 4 extends the earlier simulations to show that an associative-learning mechanism continues to be sufficient to explain how infants might learn about other causal features of the world such as about the causal factors that produce motion in objects and entities. Two broad takeaways can be gleaned from the present series of simulations. First, they demonstrate that core knowledge is not necessary to explain infants' causal knowledge about people and inanimate objects. Second, they demonstrate that associative learning is sufficient to explain how this knowledge might emerge throughout development.

General Discussion

The goal of this article was to show that domain-general associative learning—implemented in an artificial neural network—is sufficient to account for how infants learn about the various causal properties of animate entities and inanimate objects. The crux of the

Figure 14
Networks' Mean "Looking Time" (i.e., Cross-Entropy Error) to the New and Old Test Trials



Note. See the online article for the color version of this figure.

associative-learning account presented here is that infants learn to associate whatever low-level surface features distinguish animate entities from inanimate objects with multiple kinds of perceptual-based, low-level, kinematics depictions of causal action, whereas they learn to associate whatever surface features distinguish inanimate objects from animate entities with one (or a small number of) perceptual-based descriptions of causal action. This may explain why across multiple studies that examine infants' causal knowledge about animates and inanimates, infants tend to be less surprised when animate entities engage in multiple forms of causal action (e.g., action at a distance vs. action on contact) than when inanimate objects engage in multiple forms of causal action.

Simulation 1a had two aims. First, it examined whether an associative-learning mechanism could account for infants' looking behavior in Spelke et al. (1995). Second, it tested "younger" models to determine whether the learned associations by the model between different features and different actions undergo a developmental progression. The results from this simulation revealed that an associative-learning mechanism could explain infants' looking behavior in Spelke et al. (1995)—the models' pattern of looking mirrored infants' looking behavior in Spelke et al. (1995). In addition to capturing the original set of findings by Spelke et al. (1995), the model made a testable prediction: Infants' knowledge about the causal action of people and objects should undergo a developmental progression such that infants younger than 7 months of age should fail to show increased looking when objects cause other objects to move at a distance. The basis for this is simple: With less experience with objects and people in the real world, younger networks will not have had enough time to encode the relevant associations between different features of objects and people and different kinds of causal action.

Simulations 1b and 1c were carried out to resolve two interrelated issues. The first issue concerned whether an associative-learning mechanism could "discover" a causally relevant feature when that feature was as correlated with different kinds of causal action as another, less causally relevant feature. The second issue concerned how an associative-learning mechanism treated "hybrid" objects such as objects with some combination of salient and less salient animate and inanimate features. Infants are likely to encounter such instances in the real world, even if they experience them infrequently, and thus it seems important to determine how they treat those instances. This issue is especially important to resolve if, as is claimed here, the primary mechanism they use to learn about people and objects is associative learning. The results of Simulation 1b indicated that an associative-learning mechanism can extract causally relevant features if these features are made more salient than another feature; a separate mechanism was not needed. The results of Simulation 1c revealed that "real-world" experience with unmodified people and objects as well as modified objects did not influence how networks processed unmodified people and objects at test. In addition, the effect that this real-world experience had on processing modified people and objects at test depended on the nature of the features possessed by such people and objects: Networks showed heightened looking when modified stimuli with salient inanimate features caused action at a distance in other objects with salient inanimate features compared to when the same objects interacted through contact. In contrast, networks were at chance to both events when those events used modified stimuli with salient animate features. This latter result is important because it makes a testable prediction that should be explored in future research: Infants who are habituated to and tested on modified people and objects should show the same pattern of looking as networks in Simulation 1c. This prediction is theoretically significant because it is presumably at odds with what might be predicted if infants' knowledge about people and objects was subserved by core knowledge. The core knowledge account would predict that the inferences that infants make about the causal capacities of people and objects should be (at least somewhat) impervious to the surface features of those objects and people. This is because infants' causal knowledge about people and objects is thought to be abstract and not tied to the features of any one object. This means that it should be possible to determine whether core knowledge or an associative-learning mechanism underlies infants' causal knowledge about people and object by testing infants younger than 7 months of age as well as exposing infants to modified people and object stimuli.

Simulations 2 and 3 extended this simulation to show that the same associative-learning mechanism could account for the experimental data in Saxe et al. (2005) and Saxe et al. (2007). These simulations showed that associative learning alone is sufficient not only to explain how infants learn that animate entities, but not inanimate objects, can cause action at a distance in other entities (this is what Simulation 1a showed) but how they learn that animate entities, but not inanimate objects, can cause ballistic motion in other things. The networks, like the infants, expected objects with animate features not only to be capable of ballistic motion themselves but to be able to cause other objects to engage in that motion; the networks held no such expectation for objects with inanimate features. Finally, Simulation 4 accounted for the experimental data across all six experiments in Markson and Spelke (2006). This study examined whether infants could quickly learn that some objects can engage in self-propelled motion whereas other objects must be made to move through hand-generated motion. This simulation demonstrated that infants' knowledge about the self-propelled motion of objects can also be explained by the operation of an associative-learning mechanism; there is no need to assume that the capacity to quickly associate objects with different kinds of action is supported by core knowledge or core cognition. In total, the model was able to capture the behavioral data across 13 separate experiments that spanned four different studies as well as address several potential objections and make predictions for additional experiments that should be tested in the future. Together, the present series of simulations suggest that innate knowledge and specialized core systems may not be necessary to explain how infants acquire causal knowledge about animate entities and inanimate objects broadly; domain-general associative learning alone is sufficient to explain the origins of infants' causal knowledge about animates and inanimates.

Potential Criticisms

Nonetheless, three potential criticisms are worth noting. The first concerns the limited number of studies that were the focus of the present simulations. For example, one study that could have been but ultimately was not modeled here was that by Muentener and Carey (2010). This study demonstrated that 8-month-old infants ostensibly understand that inanimate objects require contact to move and that things with hands, but not inanimate objects (e.g., a toy truck), can cause state changes in other things (e.g., causing a box to break into pieces). The rationale for not including more studies in the present series of simulations was that it was not possible to model all of the potentially relevant past and present experiments—this list is simply too long. More to the point, the goal here was to show that when a subset of these studies is considered—each of which has received considerable attention in the developmental literature and thus presented ideal test cases—an associative-learning mechanism can account for the full range of data. Still, there is no reason to think that the present account could not be extended to explain other studies such as that by Muentener and Carey (2010). For instance, this could be achieved by extending the present simulations to include pretraining events in which things with animate-like features, but not things that lack such features, can cause state changes in other physical objects through direct contact. The network should subsequently show "longer looking" when things with animate features cause state changes in inert objects at a distance but not on contact, but the same network should be equally surprised when an inanimate object causes state changes on contact or at a distance.

A second criticism is that the present associative learning can only explain the narrow range of studies that were the focus of the present series of simulations but cannot be extended to explain studies that use stimuli with little-to-no cues to animacy. An example of this is a recent study by Stojnić et al. (2023). Eleven-month-old infants were familiarized to one or two of a small number of "benchmark" tasks that were designed to assess infants' common-sense reasoning abilities. Half of these tasks focused on infants' ability to attribute goals to an agent; the remaining half assessed infants' capacity to distinguish between efficient and inefficient actions. For example, in one of the goal attribution tasks, called the goal-directed task, infants were familiarized to an event in which an animated simple shape repeatedly approached one of two inert simple shapes in a simple grid world—the inert shape that the animated shape repeatedly approached can be thought of as the animated shape's goal object. Following familiarization, infants were shown two test events. In the New Location test event, the animated simple shape approached the old object, but in a new location (relative to the location seen during familiarization). In the New Object test event, the animated shape approached a new object, which was located in the old object's original, familiarization location. These authors followed a similar procedure to test infants' understanding of action efficiency. For example, in one of the rationality attribution tasks, called the Efficient-Agent task, infants were familiarized to an event in which a different but equally simple, animated shape followed a curvilinear "rational" path to reach an inert simple shape whose location was obstructed by an obstacle—the curvilinear path was justified because it allowed the animated shape to bypass the obstacle that separated it from the target object. Infants were then shown two test events. In the Efficient test event, because the target object was no longer obstructed by the obstacle, the animated shape was able to follow a direct, linear, and "rational" path to reach the target object. In the Inefficient test event, although it was the case that the target object was no longer obstructed, the animated shape nonetheless moved along a curvilinear (and now unwarranted or irrational) path to reach it. The results suggested that infants not only attributed goals to an agent but recognized whether that agent's actions were efficient: Infants looked longer at the New Object test event than at the New Location test event in the goal-directed task. In contrast, in the Efficient-Agent task, infants looked longer at the Inefficient test event than at the Efficient test event. These authors interpreted these results to mean that infants "expect agents' actions to be goal directed towards objects, not locations, and that they expect agents' goal-directed actions to be rationally efficient" (p. 4).

Although it may be tempting to conclude that these results in fact demonstrate that infants can attribute goals to objects and determine whether an action is efficient, I see no reason why these results could not also be explained by the operation of simpler, domain-general learning mechanisms such as habituation and associative learning. For example, infants may well have looked longer at the New Object test event than at the New Location test event in the goal-directed task and longer at the Inefficient test event than at the Efficient test event in the Efficient-Agent condition simply because infants failed to fully encode the events to which they were familiarized. As a result, infants may have been drawn more toward the perceptually more familiar test events (i.e., the New Object and Inefficient test events) than toward the perceptually novel events (i.e., the New Location and Efficient test events). This contention is supported by a rich history that has established that if infants fail sufficiently to encode an event, they will continue to orient toward that event to complete their encoding of it. This idea was perhaps best captured by Sokolov's (1960) neuronal model of habituation. According to this model, habituation is the process whereby learners continually compare their internal model of a stimulus to the observed stimulus until their internal model matches the observed stimulus. On this model, then, infants should be more likely to show a familiarity preference (i.e., longer looking to the familiar test event than to the novel test event) than a novelty preference (i.e., longer looking to the novel test event than to the familiar one) if there is a mismatch between the observed and internally represented stimulus; that is, infants should look longest at a familiar stimulus than at a novel one if the representation of the associative relations among the features of the familiar stimulus is not sufficiently well encoded to support looking elsewhere.

A similar idea was put forward by Hunter and Ames (1988; for empirical evidence of this perspective, see Bogartz et al., 1997, 2000; Hunter et al., 1982, 1983; Rose et al., 1982). The crux of their view is that whether an infant shows a familiarity or novelty preference depends on the amount of time that they are familiarized to a stimulus or event as well as on the complexity of that stimulus or event: Infants should be more likely to show a familiarity preference if the stimuli or events to which they are familiarized are complex *and* the amount of time that they are familiarized to those stimuli is relatively short. This is precisely the situation that infants were presented with in Stojnić et al. (2023): Although the events used simple shapes, the events themselves were relatively complex and abstract. This, combined with the fact that the study relied on experimenter-defined familiarization—which is known to induce familiarity preferences (e.g., Bogartz et al., 2000; Cohen & Marks, 2002)—rather than infant-controlled habituation, could have resulted in a familiarity preference rather than a high-level assessment about an agent's goal and the rationality of its actions.

A final potential criticism concerns the lack of behavioral data. The omission of behavioral data was a deliberate choice rather than an oversight. My primary objective was to put forward a theoretical account for how infants acquire causal knowledge about animate entities and inanimate objects as well as provide a proof of concept that the perspective—when implemented in an artificial neural network—can provide a developmentally and mechanistically inspired explanation for how infants acquire causal knowledge of animates and inanimates. In other words, my goal was to offer a mechanistically focused theoretical account of how infants might acquire causal knowledge about objects and entities that has potential to address existing gaps in the literature and that can form the basis of new research.

Two Remaining Issues

Before closing, two issues are worth addressing. First, it is worth touching on the status of the debate about the origins of infants' knowledge and concepts. Second, it is worth discussing what can be gleaned from determining whether a given learning mechanism or process is domain-general or domain-specific. In terms of the first issue, although it is the case that some researchers have begun to urge the field to move beyond the "nativist–empiricist" debate and to adopt a "developmental cascades" view of development, which emphasizes that developmental milestones represent points in a cumulative cascade of events and that development itself is best viewed as continuous, interconnected, cumulative, and context-dependent (e.g., Malachowski & Needham, 2023; Masten & Cicchetti, 2010; Oakes & Rakison, 2019; Spencer et al., 2009), the debate about the origins of early knowledge and the mechanisms that underpin that knowledge is still being hotly debated (for one extremely recent example of this, see Vong et al., 2024). One reason for this is simply that many of the foundational issues that are at the center of this debate largely remain unresolved. In the sociomoral domain, for example, there continues to be trenchant debate among researchers and theorists about how infants and young children learn to evaluate others based on their social and moral actions. In other words, how do infants learn to distinguish between prosocial and antisocial beings? A specific issue is whether sociomoral reasoning in infants—that is, their capacity to distinguish between prosocial and antisocial beings—is supported and can be described by an "innate moral core" (e.g., Hamlin, 2013; Woo et al., 2022) and rational processes such as naïve utility calculus (e.g., Jara-Ettinger et al., 2016; Powell, 2022) or whether such abilities are learned via domain-general capacities such as associative learning (Benton & Lapan, 2022; Scarf et al., 2012) that operate over low-level, perceptually based information and cues processed by the visual system (Malik & Isik, 2023; McMahon & Isik, 2023).

A similar debate is playing out in the face-perception literature. One of the questions that researchers continue to focus on in this literature is how human face expertise arises as well as what the basis is for infants' early emerging preference for facelike stimuli. On the one hand, some researchers argue that face expertise is acquired via the operation of domain-general learning mechanisms such as associative learning (e.g., Scott & Arcaro, 2023) and that the bias preferentially to attend to human faces and facelike stimuli arises from nonspecific structural properties of the infant visual system and biases such as biases for stimuli that are "top-heavy" and congruent (Macchi Cassia et al., 2008; Simion et al., 2002; Simion & Giorgio, 2015). On the other hand, other researchers maintain that face expertise and the early bias for human faces is in place at birth, is modular, is domain-specific, and has a basis in specialized neural circuits that are present in young infants (e.g., Kosakowski et al., 2022; Powell et al., 2018).

Finally, in the causal-learning domain, there is no consensus among researchers about what the processes are that underlie causal reasoning in children. For example, an open question in this field concerns whether children are "little scientists" (e.g., Gopnik & Wellman, 1992, 2012) who use explicitly structured representations, which are thought to be innate (e.g., Gopnik et al., 2004), combined with rational processes that approximate Bayesian inference to make causal judgments and decisions (e.g., Bonawitz et al., 2014; Gopnik & Bonawitz, 2015; Gopnik & Tenenbaum, 2007) or whether such judgments, inferences, and decisions reflect the operation of much simpler processes such as associative learning (e.g., Benton et al., 2021, 2024; Kloos & Sloutsky, 2013; McClelland & Thompson, 2007). Thus, it is clear that far from being a resolved (and now outdated) issue, the nativist–empiricist debate is alive and well.

In terms of the second (and related) issue, what value, if any, is there in determining whether the cognitive mechanism or mechanisms that infants rely on to acquire knowledge and concepts are either domain-general or domain-specific? One reason it is important to determine the nature of the mechanisms that infants use to learn about the world is that such knowledge has implications about the effectiveness and scope of clinical treatment and intervention. For example, if ultimately it is determined that infants' knowledge about the causal properties of people and objects arises from a domain-general learning mechanism—as I am arguing in this article—then interventions applied to aspects of that mechanism not only should impact when, whether, how, and to what extent infants acquire causal knowledge about people and objects but should extend to affect knowledge acquisition in other domains. This is because, by their very nature, domain-general learning mechanisms are processes that support learning across a wide range of domains and content areas and are triggered by a wide range of input types. However, if infants' knowledge about the causal properties of people and objects is subserved by a domain-specific learning mechanism, then interventions to different aspects of the learning mechanism should affect learning only in those domains that fall under the purview of the mechanism.

Conclusion

A central question in the field of cognitive development has concerned the origins of infants' knowledge about the causal properties of people and objects. One answer has been to assume that this knowledge may be unlearned and may derive from evolutionarily ancient, specialized mechanisms (e.g., Gelman, 1990), modules (e.g., Leslie, 1995), and core knowledge and core cognition systems (e.g., Carey, 2009; Spelke, 2022). The present simulations illustrate that appeals to such specialized processes and knowledge may not be necessary to account for how infants come to learn about the causal properties of the various things in the world. In this article, I have shown that associative learning as instantiated in an artificial neural network is sufficient to explain infants' looking behavior in four classic studies that consisted of a total of 13 experiments.

References

Ayzenberg, V., & Behrmann, M. (2024). Development of visual object recognition. *Nature Reviews Psychology*, *3*, 73–90. https://doi.org/10.1038/s44159-023-00266-w

Baillargeon, R. (1987). Object permanence in 31/2- and 41/2-month-old infants. *Developmental Psychology*, *23*(5), 655–664. https://doi.org/10.1037/0012-1649.23.5.655

Bélanger, N. D., & Desrochers, S. (2001). Can 6-month-old infants process causality in different types of causal events? *British Journal of*

Developmental Psychology, *19*(1), 11–21. https://doi.org/10.1348/026151001165930

Benton, D. T., Kamper, D., Beaton, R. M., & Sobel, D. M. (2024). Don't throw the associative baby out with the Bayesian bathwater: Children are more associative when reasoning retrospectively under information processing demands. *Developmental Science*, *27*(3), e13464. https://doi.org/10.1111/desc.13464

Benton, D. T., & Lapan, C. (2022). Moral masters or moral apprentices? A connectionist account of sociomoral evaluation in preverbal infants. *Cognitive Development*, *62*, Article 101164. https://doi.org/10.1016/j.cogdev.2022.101164

Benton, D. T., Rakison, D. H., & Sobel, D. M. (2021). When correlation equals causation: A behavioral and computational account of second-order correlation learning in children. *Journal of Experimental Child Psychology*, *202*, Article 105008. https://doi.org/10.1016/j.jecp.2020.105008

Bogartz, R. S., Shinskey, J. L., & Schilling, T. H. (2000). Object permanence in five-and-a-half-month-old infants? *Infancy*, *1*(4), 403–428. https://doi.org/10.1207/S15327078IN0104_3

Bogartz, R. S., Shinskey, J. L., & Speaker, C. J. (1997). Interpreting infant looking: The event set x event set design. *Developmental Psychology*, *33*(3), 408–422. https://doi.org/10.1037/0012-1649.33.3.408

Bonawitz, E., Denison, S., Gopnik, A., & Griffiths, T. L. (2014). Win-Stay, Lose-Sample: A simple sequential algorithm for approximating Bayesian inference. *Cognitive Psychology*, *74*, 35–65. https://doi.org/10.1016/j.cogpsych.2014.06.003

Carey, S. (2009). *The origin of concepts*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195367638.001.0001

Cohen, L. B., & Marks, K. S. (2002). How infants process addition and subtraction events. *Developmental Science*, *5*(2), 186–201. https://doi.org/10.1111/1467-7687.00220

Elman, J. L. (1990). Finding structure in time. *Cognitive Science*, *14*(2), 179–211. https://doi.org/10.1207/s15516709cog1402_1

Gelman, R. (1990). First principles organize attention to and learning about relevant data: Number and the animate-inanimate distinction as examples. *Cognitive Science*, *14*(1), 79–106. https://doi.org/10.1207/s15516709cog1401_5

Gopnik, A., & Bonawitz, E. (2015). Bayesian models of child development. *WIREs Cognitive Science*, *6*(2), 75–86. https://doi.org/10.1002/wcs.1330

Gopnik, A., Glymour, C., Sobel, D. M., Schulz, L. E., Kushnir, T., & Danks, D. (2004). A theory of causal learning in children: Causal maps and Bayes nets. *Psychological Review*, *111*(1), 3–32. https://doi.org/10.1037/0033-295X.111.1.3

Gopnik, A., & Tenenbaum, J. B. (2007). Bayesian networks, Bayesian learning and cognitive development. *Developmental Science*, *10*(3), 281–287. https://doi.org/10.1111/j.1467-7687.2007.00584.x

Gopnik, A., & Wellman, H. M. (1992). Why the child's theory of mind really is a theory. *Mind & Language*, *7*(1–2), 145–171. https://doi.org/10.1111/j.1468-0017.1992.tb00202.x

Gopnik, A., & Wellman, H. M. (2012). Reconstructing constructivism: Causal models, Bayesian learning mechanisms, and the theory theory. *Psychological Bulletin*, *138*(6), 1085–1108. https://doi.org/10.1037/a0028044

Hamlin, J. K. (2013). Moral judgment and action in preverbal infants and toddlers: Evidence for an innate moral core. *Current Directions in Psychological Science*, *22*(3), 186–193. https://doi.org/10.1177/0963721412470687

Hunter, M. A., & Ames, E. W. (1988). A multifactor model of infant preferences for novel and familiar stimuli. *Advances in Infancy Research*, *5*, 69–95.

Hunter, M. A., Ames, E. W., & Koopman, R. (1983). Effects of stimulus complexity and familiarization time on infant preferences for novel and familiar stimuli. *Developmental Psychology*, *19*(3), 338–352. https://doi.org/10.1037/0012-1649.19.3.338

Hunter, M. A., Ross, H. S., & Ames, E. W. (1982). Preferences for familiar or novel toys: Effect of familiarization time in 1-year-olds. *Developmental Psychology*, *18*(4), 519–529. https://doi.org/10.1037/0012-1649.18.4.519

Jara-Ettinger, J., Gweon, H., Schulz, L. E., & Tenenbaum, J. B. (2016). The naïve utility calculus: Computational principles underlying commonsense psychology. *Trends in Cognitive Sciences*, *20*(8), 589–604. https://doi.org/10.1016/j.tics.2016.05.011

Kloos, H., & Sloutsky, V. M. (2013). Blocking a redundant cue: What does it say about preschoolers' causal competence? *Developmental Science*, *16*(5), 713–727. https://doi.org/10.1111/desc.12047

Kosakowski, H. L., Cohen, M. A., Takahashi, A., Keil, B., Kanwisher, N., & Saxe, R. (2022). Selective responses to faces, scenes, and bodies in the ventral visual pathway of infants. *Current Biology*, *32*(2), 265–274.e5. https://doi.org/10.1016/j.cub.2021.10.064

Leslie, A. M. (1995). A theory of agency. In D. Sperber, D. Premack, & A. J. Premack (Eds.), *Causal cognition: A multidisciplinary debate* (pp. 121–149). Clarendon Press/Oxford University Press.

Leslie, A. M., & Keeble, S. (1987). Do six-month-old infants perceive causality? *Cognition*, *25*(3), 265–288. https://doi.org/10.1016/S0010-0277(87)80006-9

Lupyan, G., & Rakison, D. H. (2006). What moves in a mysterious way? A domain-general account of learning about animacy and causality. In R. Sun & N. Miyake (Eds.), *Proceedings of the 28th annual conference of the cognitive science society* (pp. 525–530). Lawrence Erlbaum.

Macchi Cassia, V., Valenza, E., Simion, F., & Leo, I. (2008). Congruency as a nonspecific perceptual property contributing to newborns' face preference. *Child Development*, *79*(4), 807–820. https://doi.org/10.1111/j.1467-8624.2008.01160.x

Malachowski, L. G., & Needham, A. W. (2023). Infants exploring objects: A cascades perspective. *Advances in Child Development and Behavior*, *64*, 39–68. https://doi.org/10.1016/bs.acdb.2022.11.001

Malik, M., & Isik, L. (2023). Relational visual representations underlie human social interaction recognition. *Nature Communications*, *14*(1), Article 7317. https://doi.org/10.1038/s41467-023-43156-8

Mareschal, D., & French, R. (2000). Mechanisms of categorization in infancy. *Infancy*, *1*(1), 59–76. https://doi.org/10.1207/S15327078IN0101_06

Mareschal, D., French, R. M., & Quinn, P. C. (2000). A connectionist account of asymmetric category learning in early infancy. *Developmental Psychology*, *36*(5), 635–645. https://doi.org/10.1037/0012-1649.36.5.635

Mareschal, D., & Shultz, T. R. (1999). Development of children's seriation: A connectionist approach. *Connection Science*, *11*(2), 149–186. https://doi.org/10.1080/095400999116322

Markson, L., & Spelke, E. S. (2006). Infants' rapid learning about self-propelled objects. *Infancy*, *9*(1), 45–71. https://doi.org/10.1207/s15327078in0901_3

Mascalzoni, E., Regolin, L., Vallortigara, G., & Simion, F. (2013). The cradle of causal reasoning: Newborns' preference for physical causality. *Developmental Science*, *16*(3), 327–335. https://doi.org/10.1111/desc.12018

Masten, A. S., & Cicchetti, D. (2010). Developmental cascades. *Development and Psychopathology*, *22*(3), 491–495. https://doi.org/10.1017/S0954579410000222

McClelland, J. L., & Thompson, R. M. (2007). Using domain-general principles to explain children's causal reasoning abilities. *Developmental Science*, *10*(3), 333–356. https://doi.org/10.1111/j.1467-7687.2007.00586.x

McMahon, E., & Isik, L. (2023). Seeing social interactions. *Trends in Cognitive Sciences*, *27*(12), 1165–1179. https://doi.org/10.1016/j.tics.2023.09.001

McMurray, B., Horst, J. S., & Samuelson, L. K. (2012). Word learning as the interaction of online referent selection and slow associative learning. *Psychological Review*, *119*(4), 831–877. https://doi.org/10.1037/a0029872

Muentener, P., & Carey, S. (2010). Infants' causal representations of state change events. *Cognitive Psychology*, *61*(2), 63–86. https://doi.org/10.1016/j.cogpsych.2010.02.001

Needham, A., & Baillargeon, R. (1993). Intuitions about support in 4.5-month-old infants. *Cognition*, *47*(2), 121–148. https://doi.org/10.1016/0010-0277(93)90002-D

Oakes, L. M. (1994). Development of infants' use of continuity cues in their perception of causality. *Developmental Psychology*, *30*(6), 869–879. https://doi.org/10.1037/0012-1649.30.6.869

Oakes, L. M., & Cohen, L. B. (1990). Infant perception of a causal event. *Cognitive Development*, *5*(2), 193–207. https://doi.org/10.1016/0885-2014(90)90026-P

Oakes, L. M., & Rakison, D. H. (2019). *Developmental cascades: Building the infant mind*. Oxford University Press. https://doi.org/10.1093/oso/9780195391893.001.0001

Plaut, D. C., & Vande Velde, A. K. (2017). Statistical learning of parts and wholes: A neural network approach. *Journal of Experimental Psychology: General*, *146*(3), 318–336. https://doi.org/10.1037/xge0000262

Powell, L. J. (2022). Adopted utility calculus: Origins of a concept of social affiliation. *Perspectives on Psychological Science*, *17*(5), 1215–1233. https://doi.org/10.1177/17456916211048487

Powell, L. J., Kosakowski, H. L., & Saxe, R. (2018). Social origins of cortical face areas. *Trends in Cognitive Sciences*, *22*(9), 752–763. https://doi.org/10.1016/j.tics.2018.06.009

Rakison, D. H., & Benton, D. T. (2019). Second-order correlation learning of dynamic stimuli: Evidence from infants and computational modeling. *Infancy*, *24*(1), 57–78. https://doi.org/10.1111/infa.12274

Rakison, D. H., & Krogh, L. (2012). Does causal action facilitate causal perception in infants younger than 6 months of age? *Developmental Science*, *15*(1), 43–53. https://doi.org/10.1111/j.1467-7687.2011.01096.x

Rakison, D. H., & Lupyan, G. (2008). Developing object concepts in infancy: An associative learning perspective. *Monographs of the Society for Research in Child Development*, *73*(1), vii, 1–110. https://doi.org/10.1111/j.1540-5834.2008.00454.x

Rakison, D. H., & Poulin-Dubois, D. (2001). Developmental origin of the animate-inanimate distinction. *Psychological Bulletin*, *127*(2), 209–228. https://doi.org/10.1037/0033-2909.127.2.209

Rakison, D. H., & Poulin-Dubois, D. (2002). You go this way and I'll go that way: Developmental changes in infants' detection of correlations among static and dynamic features in motion events. *Child Development*, *73*(3), 682–699. https://doi.org/10.1111/1467-8624.00432

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black & W. F. Prokasy (Eds.), *Classical conditioning II: Current research and theory* (pp. 64–99). Appleton-Century-Crofts.

Rogers, T. T., & McClelland, J. L. (2004). *Semantic cognition: A parallel distributed processing approach*. MIT Press. https://doi.org/10.7551/mitpress/6161.001.0001

Rose, S. A., Gottfried, A. W., Melloy-Carminar, P., & Bridger, W. H. (1982). Familiarity and novelty preferences in infant recognition memory: Implications for information processing. *Developmental Psychology*, *18*(5), 704–713. https://doi.org/10.1037/0012-1649.18.5.704

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, *323*(6088), 533–536. https://doi.org/10.1038/323533a0

Saxe, R., Tenenbaum, J. B., & Carey, S. (2005). Secret agents: Inferences about hidden causes by 10- and 12-month-old infants. *Psychological Science*, *16*(12), 995–1001. https://doi.org/10.1111/j.1467-9280.2005.01649.x

Saxe, R., Tzelnic, T., & Carey, S. (2007). Knowing who dunnit: Infants identify the causal agent in an unseen causal interaction. *Developmental Psychology*, *43*(1), 149–158. https://doi.org/10.1037/0012-1649.43.1.149

Scarf, D., Imuta, K., Colombo, M., & Hayne, H. (2012). Social evaluation or simple association? Simple associations may explain moral reasoning in infants. *PLOS ONE*, *7*(8), Article e42698. https://doi.org/10.1371/journal.pone.0042698

Scott, L. S., & Arcaro, M. J. (2023). A domain-relevant framework for the development of face processing. *Nature Reviews Psychology*, *2*(3), 183–195. https://doi.org/10.1038/s44159-023-00152-5

Shutts, K., Markson, L., & Spelke, E. S. (2009). The developmental origins of animal and artifact concepts. In B. M. Hood & L. R. Santos (Eds.), *The origins of object knowledge* (pp. 189–210). Oxford Academic.

Simion, F., & Giorgio, E. D. (2015). Face perception and processing in early infancy: Inborn predispositions and developmental changes. *Frontiers in Psychology*, *6*, Article 969. https://doi.org/10.3389/fpsyg.2015.00969

Simion, F., Valenza, E., Macchi Cassia, V., Turati, C., & Umiltà, C. (2002). Newborns' preference for up–down asymmetrical configurations. *Developmental Science*, *5*(4), 427–434. https://doi.org/10.1111/1467-7687.00237

Sirois, S., & Mareschal, D. (2002). Models of habituation in infancy. *Trends in Cognitive Sciences*, *6*(7), 293–298. https://doi.org/10.1016/S1364-6613(02)01926-5

Slater, A. (1989). Visual memory and perception in early infancy. In A. Slater & G. Bremner (Eds.), *Infant development* (pp. 43–71). Lawrence Erlbaum Associates.

Sokolov, E. N. (1960). Neuronal models and the orienting reflex. In M. A. Brazier (Ed.), *The central nervous system and behavior* (pp. 187–276). J. Macy.

Spelke, E. S. (2022). *What babies know: Core knowledge and composition* (Vol. 1). Oxford University Press. https://doi.org/10.1093/oso/9780190618247.001.0001

Spelke, E. S., & Kinzler, K. D. (2007). Core knowledge. *Developmental Science*, *10*(1), 89–96. https://doi.org/10.1111/j.1467-7687.2007.00569.x

Spelke, E. S., & Kinzler, K. D. (2009). Innateness, learning, and rationality. *Child Development Perspectives*, *3*(2), 96–98. https://doi.org/10.1111/j.1750-8606.2009.00085.x

Spelke, E. S., Phillips, A., & Woodward, A. L. (1995). Infants' knowledge of object motion and human action. In D. Sperber, D. Premack, & A. J. Premack (Eds.), *Causal cognition: A multidisciplinary debate* (pp. 44–78). Clarendon Press/Oxford University Press.

Spencer, J. P., Blumberg, M. S., McMurray, B., Robinson, S. R., Samuelson, L. K., & Tomblin, J. B. (2009). Short arms and talking eggs: Why we should no longer abide the nativist-empiricist debate. *Child Development Perspectives*, *3*(2), 79–87. https://doi.org/10.1111/j.1750-8606.2009.00081.x

Stojnić, G., Gandhi, K., Yasuda, S., Lake, B. M., & Dillon, M. R. (2023). Commonsense psychology in human infants and machines. *Cognition*, *235*, Article 105406. https://doi.org/10.1016/j.cognition.2023.105406

Vong, W. K., Wang, W., Orhan, A. E., & Lake, B. M. (2024). Grounded language acquisition through the eyes and ears of a single child. *Science*, *383*(6682), 504–511. https://doi.org/10.1126/science.adi1374

Westermann, G., & Mareschal, D. (2004). From parts to wholes: Mechanisms of development in infant visual object processing. *Infancy*, *5*(2), 131–151. https://doi.org/10.1207/s15327078in0502_2

Woo, B. M., Tan, E., & Hamlin, J. K. (2022). Human morality is based on an early-emerging moral core. *Annual Review of Developmental Psychology*, *4*(1), 41–61. https://doi.org/10.1146/annurev-devpsych-121020-023312

Received September 5, 2023
Revision received July 9, 2024
Accepted July 24, 2024 ■