Evans, V. C., Berthiaume, V. G., & Shultz, T. R. (2010). Toddlers' transitions on non-verbal false-belief tasks involving a novel location: a constructivist connectionist model. Proceedings of the Ninth IEEE International Conference on Development and Learning (pp. 225-230) Ann Arbor, MI: IEEE.

Toddlers' Transitions on Non-verbal False-belief Tasks Involving a Novel Location: A Constructivist Connectionist Model

Vanessa C. Evans, Vincent G. Berthiaume, and Thomas R. Shultz

***Abstract*—Some argue that children learn a Theory of Mind (ToM), the understanding that others have mental states, at around 3.5 years. This is evidenced by their transition from failure to success on verbal false-belief tasks, when they begin to verbally predict an actress will search for a toy where she falsely believes it to be, rather than in its actual location. However, non-verbal measures have recently been used to show that children in their second year of life may already have some understanding of others' false beliefs. We present a Sibling-Descendant Cascade-Correlation neural-network model of one study that found 25-month-old toddlers correctly anticipated an actress would search according to her false belief. Networks were trained on true- and false-belief search patterns, simulating toddlers' everyday experience with true and false beliefs, and then tested on non-verbal true- and false-belief tasks involving a novel location. Networks transitioned from incorrectly predicting true-belief searches in both true- and false-belief tasks to making correct predictions in both tasks. Our model thus (1) reproduced the transition that has been observed in older children and (2) generalized its learning to a novel location. The model can be used to refine our understanding of the transitions while again demonstrating the usefulness of SDCC as an algorithm for modeling cognitive development.**

***Index Terms*—false belief, neural networks, theory of mind.**

I. Introduction

As adults, we understand, explain and predict human behaviour in terms of people's mental states, including motivational states (e.g., goals and desires), and informational states (e.g., beliefs). This understanding that others have mental states is referred to as a Theory of Mind (ToM) [2]. A key competency for navigating social interactions, ToM is one of the hottest topics in developmental psychology [3]. An understanding of its development and its simulation in computational models may benefit the field of developmental robotics, as a ToM could be implemented to allow robots to more naturally interact with and learn from humans [4].

Here we present a connectionist model that reproduces a transition from failure to success on false-belief tasks, which test whether children expect others to act on the basis of mental representations that are outdated and false. Using Sibling-Descendant Cascade-Correlation (SDCC), our model was trained on observations of true- and false-belief search to predict where an actress would search when she had true and false belief, and could eventually generalize its predictions to searches in a novel location.

We first summarize psychological evidence pertaining to ToM development, including the empirical study on which our model is based. We then briefly review a previous model of ToM development and the SDCC algorithm before describing our model and presenting our results.

II. Psychological Evidence

A. Controversies of ToM Development

A transition in ToM development is thought to occur around 3.5 years, as children begin to succeed at verbal false-belief tasks. In the standard task [5], children see Sally place a marble into location A. Sally leaves and Anne moves the marble to location B, causing Sally to falsely believe it is still in A. Sally comes back and children are asked to say where she will search for the marble. Typically, younger children incorrectly say that Sally will search in B, the actual location of the marble [6]. These children may thus not understand that Sally has a false mental representation of the marble's location, but rather expect her to know the actual state of the world; this has been labeled an omniscient ToM. In contrast, older children typically correctly say that Sally will search in A, the false-belief location, thus seeming to understand that she has a mental representation of the marble's location; this has been labeled a representational ToM.

Although this transition on verbal false-belief tasks is well-established, the mechanisms underlying the transition are widely debated. Some posit that failure at verbal tasks reflects a conceptual deficit in understanding others' beliefs; a child's transition to making correct predictions is thus due to his learning of an adult-like interpretation of others' mental states

Manuscript received March 5, 2010. This work was supported in part by a scholarship and a grant from the Natural Sciences and Engineering Research Council of Canada to VGB and TRS, respectively.

Vanessa C. Evans is with the Department of Psychology at McGill University, Montreal, QC Canada (e-mail: vanessa.evans@mail.mcgill.ca).

Vincent G. Berthiaume is with the Department of Psychology, McGill University, Montreal, QC Canada (email: vincent.berthiaume@mail.mcgill.ca).

Thomas R. Shultz is with the Department of Psychology and School of Computer Science, McGill University, Montreal, QC Canada (514-398-6139; e-mail: thomas.shultz@mcgill.ca).

Evans, V. C., Berthiaume, V. G., & Shultz, T. R. (2010). Toddlers' transitions on non-verbal false-belief tasks involving a novel location: a constructivist connectionist model. Proceedings of the Ninth IEEE International Conference on Development and Learning (pp. 225-230) Ann Arbor, MI: IEEE.

[7]–[9]. In contrast, others posit that failure at verbal tasks is due to task-specific performance limitations that obscure an earlier emerging and perhaps domain-specific and innate understanding of others' beliefs [10]–[14].

Recent studies using non-verbal false-belief tasks support this latter view by suggesting that younger infants do have some understanding of false beliefs [15]–[20]. Compared to verbal tasks, non-verbal tasks minimize confounding linguistic and inhibitory demands by relying neither on verbal prompts nor verbal responses [21]. Thus non-verbal tasks tap into an apparent implicit, non-conscious understanding of others' false beliefs, as opposed to an explicit knowledge of those beliefs [22]. We next describe a study employing such non-verbal false-belief tasks, on which the current model is based.

B. *Southgate et al.'s Non-verbal False-belief Tasks*

Southgate and colleagues presented two non-verbal false-belief tasks to 25-month-old toddlers, using their anticipatory looks and their looking times as non-verbal measures [16].

In two familiarization trials, toddlers watched an actress behind a panel observe a puppet place a ball in either of two boxes situated in front of the panel. A tone then sounded and two doors, one above each box, lit up. These cues were presented to teach toddlers that the actress would then open the door above the location of the ball to reach for it, and were used to mark the beginning of the anticipatory look period.

Half the toddlers next watched either the false-belief 1 or false-belief 2 task. In the false-belief 1 task, the actress watched the puppet put the ball first in the left box, then move it to the right one. Then a telephone ring prompted the actress to look away, whereupon the puppet moved the ball from the right box to under the stage. The false-belief 2 task was identical, except that the phone ring occurred earlier causing the actress to only see the ball being put in the first, left box. Thus, in task 1, the actress believed the ball to be in the second box, whereas in task 2, she believed it to be in the first box. In both tasks, both locations were actually empty.

After these task-specific sequences, the actress turned back to the stage. The tone and lights indicated she would reach for the ball, and the direction of the infants' first saccades was then measured with an eye-tracking device. Over both tasks, 17 out of 20 toddlers looked to the correct location first, and for nearly twice as long as the incorrect location.

Some researchers take this and other studies, e.g., [17], [19], as support that younger children understand false beliefs, whereas others prefer more conservative interpretations. Perhaps the most compelling alternative interpretation is that infants and toddlers use simple, learned behavioral rules, e.g., "people search for objects where they last saw them" to solve the tasks, without considering her false beliefs [23], [24].

III. Computational Evidence

Several models of false-belief tasks have been developed (see [25] for a review), though one in particular has succeeded at capturing the false-belief transition [25]. This model employed SDCC [26, described later] to train neural networks on a non-verbal false-belief task in which a toy was moved from one location to another with an actress either watching or absent [15]. When networks were trained with more true- than false-belief search patterns, in accordance with the assumption that children have more experience with agents acting on true rather than false beliefs [27], they transitioned from making omniscient to representational ToM predictions.

IV. Our model

We built a constructivist connectionist model using the SDCC algorithm to examine how networks would handle the Southgate *et al.* tasks. These tasks were attractive to model because they were more challenging than previously modeled tasks [25], due to their inclusion of three consecutive object placements and an unexpected offstage location. Further, the unexpected location allowed us to evaluate networks' abilities to generalize their training to a novel location. An ability to generalize past learning to novel situations is an important competency observed in children and is often used to evaluate competing models of psychological phenomena [28]. Finally, would the model reproduce the omniscient-to-representational transition on these more challenging non-verbal false-belief tasks?

A. *The SDCC Learning Algorithm*

The SDCC algorithm is a feed-forward constructive algorithm that adjusts connection weights to reduce the discrepancy, or error, between a target output activation and the network's own output activation. If error reduction stagnates, networks recruit hidden units, which provide additional computational power. A hidden unit is recruited from a pool of candidate units whose incoming weights are trained to track (correlate with) error. Once installed in the network, the chosen hidden unit's incoming weights are frozen and its outgoing connection weights are trained to reduce error. After installing the first hidden unit on a new layer between the initial input and output layers, subsequent hidden units can be installed either as siblings on the current, highest layer or as descendants on a new layer. In this way, SDCC searches in both weight and topography space to master a task. Weight adjustment corresponds to learning (via synaptic potentiation), while hidden-unit recruitment corresponds to development (via either synaptogenesis or neurogenesis). Further mathematical and computational details of SDCC are available from several other sources, along with simulations of many phenomena in cognitive development [29]–[31].

B. *Input and Output*

Figure 1 depicts an initial network without any hidden units. Networks received as input the information an observer would potentially need to predict an actress's searching behavior in the false-belief tasks. This included the start, mid, and end location of the ball, and whether the actress was watching each time the ball was moved.

We coded each of the ball's three placements with four input units. The location of the ball was encoded by activating

Evans, V. C., Berthiaume, V. G., & Shultz, T. R. (2010). Toddlers' transitions on non-verbal false-belief tasks involving a novel location: a constructivist connectionist model. Proceedings of the Ninth IEEE International Conference on Development and Learning (pp. 225-230) Ann Arbor, MI: IEEE.

one of these four units (to 0.5) and activating the other three units (to -0.5). There were three sets of these four units representing the first, second, and final placements the ball in a given search pattern. For each placement, a single unit coded whether the actress saw the ball being placed into a location (activated to 0.5) or did not see (activated to -0.5). Four output units indicated where a network predicted the actress to search, in the same way four input units coded object placement.

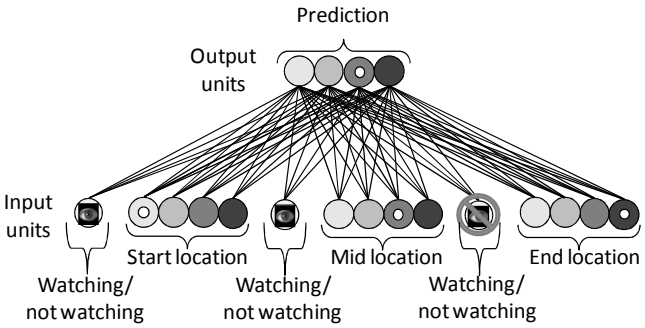

Fig. 1. A schematic diagram of an initial network. The input units encode a false-belief 1 search pattern in which an actress watches as the ball is placed in location 1 and then moved to location 3, then does not watch as the ball is moved to location 4. The correct prediction is a search in location 3.

C. Training

Whereas Southgate *et al.*'s two false-belief tasks only involved testing and not training, our neural networks first underwent a period of training before being tested. Network training can be considered to simulate the experience that toddlers have with observing others' true and false beliefs search behavior before being tested in an experiment.

Because children may observe instances of behavior mostly based on true beliefs [15], network training included twice as many true-belief as false-belief search patterns [27]. Training also included four possible object locations, as children are likely to become accustomed to objects being moved to and from more than two locations in everyday life [25].

Networks were trained on 64 different belief situations for each of the 3 tasks (false-belief 1, false-belief 2, and true-belief), obtained by combining the 4 possible start locations, 4 possible mid locations, and 4 possible end locations. To create our training set, we used different repetitions of each of these situations patterns to yield the desired 2:1 ratio of true- to false-belief search patterns. That is, networks were trained on 336 false-belief 1, 336 false-belief 2, and 1344 true-belief search patterns, giving a ratio of 1344:672 true- to false-belief search patterns and a total of 2016 search patterns in the training set. These search patterns were randomly selected from large, independent stocks of false-belief 1, 2, and true-belief search patterns independently for each network. Selecting these patterns at random from stocks ensured that each network would be trained on slightly different patterns and proportions of the unique true- and false-belief patterns, just as children gain somewhat different experiences with true and false beliefs.

For networks to qualify as having learned a given training pattern, output units need to produce an output activation that is within the score-threshold parameter (kept here at the default value of 0.4) of the target activation. Training ended once networks reach this criterion on all training patterns.

D. Testing

In testing, connection weights were frozen to prevent learning. Test patterns replicated the Southgate *et al*. tasks, and thus included only two of the original four search locations, plus a novel location mimicking the unexpected offstage location. The object was placed in location 1 (left box), then location 2 (right box), and then in the novel location (offstage). At each time step, locations 1 and 2 were coded as in training, whereas the novel location was coded by negatively activating all four units (to -0.5), indicating that the object was in neither of the expected search locations.

In addition to the false-belief 1 and 2 tasks, we included a true-belief task, in which the actress saw all the ball's placements, including its removal offstage. We included this true-belief task for two reasons. First, incorporating this task was necessary to test for an omniscient-to-representational transition, as omniscient ToM always predicts searches in the object's actual location, including in novel, offstage locations. Second, testing on a true-belief task allowed us to assess how proficient the networks became at predicting a search in the novel offstage location, given that correct false-belief task predictions were never in the actual object location.

We measured a network's search predictions by calculating the discrepancy, or error, between the four output units' activation values and the target activation values for searches in location 1, location 2, and the offstage location. Large error for a given location reflects strong "surprise" at a search in that location, whereas low error reflects a strong "expectation" of a search in that location. Error can thus be understood as analogous to the discrepancy between what a toddler expects an agent to do (in this case, to search in a certain location) and what the infant observes the agent actually doing. Although the error measure differs somewhat from Southgate *et al.*'s anticipatory looking and looking time measures, we use it here because it is widely used to evaluate network performance [25], [30], and because it can be easily related to Southgate *et al.*'s measures. That is, the toddlers were presumed to look first and longer at the location where they predicted a search; by comparison, models are expected to output lower error for the locations where they "predict" a search. Thus, networks making omniscient ToM predictions would have lower error for searches in the offstage location than for searches in locations 1 and 2 in both the true- and false-belief tasks, whereas networks making representational ToM predictions would have lower error for searches in the offstage location only in the true-belief task and lower error for searches in locations 2 and 1 for the false-belief 1 and 2 tasks, respectively.

To determine the progress the model made over training and to test for an omniscient-to-representational transition,

Evans, V. C., Berthiaume, V. G., & Shultz, T. R. (2010). Toddlers' transitions on non-verbal false-belief tasks involving a novel location: a constructivist connectionist model. Proceedings of the Ninth IEEE International Conference on Development and Learning (pp. 225-230) Ann Arbor, MI: IEEE.

networks were tested twice: once early in training, i.e., just before they recruited their first hidden unit, and once when training was complete. If networks transitioned from omniscient to representational ToM predictions, we predicted an interaction between time of test and search location in the false-belief tasks (combined for analysis). For false-belief tasks, error would be lower early in training for search in the object's actual location, whereas at the end of training, it would be lower for search in the false-belief locations. By contrast in the true-belief task we predicted a main effect of search location as error would be lower for searches in the object's actual location both early and at the end of training.

Forty networks were trained, half of which were tested on the true-belief task and half on one of the two false-belief tasks, matching Southgate *et al.*'s factorial design.

V. Results

We first provide descriptive information on training and network topologies, then report performance on the true- and false-belief tasks, which were evaluated separately.

A. Training and Network Topologies

Training took between 471 and 941 epochs to complete, $M = 710.88$, $SD = 99.42$. Networks recruited between four and eight hidden units, $M = 5.70$, $SD = 0.72$. The modal network structure, occurring in 19 of 40 networks, was six hidden units, all siblings on a single layer. The second most common structure, occurring in another 14 of 40 networks, was five hidden units, also all siblings. Other networks mostly also had varying numbers of siblings all on one layer; only five networks recruited descendant units on a second layer.

B. True-belief Task

For the true-belief task, we performed a 2-way repeated-measures Analysis of Variance (ANOVA) on network error, with the factors time-of-test (early in training and end of training) and search-location (actual, false-belief, and empty location). Where Mauchly's test indicated that the assumption of sphericity was violated, degrees of freedom were adjusted using the Greenhouse-Geisser correction. The main effects of search-location and time-of-test, and their interaction, were all significant (Figure 2).

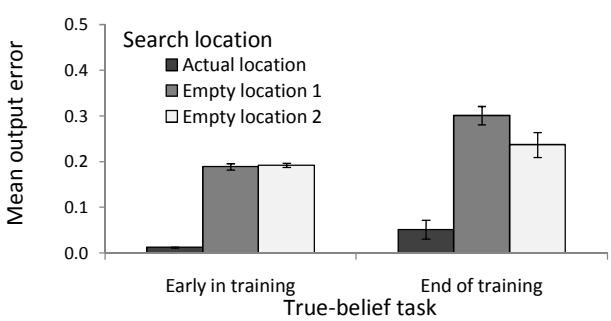

Fig. 2. Mean network error and SE bars on the true-belief task early in training and at the end of training. Networks were proficient at the true-belief task both early and at the end of training.

The main effect of search-location, $F(1.10, 20.89) = 185.89$, $p < .001$, was explored with Bonferroni-adjusted planned comparisons. Error was lower for searches in the actual than in the empty locations 1 and 2, ps $< .01$. Error was also lower for searches in empty location 1 than empty location 2, $p < .01$.

The main effect of time-of-test, $F(1, 19) = 13.166$, $p < .005$, reflected lower overall network error early than at the end of training. Indeed, network error increased for all search locations during training. This was perhaps due to the fact that by the end of training, networks had learned to accommodate a larger variety of predictions and because they had stronger expectations about where the actress would search, thus producing higher error for incorrect searches.

Finally, the time-of-test and search-location interaction, $F(1.066, 20.25) = 4.69$, $p < .05$, was explored by testing the simple main effects of search-location within each level of time-of-test using 1-way repeated-measures ANOVAs. However, both early and at the end of training, the simple main effect of search-location was significant, ps $< .01$, and error was lower for searches in the actual than in empty locations, ps $< .01$, but did not differ between the two empty locations, $p > .12$. Thus networks succeeded at the true-belief task both early and at the end of training.

C. False-belief Tasks

Before combining data for the two false-belief tasks, we performed a 3-way mixed ANOVA with the between-subjects factor Task (false-belief 1, false-belief 2) and within-subjects factors time-of-test and search-location, and found no significant main or interaction effect. As performance on false-belief tasks did not differ, following Southgate *et al.*, we combined data for the two false-belief tasks, and performed a 2-way repeated-measures ANOVA of network error with the factors time-of-test (early in training and end of training) and search-location (actual, false-belief, and empty location). The main effects of time-of-test and search-location, as well as their interaction, were all significant (Figure 3).

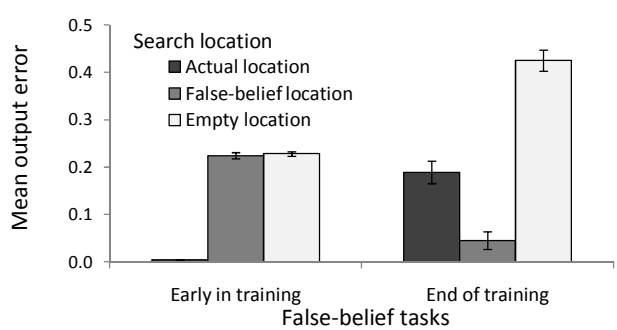

Fig. 3. Mean network error and SE bars on the combined false-belief task early in training and at the end of training. Networks' predictions indicate an omniscient-to-representational ToM transition, as they made incorrect, omniscient predictions early, but correct predictions by the end of training.

The main effect of search-location, $F(1.37, 26.09) = 99.84$, $p < .001$ was explored with Bonferroni-adjusted planned comparisons. Error was overall higher for searches in the empty location than for searches in either the actual or false-

Evans, V. C., Berthiaume, V. G., & Shultz, T. R. (2010). Toddlers' transitions on non-verbal false-belief tasks involving a novel location: a constructivist connectionist model. Proceedings of the Ninth IEEE International Conference on Development and Learning (pp. 225-230) Ann Arbor, MI: IEEE.

belief locations, *p*s < .001, while it did not differ overall between searches in the actual and false-belief locations, $p = .21$. Error was higher for the empty location because networks never expected a search there.

The main effect of time-of-test, $F(1, 19) = 49.58$, $p < .001$, reflected lower network error early that at the end of training. As for the true-belief task, this was due to error increasing for incorrect, empty location searches by the end of training, due to stronger search predictions in the correct locations.

We explored the time-of-test and search-location interaction, $F(1.44, 27.37) = 77.00$, $p < .001$, by examining the simple main effects of search-location within each level of time-of-test using 1-way repeated-measures ANOVAs. Early in training, the simple main effect of search-location was significant, $F(1.36, 27.37) = 60.63$, $p < .01$. Planned comparisons indicated that error was lower for searches in the actual location than in both the false-belief and empty locations, *p*s < .001, while error did not differ between searches in the false-belief and empty locations, $p = 1.0$ (Figure 3). Thus early in training, networks were making predictions consistent with an omniscient ToM.

At the end of training, the simple main effect of search-location was also significant, $F(1.05, 27.37) = 130.63$, $p < .01$. Planned comparisons indicated that error was lower for searches in the false-belief location than for searches in the actual and empty locations, *p*s < .005, and was lower for searches in the actual location than the empty location, $p < .001$. Thus by the end of training, networks were making predictions consistent with a representational ToM.

VI. Discussion

A. Omniscient-to-Representational ToM Transition

Our model transitioned from making predictions consistent with an omniscient ToM to making predictions consistent with a representational ToM. Early in training, networks predicted the actress would search in the object's actual location, regardless of whether she had a true or false belief about its location. However by the end of training, networks predicted the actress would search according to both her true and false beliefs—in the ball's actual location in the true-belief task, and in the empty location where she falsely believed the ball to be in the false-belief tasks.

Our results concur with those of a previous model [25], where SDCC networks made an omniscient to representational transition on a simpler non-verbal false-belief task. Our current model suggests that such a transition on non-verbal tasks could be found in children younger than 25 months, the youngest documented age of success on these particular tasks.

As in other SDCC simulations, our networks made this transition through a combination of learning (synaptic weight adjustment) and development (recruiting new units and synapses in order to cope with non-linearity in the training patterns). Before being tested on false-belief tasks, networks were trained on patterns simulating realistic observations of actors searching for objects that change locations. For instance, because most of our everyday beliefs are true [15], networks got trained on twice as many true- as false-belief situations, resulting in success first on the true-belief task, and later on false-belief tasks. The transitions in children may thus be due to similar mechanisms and experiences, e.g., they may understand true beliefs before false beliefs. Researchers have no direct access to psychological mechanisms, and thus resort to covering documented psychological phenomena and predicting new ones, as we do here in our computational modeling. Further modeling and experimentation may establish whether this or newer models will best characterize false-belief task transitions.

Coverage of behavioral phenomena and empirical confirmation of model predictions can be supplemented by analyses of knowledge representations in networks at various points in development. Such analyses in a similar, previous SDCC model indicated that networks effected the omniscient-to-representational transition by distinguishing between true- and false beliefs [25]. It is interesting that computational models trained on patterns of search can develop knowledge representations yielding distinctions between true and false beliefs. Such distinctions may prove to be fundamental in developing a full blown ToM. Computational modeling could play an important role in filling in the gaps in such a theory.

Evidence gleaned from computational modeling is particularly relevant now that a mounting body of evidence suggests that children in their second year of life are already attributing mental states, including false-beliefs, to others [20]. Indeed, recent empirical results suggest that toddlers much younger than 3.5 years understand that others can hold false beliefs about an object's location [15]–[17] and identity [19], arrive at false conclusions about an object based on misleading perceptual information [18], and even actively help in an appropriate manner given an adult's false belief [32]. This evidence lends support to the view that some aspects of ToM may be in place much earlier than previously believed.

B. Networks' Generalization

In succeeding at the Southgate *et al.* tasks, networks not only made correct predictions of search behavior, but did so even though a novel location, with which networks had no experience in training, was involved. In false-belief tasks, networks had to correctly interpret search patterns with this novel location to predict searches in other familiar locations. In the true-belief task, networks had to correctly interpret this input and also predicted a search in the novel location.

Because networks' connection weights were adjusted to perform appropriate transformations on their specific training input, networks that make correct predictions in test when given novel input must have developed representations that work on principles that are partially independent of the exact input. That is, networks developed a robust pattern of connection weights that allowed them to generalize their training to test patterns involving a novel location and even compute this novel location. Generalization is a critical ability by which computational models are evaluated and compared

Evans, V. C., Berthiaume, V. G., & Shultz, T. R. (2010). Toddlers' transitions on non-verbal false-belief tasks involving a novel location: a constructivist connectionist model. Proceedings of the Ninth IEEE International Conference on Development and Learning (pp. 225-230) Ann Arbor, MI: IEEE.

[28]; this model is the first to demonstrate generalization of this sort and demonstrates yet another advantage that SDCC offers modelers of developmental psychological studies [30].

C. Future Work

Three prospective additions may improve and extend the model. First, incorporating erroneous search patterns in training would render training more realistic, because infants and toddlers periodically observe people committing errors and failing to act according to their beliefs, due to fatigue, distraction, etc. [25]. Second, holding a proportion of belief situations out of training and using them in a generalization test set to see whether networks may succeed at these situations as well may further show the generalization capabilities of our model. Third, we could test networks on an "avoidance" version of this task, in which the actress wishes to avoid a noxious object rather than approach a desirable one. Indeed, children have been observed to undergo a second transition on verbal false-belief tasks, from succeeding at approach tasks but failing on avoidance tasks, to succeeding on both approach and avoidance tasks [27]. As of yet, no empirical studies of non-verbal avoidance tasks have been conducted, and this thus remains an open area for both empirical and model-based investigation. This and further investigation will be necessary if the combined insights of empirical research and computational models are to expand our understanding of ToM development.

Ultimately, this work should be useful for designing robots that develop and learn as human infants do. Our findings suggest that such developmental robots should be equipped with the learning ability characteristic of constructive neural networks and have experience with how other agents search for missing objects. This combination of learning ability and experience should allow them to develop human-like ToM abilities, enabling more realistic social interactions.

REFERENCES

[1] G. Gergely and G. Csibra, "Teleological reasoning in infancy: The naive theory of rational action," *Trends in Cognitive Sciences,* vol. 7, pp. 287-292, Jul. 2003.

[2] D. Premack and G. Woodruff, "Does the chimpanzee have a theory of mind," *Behavioral & Brain Sciences,* vol. 1, pp. 515-526, 1978.

[3] P. L. Harris, "Social cognition," in *Cognition, perception, and language*, 6th ed, D. Kuhn and R. S. Siegler, Eds. Hoboken, NJ: Wiley, 2006, pp. 811-858.

[4] B. Scassellati, "Theory of Mind for a Humanoid Robot," *Autonomous Robots,* vol. 12, pp. 13-24, 2002.

[5] S. Baron-Cohen, A. M. Leslie, and U. Frith, "Does the autistic child have a 'theory of mind'?," *Cognition,* vol. 21, pp. 37-46, 1985.

[6] H. M. Wellman, D. Cross, and J. Watson, "Meta-analysis of theory-of-mind development: The truth about false belief," *Child Development,* vol. 72, pp. 655-684, 2001.

[7] J. Perner, *Understanding the representational mind*. Cambridge, Mass: MIT Press 1991.

[8] A. Gopnik, "How we know our minds: The illusion of first-person knowledge about intentionality," *Readings in Philosophy and Cognitive Science,* pp. 315-346, 1993.

[9] H. M. Wellman and J. D. Woolley, "From simple desires to ordinary beliefs: The early development of everyday psychology," *Cognition,* vol. 35, pp. 245-275, Jun. 1990.

[10] S. Baron-Cohen, *Mindblindness: An essay on autism and theory of mind*. Cambridge, Mass.: MIT Press, 1995.

[11] A. M. Leslie, "Pretense and representation: The origins of 'theory of mind.'," *Psychological Review,* vol. 94, pp. 412-426, 1987.

[12] A. M. Leslie and L. Thaiss, "Domain specificity in conceptual development: Neuropsychological evidence from autism," *Cognition,* vol. 43, pp. 225-251, 1992.

[13] A. M. Leslie, "ToMM, ToBy, and agency: Core architecture and domain specificity," in *Mapping the mind: Domain specificity in cognition and culture.*, L. A. Hirschfeld and S. A. Gelman, Eds. New York: Cambridge University Press, 1994, pp. 119–148.

[14] A. M. Leslie, "Developmental parallels in understanding minds and bodies," *Trends in cognitive sciences,* vol. 9, pp. 459-462, 2005.

[15] K. H. Onishi and R. Baillargeon, "Do 15-month-old infants understand false beliefs?," *Science,* vol. 308, pp. 255-258, 2005.

[16] V. Southgate, A. Senju, and G. Csibra, "Action anticipation through attribution of false belief by 2-year-olds," *Psychological Science,* vol. 18, pp. 587-592, 2007.

[17] H. Song, K. H. Onishi, R. Baillargeon, and C. Fisher, "Can an agent's false belief be corrected by an appropriate communication? Psychological reasoning in 18-month-old infants," *Cognition,* vol. 109, pp. 295-315, 2008.

[18] H. Song and R. Baillargeon, "Infants' reasoning about others' false perceptions," *Developmental Psychology,* vol. 44, pp. 1789-1795, Nov. 2008.

[19] R. M. Scott and R. Baillargeon, "Which penguin is this? Attributing false beliefs about object identity at 18 months," *Child Development,* vol. 80, pp. 1172-1196, 2009.

[20] R. Baillargeon, R. M. Scott, and Z. He, "False-belief understanding in infants," *Trends in Cognitive Sciences,* vol. 14, pp. 110-118, 2010.

[21] P. Bloom and T. P. German, "Two reasons to abandon the false belief task as a test of theory of mind," *Cognition,* vol. 77, pp. B25-B31, Oct. 2000.

[22] T. Ruffman, W. Garnham, A. Import, and D. Connolly, "Does eye gaze indicate implicit knowledge of false belief? Charting transitions in knowledge," *Journal of Experimental Child Psychology,* vol. 80, pp. 201-224, Nov. 2001.

[23] J. Perner and T. Ruffman, "Infants' insight into the mind: How deep?," *Science,* vol. 308, pp. 214-216, Apr. 2005.

[24] T. Ruffman and J. Perner, "Do infants really understand false belief? Response to Leslie," *Trends in Cognitive Sciences,* vol. 9, pp. 462-463, 2005.

[25] V. G. Berthiaume, K. H. Onishi, and T. R. Shultz, "A computational developmental model of the implicit false belief task," in *Proceedings of the 30th Annual Conference of the Cognitive Science Society*, B. C. Love, K. McRae, and V. M. Sloutsky, Eds. Washington, D.C.: Cognitive Science Society, 2008, pp. 825-830.

[26] S. Baluja and S. E. Fahlman, "Reducing network depth in the cascade-correlation," Carnegie Mellon University, Pittsburgh, Technical report CMU-CS-94-209, 1994.

[27] A. M. Leslie, T. P. German, and P. Polizzi, "Belief-desire reasoning as a process of selection," *Cognitive Psychology,* vol. 50, pp. 45-85, 2005.

[28] F. Dandurand, V. G. Berthiaume, and T. R. Shultz, "A systematic comparison of flat and standard cascade-correlation using a student-teacher network approximation task," *Connection Science,* vol. 19, pp. 223-244, 2007.

[29] Fahlman, S.E. and C. Lebiere, *The cascade-correlation learning architecture*, in *Advances in neural information processing systems 2* D.S. Touretzky, Ed. 1990, Morgan Kaufmann: Los Altos, CA. p. 524-532.

[30] Shultz, T.R., *Computational developmental psychology*. 2003, Cambridge, MA: MIT Press.

[31] Shultz, T.R. and S.E. Fahlman, *Cascade-Correlation*, in *Encyclopedia of Machine Learning*, C. Sammut and G. I. Webb, Eds. To be published, Springer-Verlag: Heidelberg, Germany.

[32] D. Buttelmann, M. Carpenter, and M. Tomasello, "Eighteen-month-old infants show false belief understanding in an active helping paradigm," *Cognition,* vol. 112, pp. 337-342, Aug. 2009.