Using physical simulations to predict children's and adults' preference for physical interactions

Anonymous CogSci submission

Abstract

Curiosity is a fundamental driver of human behavior, and yet because of its open-ended nature and the wide variety of behaviors it inspires in different contexts, it is remarkably difficult to study in a laboratory context. A promising approach to developing and testing theories of curiosity is to instantiate them in artificial agents that are able to act and explore in a simulated environment, and then compare the behavior of these agents to humans exploring the same stimuli. Here we propose a new experimental paradigm for examining children's–and curious AI agents'–curiosity about objects' physical interactions, using a task that is both open-ended enough to allow room for curiosity, but also constrained enough to make detailed behavioral comparisons. We compared children's choices (N=66; 3-7 years of age) to those of adults (N=155), and find increasingly strong preference for choices affording particular relations (e.g., containment) across development, as well as an adult bias for choosing a support relation.

Keywords: curiosity; novel objects; object interactions; intuitive physics

Introduction

Curiosity is a hallmark aspect of human intelligence. From infants exploring the objects in their environment to scientists exploring the frontiers of our solar system, humans are highly motivated to seek out new knowledge and experiences. However, although such exploratory behavior has long been recognized as a critical component of human learning (James, 1983) and cognitive development (Gopnik, Meltzoff, & Kuhl, 2009; Piaget, 1952), formal theories that explain human curiosity and how it drives exploratory behavior have remained elusive (Kidd & Hayden, 2015). Moreover, extant theories have rarely provided quantitatively precise enough predictions to be directly compared to empirical measurements of curiosity-driven behavior in humans.

The goal of the current paper is to help close this gap by proposing a framework for advancing our theoretical understanding of curiosity. First, we present an empirical investigation of the pattern of actions taken by children and adults in a novel physical exploration task. We then present a set of heuristic models of curiosity based on a variety of metrics of simulated interactions (e.g., measuring likelihood of a dropped object coming to rest on a target object), and test whether these features predict people's preferences in the same task. Finally, we evaluate the quantitative correspondence between the exploratory actions taken by an artificial agent that instantiates our computational theory and those taken by human children on the same set of tasks.

Our empirical study is motivated by a large body of prior work in developmental psychology investigating the development of knowledge about physical objects, their properties, and how they interact (CITE Spelke, Baillargeon (2007)), thus making exploration of physical objects a natural choice of domain to explore the implications of our theory. However, the measures used in this literature have typically focused on either children's ability to discriminate types of events, or their ability to carry out particular tasks (e.g., shape sorting), but have not examined which types of interactions children are most interested in testing. For instance, a typical measure of physical curiosity has been longer looking times to surprising events than to expected events (e.g., objects appearing to pass through a wall; Stahl & Feigenson (2015); Baillargeon (2007)). Other work that has investigated exploratory behavior in children have used more granular measures, such as counting the number of functions discovered while playing with novel and complex toy (e.g., Cook, Goodman, & Schulz, 2011; Bonawitz, Schijndel, Friel, & Schulz, 2012; Gweon, Pelton, Konopka, & Schulz, 2014). Moreover, these studies have yet to directly constrain theories that predict which functions children will discover, and in what sequence. What both of the standard looking-time and toy-exploration measures fail to capture when characterizing children's curiosity are the ways in which children actively learn by intervening on the world and observing the consequences of their actions (Gopnik et al., 2009; Gureckis & Markant, 2012). To address these limitations, we developed a novel physical exploration task in which children, adults, or AI agents choose which series of physical experiments to perform and observe the results of each in real time.

In sum, this paper presents a paradigm for testing and building theories of curiosity, and defining its role in guiding the development of knowledge about the physical world. We present a study that measures adults' (Experiment 1) and children's (Experiment 2) preferences in a novel physical exploration task, a set of 20 heuristic models of curiosity that operate on simulations of this same task in a 3D environment, and comparisons between the preferences of children, adults, and these heuristic models on the same set of trials. Overall, our approach of coordinating the development of task-performing computational models and detailed measurement of human action selection on the same tasks has the promise to lead to more robust and precise theories of how curiosity guides cognitive development.

Experiment 1: Adults

To investigate the systematicity of people's preferences for physical interactions between objects, we began by studying adults, whom we might expect to be less idiosyncratic and thus more consistent in their choices than children. Our design is motivated by the results of a pilot study conducted in-person in January, 2020, in which we asked adults to select which of a pair of 3D-printed toy blocks (see Figure 1) they would like to drop on a given target object from the set, or vice-versa: on which of a given pair of target objects would they like to drop a given object. In the pilot study (N=15), the pairs of target or drop objects were chosen essentially at random, but were the same for all participants. We were surprised to find consistency in adults' preferences for many of the trials: especially when given a drop object and asked to choose one of two target objects, adults were quite often (75-90%) targeting that object that would either *contain* or *support* the dropped object (i.e., the pipe could contain the cone; the pentagonal prism could support the octahedron). This consistency in their choices is remarkable given that there are many other possible goals that people might choose in order to make something "interesting" happen: they might attempt to make the dropped object roll or bounce far from the target, or rebound in an unexpected direction, but in fact most people settled on attempting support or containment relations. Thus, we set out to examine the development of these preferences, first in a large adult sample, and then in young children.

Method

Participants Participants were 200 adults recruited online via Amazon Mechanical Turk who were paid $1 for their participation.

Materials Stimuli were images of 3D objects produced using Blender 3D-modeling software. The nine objects, depicted in Figure 1, were bowl, cone, dumbbell, octahedron, pentagon (pentagonal prism), pipe, pyramid, torus, triangular prism, and ball (sphere; not pictured).

Design The experiment consisted of 20 drop trials, completed by each participant in one of four pseudorandom orders. Participants were randomly assigned to each order. Each drop trial displayed an object X that is to be dropped on one of two other objects (Y or Z), as described in the following procedure. In the 20 trials, each of the above 3D objects was used twice as the drop object (X), and appeared four times as potential targets. Instead of randomly sampling the target objects (Y or Z), each trial was designed to allow for one of two goal relations (or affordances): containment, or support. That is, the target objects were chosen so that either (but not both) of them would be capable of either supporting or (partially) containing the dropped object. For example, if asked to drop the cone on either the torus or the octahedron,

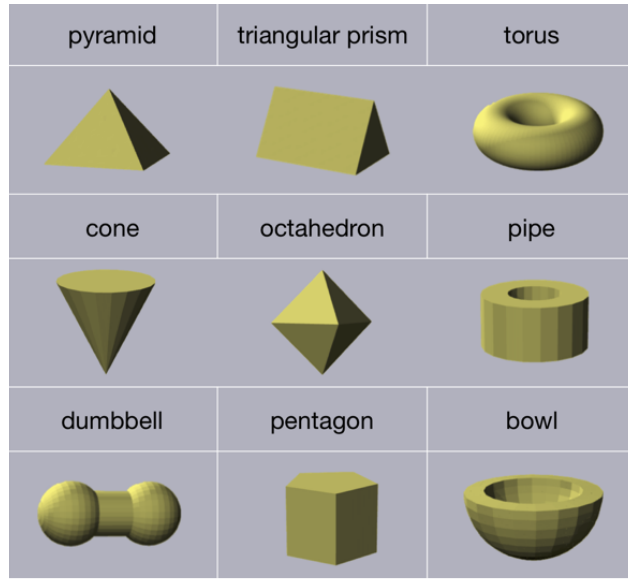

Figure 1: Nine of the 10 3D objects used for dropping and as targets (not pictured: ball).

a containment relation could be achieved between the cone and the torus, but not the octahedron (and support is unlikely due to the small surface area of the octahedron). An example support relation is choosing to drop the pentagonal prism on the (equal diameter) pipe rather than the pyramid. Half of the 20 trials were designed to afford containment relations, and the other half afforded support relations. We refer to these designed containment and support relations as the "targets", and the alternative objects as the "distractors". Distractors for each trial were the same for all participants, and were selected to 1) not afford containment/support of the dropped object, and 2) roughly equate the frequency of appearance of all 10 objects.

Procedure Participants were instructed that they would be helping choose which toys to include in a new set of children's blocks in order to make them the most interesting. Participants were then given a practice trial (see Figure 2), in which they were told that they should imagine dropping a given object (the torus) on each of two toys (the dumbbell or the pentagonal prism) in a bin. They were then asked to choose which toy to drop the torus on, and prompted to choose the most interesting or surprising combination. After making their choice on the practice trial, participants were shown a 10-second video recording of 3D-printed plastic toys carrying out their chosen interaction (e.g., if they chose the dumbbell, they would see the torus dropped on the dumbbell). This was done to ensure that they understood the consequences of the choice they made. After the video of the torus drop, participants were given a sequence of 20 more drop trials asking them to choose which toy (X or Y) they would like to drop object Z on, but without videos. Four catch trials were interspersed among the 20 drop trials, which asked them to indicate which object they had just dropped on the previous trial (3-alternative forced choice). The goal of the catch trials was to encourage participants to attend to the objects and the choices they made, and to exclude any participants who were incorrect on 2 or more of the catch trials. Finally, there were two free response explanation trials mixed among the drop trials, querying participants, “Why did you make that choice?” These free responses were examined to determine if participants had particular goals in mind when making a choice, and how prevalent those goals were among participants.

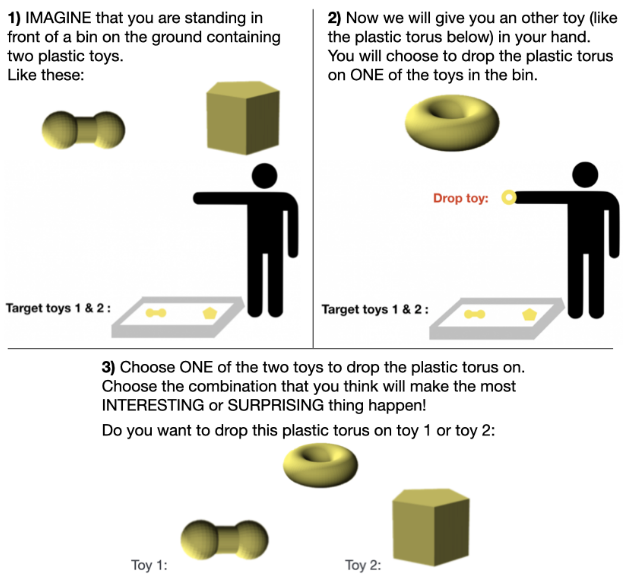

Figure 2: Example trial from Experiment 1 with adults.

Results

We analyzed data from the 155 adults who completed the experiment and got at least 3 of the 4 catch trials correct (45 participants were excluded for not meeting this criterion). Table 1 shows the proportion of participants that chose the designed target relation for each of the 20 trials. Based on the binomial distribution, any of the trials on which more than 90 of the 155 participants agree (i.e., >0.59 or <0.41) significantly differ from chance [adjust criterion for multiple comparisons?]. As can be seen in Table 1, adults participants (“adult” column) significantly preferred the target relation for all 10 of the containment trials, and significantly preferred five targets on the 10 support trials; the other five support trials did not significantly differ from chance.

Were adults more consistent in their choices on support or containment trials? Averaging each subject’s binary responses (1=target relation, 0=alternative) for each trial type revealed a stronger preference for containment relations ($M = 0.87$) than support relations ($M = 0.63$; paired $t(154) = 12.69$, $p < .001$, $d = 1.02$).

relation	drop	target	child	adult
contain	bowl	pyramid	0.72	0.96
contain	cone	torus	0.73	0.94
contain	dumbbell	pipe	0.67	0.92
contain	octahedron	pipe	0.58	0.90
contain	pentagon	bowl	0.69	0.88
contain	pipe	cone	0.58	0.88
contain	pyramid	torus	0.71	0.87
contain	sphere	bowl	0.80	0.86
contain	torus	cone	0.67	0.81
contain	trig prism	bowl	0.65	0.71
support	bowl	torus	0.71	0.85
support	cone	trig prism	0.45	0.82
support	dumbbell	pentagon	0.38	0.74
support	octahedron	pentagon	0.45	0.64
support	pentagon	pipe	0.57	0.62
support	pipe	torus	0.60	0.56
support	pyramid	trig prism	0.52	0.55
support	sphere	pipe	0.58	0.54
support	torus	bowl	0.70	0.53
support	trig prism	pentagon	0.62	0.46

Table 1: Children’s and adults’ target preferences per trial.

Experiment 2: Children

Experiment 1 demonstrated that adults show consistent preferences for dropping objects on target objects that afford containment relations, and to a lesser extent support relations. Experiment 2 investigates whether these strong preferences are present even in young children, recognizing that children may find different types of physical interactions (e.g., rolling, bouncing, unpredictability) of greater interest.

Method

We adapted the same materials and procedure used in Experiment 1 for use in an online experiment with children, in order to directly compare the results for children and adults.

Participants Participants were 73 children recruited online via outreach through a local nursery school and targeted Facebook Ads. Participant exclusions were made based on cases where i) the participant did not complete more than half of the study play session or ii) the parent did not consent for video recording of study. After exclusions, results from 66 were analyzed, including 17 3-year-olds, 15 4-year-olds, 16 5-year-olds, 16 6-year-olds, and 2 7-year-olds.

Materials The materials were the same as those used in Experiment 1, except that the trials were adapted for presentation via Zoom in a Keynote. Introductory slides were added for informed consenting process and screen configuration, where researcher’s video was placed directly below drop object X and in between target objects Y and Z (see Figure 3).

Procedure After the parent provided informed consent, children were assigned to one of two pseudorandom trial orders in counterbalanced order. A video recording was taken

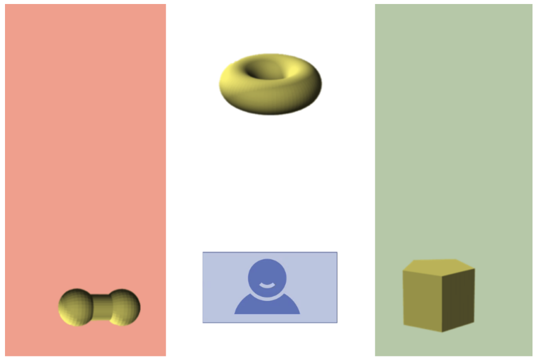

Figure 3: Zoom screen configuration for children in Experiment 2. Parents were asked to place the experimenter's video feed in the bottom center so the experimenter could point up to the drop object and left and right at each target choice.

of the online session.

To accompany the practice trial, children were shown the three physical objects on video, to ensure understanding of their 3D properties. Verbal responses were collected from children to indicate the selection of target object Y or Z. To prevent left/right mapping confusion, target objects Y and Z were overlaid on thirds of the screen with either red or green background color, which were counterbalanced across trials. The child was asked to verbally select target object and then to indicate whether it was on the "red side or the green side."

Results

We examined children's choice of target objects to determine if children were choosing randomly, or had consistent preferences for some objects. Table 1 shows the proportion of children ("child" column) choosing the designed target relation for each of the 20 trials, alongside the adult choice proportions from Experiment 1 [maybe just drop Table 1?]. Based on the binomial distribution, any of the trials on which more than 41 of the 66 participants agree (i.e., >0.64 or <0.36) significantly differ (p<.05) from chance. Children significantly preferred the designed relation on eight of the 10 containment trials, while only significantly preferring two of the 10 support relation choices. On the remaining two containment trials and the other eight support trials, children's preferences also did not significantly differ from chance against the designed relation.

Were children more consistent in their choices on support or containment trials? As for adults, averaging each child's binary responses for each trial type revealed a stronger preference for containment relations ($M = 0.68$) than support relations ($M = 0.56$; paired $t(154) = 12.69$, $p < .001$, $d = 1.02$). Overall, children did not show a significant preference for choosing the support relation [should we add a regression with trial_type * age (3 - 18=adult)?]. Figure 4 shows the proportion of participants' choice of the designed target relation broken down by relation type and age group, showing a preference for the designed relations that increased with age, and that was stronger for containment.

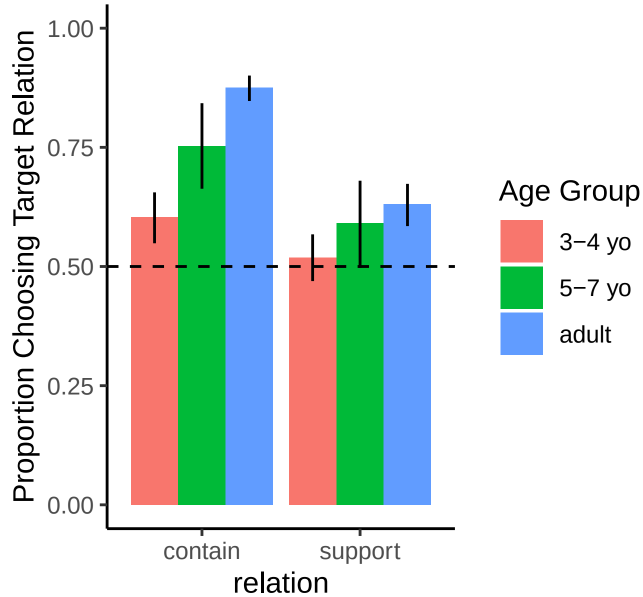

Figure 4: Comparison of children's vs. adults' choices for each type of trial, with bootstrapped 95% confidence intervals.

Comparison of Children's vs. Adults' Preferences

Finally, we compare the trial-level preferences of children and adults. Figure 5 shows adults' (Experiment 1) vs. children's (Experiment 2) choice proportions for the target relations, colored by relation type (visualizing the same data presented in Table 2). If they corresponded perfectly, they would fall along the dotted $y = x$ line, but children's preferences mostly fall short of that line, lying closer to $y = 0.5$. At a glance, this seems to support the hypothesis that "children are noisy adults." We will re-examine this hypothesis with model-based predictions for each trial.

We also examined the split-half reliability of children's and adults' choices by repeatedly splitting the samples in half and testing the correlation between the halves. Based on 100 random samples, the split-half reliability for adults is r = 0.95 (sd=0.02), and for children is r = 0.61 (sd=0.09). The overall split-half reliability on the containment trials is r = 0.82 (sd=0.07), and for support trials is r = 0.76 (sd=0.09). [should we separately report these for adults and children?]

Figure 4 shows participants' mean proportion of unique objects dropped as a function of age. Children of all ages sampled approximately equal proportions of the objects for dropping–roughly 70%, which is close to the 75% that would be expected if they were selected by chance (9 unique object occurring across 12 trials).

Comparison to Physical Simulations

Can children's and adults' preferences be predicted by heuristic models of the simulated physical interactions? Using a physics engine and 3D models of the objects we showed

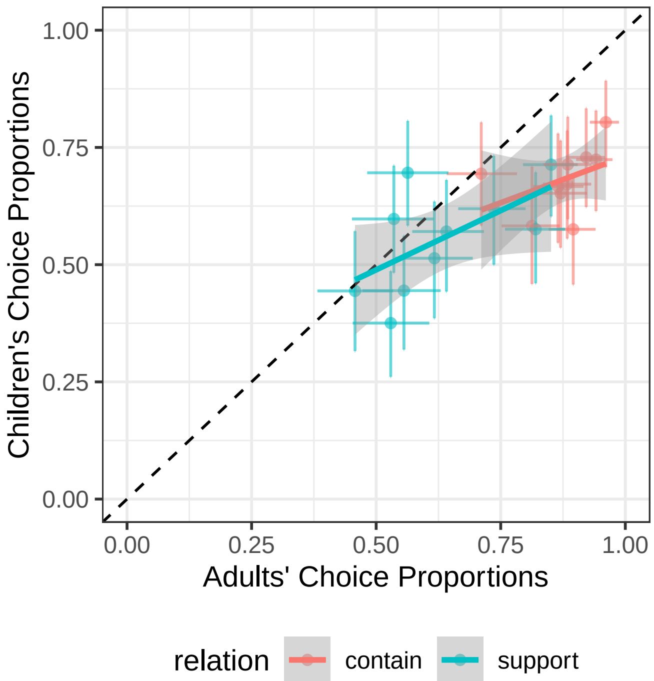

Figure 5: Comparison of children's vs. adults' preferences on each trial (dot), with bootstrapped 95% confidence intervals.

to participants, we constructed a simulation of each trial of the experiment. For each trial's the two possible drop object choices, we ran 100 simulated drops, and for each drop calculated a variety of features measuring the state of the model after the drop was completed. We selected features that we thought may provide good metrics for what people would found interesting. For example, we collected for each drop the mean length of time before both objects come to rest (*avg_len*), as well as the standard deviation (SD) of that time (*len_std*). Some features were calculated separately for drop and target objects, such as: the SD of the object's final position (*obj_final_position_std_objects*) as well as the inverse of that (*obj_final_position_invstd_objects*), the mean final distance from the drop location (*avg_final_radius_objects*), the maximum final distance from the drop location (*avg_max_radius_objects*), the ... (*max_radius_std_objects*), ... (*normed_velocity_std_after_first_collision_objects*)

We measure each of these features across each possible drop in the experiment 100 times, and then use the resulting metrics to generate relative preferences for each of the drop choices on each trial. Finally, we compare how well each of these simple heuristic models account for children's and adults' drop preferences on each trial.

Discussion

This study aimed to investigate what physical interactions between objects most evoke curiosity. After finding surprising consistency in an in-person pilot study of adults, an online study designed to make possible the same support and containment relations between dropped and target objects replicated this finding of consistency: adults are interested in dropping objects into containers, and to a lesser degree on top of supporting objects. Experiment 2 found the same pattern in children, but with greater noise – even to the extent that younger children were at-chance at choosing the support relation on most trials.

Thus, we find an early-emerging interest in containment (Casasola, Cohen, & Chiarello, 2003) that persists into adulthood, along with a weaker interest in support relations. Is containment inherently more satisfying, or is there greater interest because it is more achievable? (Less susceptible to small amounts of error) Are children less interested in relations that are more difficult to achieve with more motor noise? Or are they less aware (or simply less interested in) support? Or do children find some other goal relatively more enticing?

Thexe results will serve as a test set for evaluating computational theories of curiosity, which in turn are inspired by classic descriptive theories positing that curiosity-driven behavior was the consequence of attentional capture by novel stimuli (e.g., Berlyne, 1954; Fantz, 1964), as well as theories providing qualitative explanations for curiosity-driven behavior as the result of learners' preferences for stimuli that are "moderately discrepant" in relation to their current knowledge state, and thereby provide opportunities to learn (Kinney & Kagan, 1976). Recent studies have provided empirical support for this basic idea: 7- to 8-month-old infants are sensitive to stimulus complexity and prefer to look at moderately complex stimuli rather than simple or highly-complex stimuli (Kidd, Piantadosi, & Aslin, 2012, 2014). According to these theories, curiosity-driven behavior reflects a gap between the learner's knowledge and the state of the world (Loewenstein, 1994), and thus they predict that as a learner gains additional knowledge, their preferences will shift toward more complex stimuli (Dember & Earl, 1957). Although these theories have provided important qualitative insights, one of their major limitations is that they do not provide precise ways to characterize a learner's current knowledge state, how physical states of the world are represented in their mind, nor how discrepancies between a learner's knowledge and such states are compared.
Our approach leverages theory and insights from reinforcement learning (RL), including the value of mechanisms that equip RL agents with "intrinsic motivation" — a drive to explore the state space even when rewards are sparse or absent, and favor actions with uncertain outcomes, because they may result in the discovery of new policies with high expected values (Schmidhuber, 2010). Our modeling approach is resonant with recent work that has instantiated such intrinsic-motivation mechanisms in robots to help them learn robust ways to predict physical events in the world world (Oudeyer, Baranes, & Kaplan, 2013; Oudeyer & Kaplan, 2007), although this work typically involves pretraining models on a separate physical prediction task before implementing curiosity-driven learning, and has not yet been directly compared to human behavior in the same prediction tasks.

Acknowledgements

This work was funded by HAI seed grant #. We thank X and Y for helpful comments.

References

Baillargeon, R. (2007). The acquisition of physical knowledge in infancy: A summary in eight lessons. In *Blackwell handbook of childhood cognitive development* (pp. 47–83). Blackwell Publishers Ltd.

Berlyne, D. E. (1954). A theory of human curiosity. *British Journal of Psychology*, *45*, 180–191.

Bonawitz, E. B., Schijndel, T. van, Friel, D., & Schulz, L. (2012). Children balance theories and evidence in exploration, explanation, and learning. *Cognitive Psychology*, *64*, 215–234.

Casasola, M., Cohen, L. B., & Chiarello, E. (2003). Six-month-old infants' categorization of containment spatial relations. *Child Development*, *74*(3), 679–693.

Cook, C., Goodman, N. D., & Schulz, L. E. (2011). Where science starts: Spontaneous experiments in preschoolers' exploratory play. *Cognition*, *120*, 341–349.

Dember, W. N., & Earl, R. W. (1957). Analysis of exploratory, manipulatory, and curiosity behaviors. *Psychological Review*, *64*, 91–96.

Fantz, R. L. (1964). Visual experience in infants: Decreased attention to familiar patterns relative to novel ones. *Science*, *146*(3644), 668–670. Retrieved from `http://www.jstor.org/stable/1714550`

Gopnik, A., Meltzoff, A. N., & Kuhl, P. K. (2009). *The scientist in the crib: Minds, brains, and how children learn*. HarperCollins.

Gureckis, T. M., & Markant, D. B. (2012). Self-directed learning: A cognitive and computational perspective. *Perspectives on Psychological Science*, *7*(5), 464–481.

Gweon, H., Pelton, H., Konopka, J. A., & Schulz, L. E. (2014). Sins of omission: Children selectively explore when teachers are under-informative. *Cognition*, *132*, 335–341.

James, W. (1983). *Talks to teachers on psychology and to students on some of life's ideals* (Vol. 12). Harvard University Press.

Kidd, C., & Hayden, B. Y. (2015). The psychology and neuroscience of curiosity. *Neuron*, *88*(3), 449–460.

Kidd, C., Piantadosi, S. T., & Aslin, R. N. (2012). The goldilocks effect: Human infants allocate attention to visual sequences that are neither too simple nor too complex. *PLOS ONE*, *7*(5), 1–8. `http://doi.org/10.1371/journal.pone.0036399`

Kidd, C., Piantadosi, S. T., & Aslin, R. N. (2014). The Goldilocks effect in infant auditory attention. *Child Development*, *85*(5), 1795–1804.

Kinney, D. K., & Kagan, J. (1976). Infant attention to auditory discrepancy. *Child Development*, *47*, 155–164.

Loewenstein, G. (1994). The psychology of curiosity: A review and reinterpretation. *Psychological Bulletin*, *116*, 75–98.

Oudeyer, P.-Y., Baranes, A., & Kaplan, F. (2013). Intrinsically motivated learning of real-world sensorimotor skills with developmental constraints. In *Intrinsically motivated learning in natural and artificial systems* (pp. 303–365). Springer.

Oudeyer, P.-Y., & Kaplan, F. (2007). What is intrinsic motivation? A typology of computational approaches. *Frontiers in Neurorobotics*, *1*(6).

Piaget, J. (1952). The origins of intelligence in children, *8*.

Schmidhuber, J. (2010). Formal theory of creativity, fun, and intrinsic motivation (1990 – 2010). *IEEE Transactions on Autonomous Mental Development*, *2*(3), 230–247. `http://doi.org/10.1109/TAMD.2010.2056368`

Stahl, A. E., & Feigenson, L. (2015). Observing the unexpected enhances infants' learning and exploration. *Science*, *348*(6230), 91–94.