Neural Network Modeling of Pure Reasoning in Preverbal Infants

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 45(45)

Authors
Nobandegani, Ardavan S.
Shultz, Thomas

Publication Date
2023

Peer reviewed

Neural Network Modeling of Pure Reasoning in Preverbal Infants

Ardavan S. Nobandegani[1,3] **& Thomas R. Shultz**[2,3]
{ardavan.salehinobandegani, thomas.shultz}@mcgill.ca
[1]Mila – Quebec AI Institute
[2]School of Computer Science, McGill University
[3]Department of Psychology, McGill University

Abstract

Recent empirical work has provided evidence for *pure reasoning* in infancy, a capacity permitting flexible integration of multiple sources of information to form rational expectations about novel events (Teglas et al., 2011). However, the neural underpinnings of this capacity have remained elusive. In this work, we present the first ecologically rational, neural-level account of these findings on pure reasoning in human infants. Our work bridges two dominant approaches in computational developmental psychology, i.e. neural-network models and Bayesian modeling, substantiating the view that intuitive physics in infancy might, at least partly, involve heuristics: a set of simple, fast, resource-efficient, approximation algorithms that yield sufficiently good results.

Keywords: neural networks; pure reasoning; ecological rationality; heuristics; infants

Introduction

Building on extensive research on infant reasoning abilities, recent experimental work (Teglas et al., 2011) has provided evidence for *pure reasoning* in preverbal infants: the ability to flexibly combine multiple sources of information to form rational expectations about events that the infant has never directly experienced. This finding was used to support the view that preverbal infants are equipped with a rich, abstract system of physics knowledge capable of near-perfectly simulating object movements (the Bayesian ideal observer model) (Teglas et al., 2011). That interpretation is a simplified version of the *intuitive physics engine* hypothesis which maintains that humans possess a rich, symbolic, intricate knowledge of physics, a system of knowledge analogous to computer game engines used for simulating realistic scenarios (Battaglia, Hamrick, & Tenenbaum, 2013; Ullman, Spelke, Battaglia, & Tenenbaum, 2017). Here, we consider an alternative explanation based on heuristics: a set of simple, fast, resource-efficient, approximation algorithms that yield sufficiently good results (Gigerenzer & Todd, 2000).

In this work, we present the first neural-level account of the empirical findings supporting pure reasoning in preverbal infants (Teglas et al., 2011). Our model focuses on learning simple, high-level physical principles by young infants, allowing them to effectively reason about the physics of their environment. This is consistent with a substantial body of work maintaining that cognition is predominantly driven by heuristics — a set of simple, fast, resource-efficient, approximation algorithms that often work quite well in practice (Gigerenzer & Gaissmaier, 2011; Gigerenzer & Selten, 2002; Gigerenzer & Todd, 2000).

Additionally, we showcase how simple heuristics could be mechanistically implemented using Neural Probability Learner and Sampler (NPLS), a neural network model that has provided a unified account of human adults', human infants', and chimpanzees' learning and use of probabilities across a range of probabilistic tasks (Shultz & Nobandegani, 2022a, 2022b).

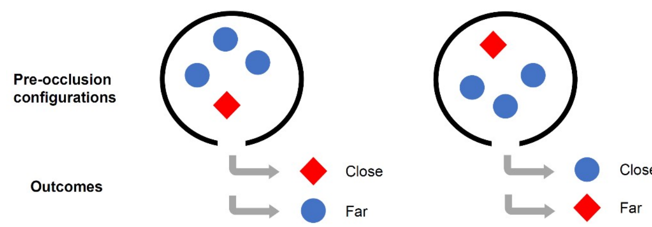

Figure 1: Schematic illustration of the stimuli used in (Teglas et al., 2011). After an occluder hid the balls for 0, 1, or 2 s, a ball was seen to fall out of a container. The exiting ball was either a majority-colored ball or a minority-colored ball, and either close to, or far from, the exit when last visible. Figure adapted from (Teglas et al., 2011).

In the experiment on pure reasoning in preverbal infancy (Teglas et al., 2011), 12-month-old infants saw movies of 3 blue balls and 1 red ball bouncing randomly inside a container, which had an opening at the bottom (Figure 1). After an occluding screen hid the balls for 0, 1, or 2 s, a ball was seen to fall out of the container. The exiting ball was either blue (more numerous) or red (less numerous), and either close to, or far from, the exit when last visible. Infants' surprise at the outcome was assessed by recording their looking time after the resulting exit. How the infants integrated number and closeness was conditional on the amount of time the balls were hidden from view. With a near 0 s occlusion, infants used only distance information to predict the color of the exiting ball. With a 2 s occlusion, they used only number information. And with a 1 s occlusion, they used a combination of distance and number. There was no significant interaction between number and distance in any of the three occlusion-time conditions, suggesting that infants considered the two cues additively. Infants were more surprised, and thus looked

In M. Goldwater, F. K. Anggoro, B. K. Hayes, & D. C. Ong (Eds.), *Proceedings of the 45th Annual Conference of the Cognitive Science Society*.

longer, at an unexpected than at an expected outcome.

Methods

Neural Probability Learner and Sampler (NPLS)

Our proposed neural-network model has two modules, each implemented using the NPLS system (Shultz & Nobandegani, 2022a). NPLS has two major components: an enhanced version of sibling-descendant cascade-correlation (SDCC) for learning probabilities, and a version of Markov-chain Monte Carlo (MCMC) to use the network connection weights in reverse to make inferences about events (Nobandegani & Shultz, 2017).

SDCC is a deterministic, discriminative, neural-network algorithm that constructs networks in a relatively autonomous manner (Baluja & Fahlman, 1994). As in its predecessor, cascade-correlation (CC), SDCC training starts with a two-layer network comprising only input and output layers, and gradually constructs a deeper network by recruiting single hidden units individually, as needed, to reduce network sum-of-squared error. Each recruited hidden unit has been trained to correlate its activation with current network error so that it can readily use that unit's error-detection ability to eventually reduce that error. Unlike classical cascade-correlation (CC), which installs each hidden unit on its own unique layer, SDCC installs the best candidate on the current highest hidden layer (as a sibling), or on the next higher layer (as a descendant), depending on which candidate correlates best with current network error. Sibling units do not receive any input from each other. SDCC thus constructs a variety of network topologies, dependent on the training set and the network's individual construction history. Like its predecessor CC, SDCC uses symmetric sigmoidal activation functions on its hidden units (with range -0.5 to +0.5). For probability learning, NPLS uses an asymmetric sigmoidal activation function for output units (with range 0 to 1).

SDCC learning and development alternates between two phases: output phase and input phase. In output phase, SDCC adjusts output connection weights to reduce network error. When error reduction stagnates, the algorithm changes to input phase to recruit a new hidden unit, adjusting weights entering candidate units to increase the covariance between their activations and network error. In each of these two phases, stagnation is detected when there is no progress greater than a threshold parameter for the number of training epochs specified by a patience parameter. There is also an outer loop with its own threshold and patience parameters to monitor progress over learning cycles, where each such cycle comprises an input phase and the next output phase (Shultz & Doty, 2014). With this ability, NPLS can learn any unnormalized multivariate probability distribution from examples that specify whether or not an output occurs in the presence of a particular input (Kharratzadeh & Shultz, 2016). Importantly, in NPLS, the relevant probabilities are not supplied as learning targets, but rather emerge naturally from neural-network learning of event sequences.

CC and SDCC have been used to simulate many empirical phenomena in cognitive development (Shultz, 2003, 2012; Shultz & Fahlman, 2010). They both recruit as many hidden units as needed to solve the problem being learned, capturing both development (via unit recruitment) and learning (via weight adjustment), thus showing how learning and development can work together. Their coverage of developmental phenomena is typically better than that achieved by symbolic rules or static neural-networks (Shultz, 2013, 2017). In a constructive fashion, they start small and build new knowledge on top of existing knowledge. These constructive networks simulate qualitatively distinct stages because of naturally focusing on the largest current source of error with their current computational power before having to extend that power. To permit learning of probabilistic outcomes, SDCC within NPLS stops when learning progress stagnates. It turns out that this is also the point at which output activations match the probabilities being learned (Shultz & Nobandegani, 2022a).

A useful parameter in NPLS is score-threshold (ST). Technically, ST is the maximum distance from target training values (in this case 0 or 1) considered to be correct. The default value of ST in NPLS is .4, providing a region of uncertainty around the .5 midpoint of the asymmetric sigmoid activation function. NPLS networks are run in learning-cessation mode to ensure that they quit learning when no further progress is being made in error reduction. Here we set ST to .52 to introduce some realistic additional variation into the learning process. With the default .4 ST, NPLS processing is typically much more precise than human processing.

The second major component in NPLS is the use of an MCMC sampling algorithm to simulate how infants could use the learned connection weights in reverse to generate example predictions of which balls would exit the container at each of the three time periods. Effectively, this converts a deterministic neural network into a probabilistic generative model (Nobandegani & Shultz, 2017).

Simulations

We simulate the infant results reported by Teglas et al. (2011) with a neural-network model comprising two modules (Figure 2). These modules are implemented by a pair of NPLS networks that each receive binary-coded versions of the same input used in the infant experiments, i.e., information on number and distance. See Shultz & Nobandegani (2022a) for details on binary-coding of inputs in NPLS.

The number network implements an instantiation of a simple *numerosity heuristic*, operationalizing a high-level principle of physics according to which a majority-colored item is more likely to randomly exit a container than a minority-colored item, provided that the content of the container is sufficiently shuffled over time. The distance network implements an instantiation of *distance heuristic*, operationalizing another high-level principle of physics according to which an item closest to an exit is more likely to randomly exit than items that are farther away.

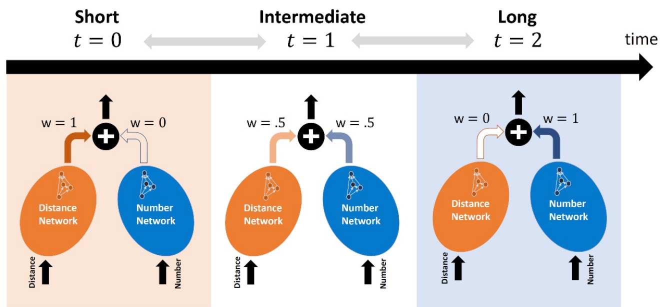

Figure 2: Schematic illustration of the neural-network model accounting for the empirical findings of (Teglas et al., 2011). The *number network* (blue oval) and the *distance network* (orange oval) are implemented as specialized networks, each implementing a simple heuristic. For the two extreme experimental conditions ($t = 0, 2$), one specialized network is fully operative (full weight, $w = 1$), while the other is suppressed (null weight, $w = 0$). For the intermediate condition ($t = 1$), the model takes the average of the recommendations made by the specialized number and distance networks.

Prior to simulating the reasoning experiment, the number network (implementing an instantiation of the numerosity heuristic) is trained to predict that a majority-colored ball is more likely to be randomly selected than a minority-colored ball. With a frequency ratio of 3:1, those probabilities are .75 vs .25. Similarly, the distance network (implementing an instantiation of the distance heuristic) is trained to predict that a ball closest to an exit is more likely to randomly exit than balls that are farther away; using again a 3:1 distribution, those probabilities are .75 vs .25. These target probabilities are never provided to the networks but instead emerge naturally from neural-network learning of event sequences. We assume that infants have such visual experiences independent of any participation in a psychology experiment that pits distance and number cues against each other. The choice of the 3:1 distribution for the distance network allows for *minimal assumptions* about the effect of closeness, by having identical distributions for effects of numerosity and closeness, hence liberating the simulated infants from having to make currently unsupported assumptions about whether the distance network should be trained on a different distribution, and if so, what that distribution should be. Importantly, as long as the distance network captures the simple, intuitive understanding that a closer ball has a higher chance of exiting the container, all our simulation findings qualitatively hold, highlighting the robustness of our results.

The rationality of the network implementing the distance heuristic follows from the intuitive principle that physical objects barely move during an extremely short time interval. The rationality of the network implementing the numerosity heuristic rests on the high-level principle that drawing from a collection of objects shuffled many times is mathematically equivalent to random sampling. Therefore, the use of the number network is well-justified for relatively long occlusion periods, while the use of the distance network is well-justified for extremely short occlusion periods.

In simulating the infant experiment (Teglas et al., 2011), we assume that the distance network is invoked in the shortest occlusion condition (i.e., the near 0 s occlusion period), while the number network is invoked in the longest occlusion condition (i.e., the 2 s occlusion period). As such, network selection is justified by *ecological rationality* (Todd & Gigerenzer, 2007, 2012; Nobandegani & Shultz, 2019), according to which selected heuristics are well-adapted to the environmental conditions in which they are used. Relatedly, recent work has effectively cast heuristic selection as a form of rational meta-reasoning (Lieder & Griffiths, 2017).

Our use of a number network is further supported by recent evidence that young infants form accurate probabilistic expectations based on numerosity, performing accurate probabilistic inferences similar to those studied here (Denison, Reed, & Xu, 2013; Xu & Garcia, 2008). Also, several studies showed children's and preverbal infants' sensitivity to distance in their causal judgments (Schlottmann & Surian, 1999; Scholl & Tremoulet, 2000), physical reasoning (Teglas et al., 2011), and integration of physical information (Wilkening, 1981), lending support to our use of a distance network.

Each NPLS network begins learning with just 1 input unit and 1 output unit. With an error-inducing score-threshold of .52, these networks recruit from 0-4 hidden symmetric-sigmoid units, whose activations range from -.5 to .5. The output unit has an asymmetric sigmoid activation function with output ranging from 0-1. The most common network topologies are 3 hidden units on 1 layer (for a 1-3-1 structure), or 2 hidden units on 1 layer and 1 on another layer (for a 1-2-1-1 or 1-1-2-1 structure). We run 20 networks in each of the 12 cells of the experiment: 2 numbers x 2 distances x 3 time periods. After training, these networks are tested on the ball exiting in each of the 12 experimental conditions. Both output activations (reflecting probability learning) and network error (reflecting looking time and surprise) are recorded.

Results

Under these parameters, the networks learn their probability distributions to a high level of accuracy (Figure 3). Mean output activations (with SDs) over 20 networks trained on number are presented in Figure 3 (top). These networks learn the 3:1 probability distribution of relative numbers of balls inside the container, ignoring information on distance from the exit. As such, these networks come to expect that a majority-colored-ball exit is about 3 times more likely than a minority-colored-ball exit, indicative of the numerosity heuristic. A repeated measures ANOVA reveals a main effect of number, $F(1,19) = 392, p < .001, \eta_p^2 = .954$, no effect of distance, $F(1,19) = .879, p = .36, \eta_p^2 = .044$, and no interaction between number and distance, $F(1,19) = 1.077, p = .328, \eta_p^2 = .05$.

Comparable results for networks trained on distance from the exit are shown in Figure 3 (bottom) in the 0 s condition. These networks come to expect exit of the closest ball with a probability of about .75, regardless of ball color frequency, indicative of the distance heuristic. In this case, there is a main effect of distance, $F(1,19) = 428, p < .001, \eta_p^2 = .957$, no effect of number, $F(1,19) = 1.335, p = .262, \eta_p^2 = .066$, and no interaction, $F(1,19) = 1.05, p = .317, \eta_p^2 = .053$.

Considering the high levels of statistical significance reached, the networks undoubtedly perform more accurately than the infants they are simulating. This shows that these networks are well suited to learning and representing probability distributions. To match the lesser precision of infants, we could run the networks with higher score-thresholds, resulting in less learning and fewer hidden units.

The probability distributions in Figure 3 document the knowledge required to implement and explain the surprise reactions. Neural network modelers often use network error as a measure of surprise at unexpected vs. expected outcomes (Althaus, Gliozzi, Mayor, & Plunkett, 2020; Oakes, Madole, & Cohen, 1991; Shultz & Cohen, 2004). This modeling choice is well justified as network error quantitatively captures the gap between observation (the stimuli presented to participants) and expectation (what stimuli participants expect to receive). Surprise in infants is typically measured as increased looking time at an unexpected outcome, as compared to an expected outcome.

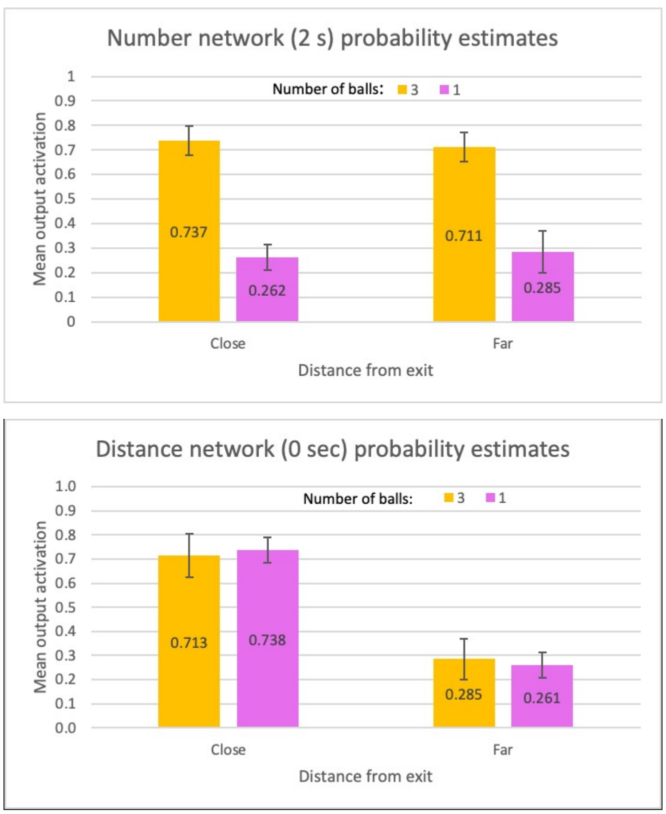

Figure 3: (Top) Mean output activations over 20 networks trained on number. These networks learn the 3:1 probability distribution of relative numbers of balls inside the container (see Figure 1), and ignore any information regarding their distance from exit before the occurrence of occlusion. (Bottom) Mean output activations over 20 networks trained on distance from exit. These networks learn a 3:1 probability distribution and come to expect exit of the closest ball with a probability of about .75, regardless of ball color, indicative of the distance heuristic.

Figures 4 (top) and 4 (middle) plot network error to expected and unexpected outcomes for number and distance networks, respectively. Like the infants in the 2 s occlusion condition (Teglas et al., 2011), number networks show more error, indicating surprise, to exit of a minority-colored ball than a majority-colored ball (Figure 4 (top)). A repeated-measures ANOVA reveals a main effect of number, $F(1,19) = 393, p < .001, \eta_p^2 = .954$, but not distance, $F(1,19) = 1.0, p = .330, \eta_p^2 = .05$, and no interaction between distance and number, $F(1,19) = 1.0, p = .330, \eta_p^2 = .05$, as in the infant experiments (Teglas et al., 2011).

Analogously, like infants in the 0 s condition (Teglas et al., 2011), distance networks display higher error, signaling surprise, to exiting of a ball positioned far from the exit than a ball close to the exit (Figure 4 (middle)). ANOVA results here show a main effect of distance, $F(1,19) = 428, p < .001, \eta_p^2 = .958$, no effect of number, $F(1,19) = 1.034, p = .322, \eta_p^2 = .052$, and no interaction, $F(1,19) = 1.064, p = .315, \eta_p^2 = .053$, as in the infant experiments (Teglas et al., 2011).

To simulate the condition with a 1 s delay between exit and occlusion removal, there is no training and no special network. Instead, we assume that our system infers what happens when both number and distance effects are operative, by taking the mean of the error predictions from each pair of number and distance networks. As a 1 s delay is halfway between a 0 s and a 2 s delay, it is plausible to assume that the infants are maximally uncertain as to which of the two networks should be used, and hence rely equally on both for making an inference. These means are plotted in Figure 4 (bottom), along with SDs.

As with the infants in the 1 s condition (Teglas et al., 2011), there are now main effects for both number, $F(1,19) = 119, p < .001, \eta_p^2 = .862$, and distance, $F(1,19) = 370, p < .001, \eta_p^2 = .951$. As with the infants (Teglas et al., 2011), there is no significant interaction effect, $F(1,19) = 0.00$, reflecting additive use of number and distance information.

The Pearson correlation between mean network error and mean infant looking time across the 12 experiment conditions is $.94, p < .01$. This is identical to the Pearson correlation of .94 between infant looking time and the Bayesian ideal observer model (Teglas et al., 2011).

Discussion

We simulate these infant results with a novel neural-network model comprising two modules (Figure 2). These modules are implemented by a pair of NPLS networks that each receive coded versions of the same number and distance input used in the infant experiments (Teglas et al., 2011). A number network learns an instantiation of the number heuristic according to which a majority-colored item is more likely to randomly exit a container than a minority-colored item, provided that the content of the container is sufficiently shuffled, while a distance network learns an instantiation of the distance heuristic according to which an item closest to an exit is more likely to randomly exit than items that are farther away. This probabilistic knowledge enables the networks to simulate surprise at unexpected outcomes, modeled as network error, the discrepancy between what is expected and what actually happens.

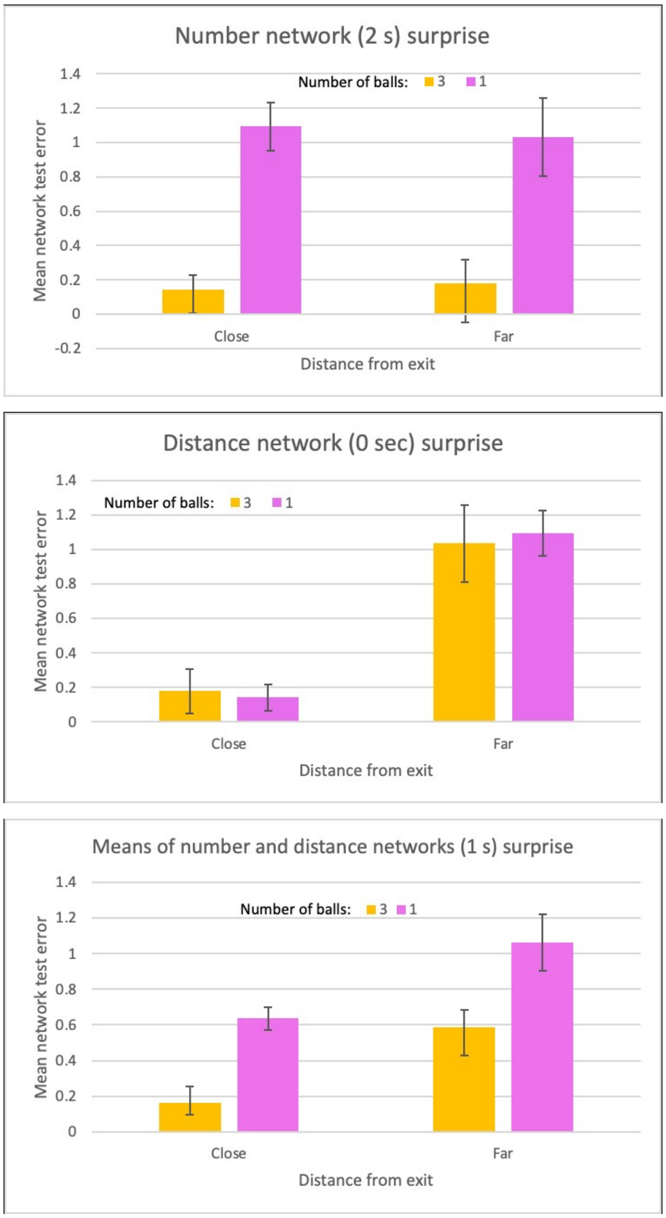

Figure 4: (Top) Mean network error over 20 number networks, as a function of number and distance information, along with SDs. (Middle) Mean network error over 20 distance networks, as a function of number and distance, along with SDs. (Bottom) Mean network error (and SDs) over 20 networks. Our system infers what happens when both number and distance effects are operative, by taking the mean of the error predictions from each pair of number and distance networks.

Importantly, as we discussed in the Simulations section, the use of the number and distance heuristics in our modeling work is well-justified by *ecological rationality*, according to which heuristics are well-adapted to the environmental conditions in which they are used (Todd & Gigerenzer, 2007, 2012).

Past work (Shultz & Nobandegani, 2022a) has shown that models similar to those used here can account for an extensive series of empirical findings on infant learning and reasoning with probabilities, including sample-to-population and population-to-sample generalizations (Xu & Garcia, 2008), the emergence of the ability to generalize from samples to populations at about 6 months of age (Denison et al., 2013), and how probabilistic knowledge guides preverbal infant's choice behavior (Denison & Xu, 2010, 2014).

We also favor the neural network model presented here because it shows how relevant probability distributions can be established. In contrast, Bayesian researchers typically supply probability distributions for free, even if they differ considerably across different tasks. It is doubtful that specific probability distributions would be supplied by biological evolution because humans can succeed on tasks involving many different distributions. It is more likely that human infants are innately equipped with an ability to learn such distributions from some experiences with the distributions. The idea that innateness proposals should not be restricted to knowledge representations has been well discussed elsewhere (Elman et al., 1996; Shultz, 2003).

It is also worth noting that both the training and test phases of the SDCC networks used in our work show remarkable speed and efficiency and the relevant probabilities emerge naturally at the network outputs, all of which elevates the psychological plausibility of the proposed model.

It remains unknown how much of human intuitive physics is innate and how much of it is learned through the course of development. Perhaps, much like in the case of many other areas of science dealing with successful, but radically different and largely opposing, theories of the same empirical phenomenon, the truth likely lies somewhere in between. Presumably, human infants come to this world with some innate, but somewhat primitive, core knowledge of physics, e.g., that the physical world comprises objects (Spelke, 2000). Infants then expand on that knowledge presumably by way of learning simple, resource-efficient rules of thumb (aka heuristics), collectively allowing them to make inferences about ordinary physical events. Future research should explore how infants' primitive physics knowledge supports the learning of formal physics knowledge later in development.

While many questions remain open, the work presented here provides a fresh perspective on possible neural underpinnings of intuitive physics in preverbal infancy, substantiating the view that the remarkable array of physical knowledge demonstrated by infants might be accounted for by a set of simple, fast, resource-efficient heuristics approximating high-level, intuitive physics principles. In the spirit of an extensive body of work on the role of heuristics in human cognition (Gigerenzer & Gaissmaier, 2011; Gigerenzer & Selten, 2002; Gigerenzer & Todd, 2000), our work advocates understanding a good deal of infant physics knowledge in terms of a bag of heuristics: a set of simple, fast, rules of thumb that yield sufficiently good results.

Acknowledgments

This work was supported in part by an operating grant to TRS from the Natural Sciences and Engineering Research Council of Canada.

References

Althaus, N., Gliozzi, V., Mayor, J., & Plunkett, K. (2020). Infant categorization as a dynamic process linked to memory. *Royal Society Open Science*, *7*(10), 200328.

Baluja, S., & Fahlman, S. E. (1994). *Reducing network depth in the cascade-correlation learning architecture* (Tech. Rep.). Pittsburgh, PA.

Battaglia, P. W., Hamrick, J. B., & Tenenbaum, J. B. (2013). Simulation as an engine of physical scene understanding. *Proceedings of the National Academy of Sciences*, *110*(45), 18327–18332.

Denison, S., Reed, C., & Xu, F. (2013). The emergence of probabilistic reasoning in very young infants: evidence from 4.5-and 6-month-olds. *Developmental Psychology*, *49*(2), 243.

Denison, S., & Xu, F. (2010). Twelve-to 14-month-old infants can predict single-event probability with large set sizes. *Developmental Science*, *13*(5), 798–803.

Denison, S., & Xu, F. (2014). The origins of probabilistic inference in human infants. *Cognition*, *130*(3), 335–347.

Elman, J. L., Bates, E. A., & Johnson, M. H. (1996). *Rethinking innateness: A connectionist perspective on development* (Vol. 10). Cambridge, MA: MIT press.

Gigerenzer, G., & Gaissmaier, W. (2011). Heuristic decision making. *Annual Review of Psychology*, *62*(1), 451–482.

Gigerenzer, G., & Selten, R. (2002). *Bounded rationality: The adaptive toolbox*. Cambridge, MA: MIT press.

Gigerenzer, G., & Todd, P. M. (1999). *Simple heuristics that make us smart*. Oxford: Oxford University Press, USA.

Kharratzadeh, M., & Shultz, T. (2016). Neural implementation of probabilistic models of cognition. *Cognitive Systems Research*, *40*, 99–113.

Lieder, F., & Griffiths, T. L. (2017). Strategy selection as rational metareasoning. *Psychological Review*, *124*(6), 762.

Nobandegani, A. S., & Shultz, T. R. (2017). Converting cascade-correlation neural nets into probabilistic generative models. In: G. Gunzelmann, A. Howes, T. Tenbrink, & E. J. Davelaar (Eds.), *Proceedings of the 39th Annual Conference of the Cognitive Science Society* (pp. 1029-1034). Austin, TX: Cognitive Science Society.

Nobandegani, A. S., & Shultz, T. R. (2018). Example generation under constraints using cascade correlation neural nets. *Proceedings of the 40th Annual Conference of the Cognitive Science Society* (pp. 2385-2390). Austin, TX: Cognitive Science Society.

Nobandegani, A. S., & Shultz, T. R. (2019). Toward a formal science of heuristics. *Proceedings of the 41st Annual Conference of the Cognitive Science Society* (pp. 877-882). Austin, TX: Cognitive Science Society.

Oakes, L. M., Madole, K. L., & Cohen, L. B. (1991). Infants' object examining: Habituation and categorization. *Cognitive Development*, *6*(4), 377–392.

Schlottmann, A., & Surian, L. (1999). Do 9-month-olds perceive causation-at-a-distance? *Perception*, *28*(9), 1105–1113.

Scholl, B. J., & Tremoulet, P. D. (2000). Perceptual causality and animacy. *Trends in Cognitive Sciences*, *4*(8), 299–309.

Shultz, T., & Nobandegani, A. S. (2022a). A computational model of infant learning and reasoning with probabilities. *Psychological Review*, *129*(6), 1281-1295.

Shultz, T., Nobandegani, A. S., & Fahlman, S. E. (2020). Cascade-Correlation. In D. Phung, G. I. Webb, & C. Sammut (Eds.), *Encyclopedia of Machine Learning and Data Science* (pp. 1-11). New York: Springer.

Shultz, T. R. (2003). *Computational developmental psychology*. Cambridge, MA: MIT Press.

Shultz, T. R. (2012). A constructive neural-network approach to modeling psychological development. *Cognitive Development*, *27*(4), 383–400.

Shultz, T. R. (2013). Computational models in developmental psychology. In P. D. Zelazo (Ed.), *Oxford handbook of developmental psychology, Vol. 1: Body and mind (pp. 477–504).* New York: Oxford University Press.

Shultz, T. R. (2014). Knowing when to quit on unlearnable problems: Another step towards autonomous learning. In J. Mayor & P. Gomez (Ed.), *Computational Models of Cognitive Processes* (pp. 211-221). London: World Scientific.

Shultz, T. R. (2017). Constructive artificial neural-network models for cognitive development. In N. Budwig, E. Turiel, & P. D. Zelazo (Eds.), *New Perspectives on Human Development* (pp. 1326). Cambridge: Cambridge University Press.

Shultz, T. R., & Cohen, L. B. (2004). Modeling age differences in infant category learning. *Infancy*, *5*(2), 153–171.

Shultz, T. R., & Nobandegani, A. S. (2022b). Modeling the learning and use of probability distributions in chimpanzees and humans. *Proceedings of the 44th Annual Con-*

ference of the Cognitive Science Society (pp. 1947-1953). Austin, TX: Cognitive Science Society.

Spelke, E. S. (2000). Core knowledge. *American Psychologist*, *55*(11), 1233-1243.

Téglás, E., Vul, E., Girotto, V., Gonzalez, M., Tenenbaum, J. B., & Bonatti, L. L. (2011). Pure reasoning in 12-month-old infants as probabilistic inference. *Science*, *332*(6033), 1054–1059.

Todd, P. M., & Gigerenzer, G. (2007). Environments that make us smart: Ecological rationality. *Current Directions in Psychological Science*, *16*(3), 167–171.

Todd, P. M., & Gigerenzer, G. E. (2012). *Ecological rationality: Intelligence in the world*. Oxford: Oxford University Press.

Ullman, T. D., Spelke, E., Battaglia, P., & Tenenbaum, J. B. (2017). Mind games: Game engines as an architecture for intuitive physics. *Trends in Cognitive Sciences*, *21*(9), 649–665.

Wilkening, F. (1981). Integrating velocity, time, and distance information: A developmental study. *Cognitive Psychology*, *13*(2), 231–247.

Xu, F., & Garcia, V. (2008). Intuitive statistics by 8-month-old infants. *Proceedings of the National Academy of Sciences*, *105*(13), 5012–5015.