

UC Merced

Proceedings of the Annual Meeting of the Cognitive Science Society

Title

Relations between number-knowledge and causal reasoning about number in young children: A preliminary investigation

Permalink

<https://escholarship.org/uc/item/6579v28m>

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 47(0)

Author

Sobel, David M.

Publication Date

2025

Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at <https://creativecommons.org/licenses/by/4.0/>

Peer reviewed

Relations between number-knowledge and causal reasoning about number in young children: A preliminary investigation

David M. Sobel (dave_sobel@brown.edu)

Department of Cognitive and Psychological Sciences, Brown University, 190 Thayer St. Providence RI 02912

Abstract

Three experiments investigated preschoolers' ability to infer that numbers can be causally efficacious. Preschoolers observed that one of two quantities of objects activated a machine (i.e., a container holding 2 blocks activated a machine while a container holding 3 did not). Children were asked to determine whether novel containers with either 2 or 3 objects would activate the machine, and then construct their own container of objects that would do so. Four-year-olds, but not 3-year-olds, were above chance at both tasks given a contrast between the numbers '2' and '3' (Study 1), but not as good when the contrast was between the numbers '4' and '6' (Study 2). The effect of age on understanding '2' was mediated when children's numerical knowledge was considered (Study 3). These results are interpreted in terms of children's causal reasoning and hypothesis-formation abilities, but also their developing knowledge of numbers.

Keywords: Causal reasoning; Conceptual Development; Numeracy

Introduction

By the time children enter elementary school, they have sophisticated causal reasoning abilities. Infants register causal relations in object perception (e.g., Leslie & Keeble, 1987), and can make inferences about unseen events and relations (e.g., Gweon & Schulz, 2011; Sobel & Kirkham, 2006). Causal inference becomes more explicit during the preschool years, during which children can engage in prediction and diagnosis (e.g., Bullock et al., 1982), explanation (e.g., Schult & Wellman, 1997), and counterfactual reasoning (e.g., Harris et al., 1996).

At the same time, children come to possess a great deal of knowledge about numbers. Newborns possess sophisticated capacities for representing quantities (e.g., Izard et al., 2009; Xu, 2003), which allows them to perform computations (e.g., Wynn, 1998). Children whose native language is English learn to count (i.e., they can recite number words in an ordered list, at least to ten) usually by age 2, but only afterwards do they learn the cardinal principles that underlie the relation between position on that list and quantity (Carey, 2009; Sarnecka & Carey, 2008). Between the ages of 18 months and 4, they progress through stages during which they only understand the meaning of the first few number words. That is, as a *subset knower*, young children understand the meaning of only a subset of small number words (usually *one* first, then *one* and *two*, and then *one*, *two*, and *three*, called *one-knowers*, *two-knowers*, and *three-knowers* respectively). Sometime around age 4, children learn that a number word's meaning reflects its position in the count list, and that

numerals have particular conceptual relations to one another. That is, children become *cardinal principle (CP) knowers* – they develop more abstract knowledge about the natural numbers, such as an understanding of succession (i.e., adding an object changes numerosity in a predictable way) and other relations among quantities (e.g., taking one away from 7 makes 6, see e.g., Gelman & Gallistel, 1978; Sarnecka & Carey, 2008; Wynn, 1992; see Carey, 2009, for a review).

Subset-knowers clearly have a different understanding of the meaning of number words than CP-knowers. But there are other differences between subset and CP knowers. For example, CP-knowers, but not subset-knowers, recognize relations between number words and quantities, including aspects of the semantic structure of numerical words as opposed to adjectives (e.g., Slusser & Sarnecka, 2011). CP-knowers are more likely than subset-knowers to recognize that when there is a transformation to a quantity of objects, the corresponding number word used to refer to that quantity changes (e.g., Condry & Spelke, 2008). If becoming a CP-knower allows children to “infer the meanings of all higher [*higher than 3*] number words through counting” (Negen et al., 2012, p. 58), then CP-knowers might apply a more abstract process than subset knowers when interpreting the number of objects that they observe.

Indeed, Wynn (1990, 1992) observed that CP-knowers will typically generate a single number word when asked how many objects are in a set they have already counted. Subset-knowers, in contrast, might recount. When CP knowers make mistakes in their counting, they often correct themselves while subset-knowers do not (LeCorre et al., 2006; see also Carey, 2009). LeCorre et al. also showed that CP-knowers, but not subset-knowers could make an inference about numeracy; for example, if a character specifically wanted a certain large quantity of objects (say 7), and another character counted out 6, 7, or 8 objects, CP-knowers would respond accurately when asked if the quantity in question was the appropriate value; subset-knowers were less accurate.

These findings point to a novel hypothesis, which is that CP-knowers, but not subset-knowers, might interpret numbers of objects in general as potential causes. That is, in the LeCorre et al study, when the character wants exactly seven objects, there is something causal about that quantity, but not others (i.e., there's a reason why the character wants 7, but not another number – what that reason is might be unknown, but the quantity itself has some efficacy). Here, I test this hypothesis by introducing children to a novel causal

system (a *blicket detector*, Gopnik & Sobel, 2000), which activates when a particular number of objects is placed on it.

In Study 1, three- and 4-year-olds observed that trays with either two objects or three objects activated a machine, and trays with the other quantity did not. They were then asked to infer whether novel containers of two or three objects would activate the machine, and then finally construct a container of objects that was efficacious. In this first study, in the demonstration, I also contrasted whether the objects were all the same size or if the container with two objects was clearly heavier than the container with three objects (by virtue of presenting two large objects vs. three small ones). I hypothesized that the older children would be more likely to generalize the number in the container as causally efficacious, given that more 4-year-olds than 3-year-olds would be CP-knowers. Study 2 replicated this procedure, contrasting containers with 4 vs. 6 objects on it, however given the results of Study 1, all the objects were the same size. Finally, Study 3 reinvestigates the procedure in Study 1, but also considers whether children are subset- or CP-knowers, with the hypothesis that, independent of age, the CP-knowers would be more likely to recognize the efficacy of a particular number.

Study 1

Three- and 4-year-olds were introduced to a novel machine that activated when a certain number of objects in a container was placed upon it. Children observed that containers with either two or three objects activated the machine (and that the other number did not). They were asked to generalize that efficacy to novel containers of two vs. three objects. To discourage children from thinking that weight or identity was a critical factor, the size of the objects in the novel containers varied, such that children saw objects that were either all the same size or two large objects and three small ones.

Participants

Forty-six 3-year-olds ($M=40.46$ months) and 42 four-year-olds ($M=54.14$ months) participated (41 girls and 47 boys). Seventy-six children were Caucasian, three children were of Asian descent, two children were Black, three children were Hispanic, one was of mixed descent (three did not respond). Four additional children were tested, but not included because of parental interference or experimental error.

Materials and Procedure

Children were introduced to a *blicket machine* (Gopnik & Sobel, 2000). The machine was 15x13x6cm, made of wood, painted grey, with a red Lucite top that depressed. The machine could light up and play music, controlled by the experimenter, when objects were placed on it.

Children also observed nine transparent plastic containers 9x9x13cm. Four contained 2 wooden blocks that were either identical or different; four contained 3 blocks that were either identical or all different, either in color/shape or in size. The final box contained five different blocks. Figure 1 shows an example of a pair of containers

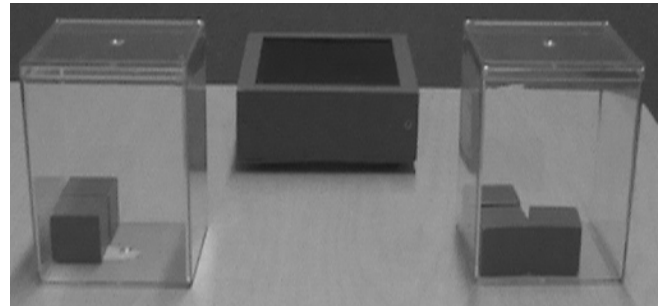


Figure 1. The machine and a pair of stimuli that were used in the procedure in which all objects were the same size.

After a brief warmup with an experimenter, children were introduced to the machine and told that some things made it go. They were shown a pair of containers – one with two objects and another with three. Children were randomly assigned to the ‘2’ or ‘3’ condition in which the container with that number of objects inside activated the machine and the other did not. Children were asked to confirm which container activated the machine. This familiarization was repeated. Children responded correctly 98% of the time and corrective feedback was given if they erred.

Children were then given generalization trials in which they were shown two novel containers, one with 2 objects and one with 3, and were asked which one made the machine go. For approximately half of the children ($N=48$), the objects in the box were roughly the same size as the familiarization objects. For the other children ($N=40$), the objects differed in size, such that the container had either two large objects or three small ones (or vice versa). Children were given two of these generalization trials.

After these trials, children were shown a final empty container and five objects, all of different sizes. The experimenter showed the child that when only one of the objects (selected at random) was in the container, the container did not activate the machine. That object was removed, and children were asked to make the machine go. They were given three opportunities to put objects in the container and test the container on the machine.

Results and Discussion

Raw data and analysis code are available at https://osf.io/p5vuy/?view_only=5e15a835bec943f3bba2221e9744be68. Overall, children generalized the appropriate number as efficacious 73% of the time. No difference was observed when the objects were all the same size or there was a conflict between the quantity and size, $\chi^2(2) = 0.45, p = .80$. This will not be considered further.

These data were analyzed via a generalized linear mixed model assuming a binomial logistic response, with age, numeracy, and trial number as fixed effect, and participant ID was the random effect. This model showed a significant effect of age, with children more likely to generalize the efficacious number as they got older, $\beta = 0.12, SE = 0.05, t = 2.56, p = .01$. Children responded at greater than chance

levels on each of the two test trials, Binomial tests, $p < .001$, but this was due to performance by the 4-year-olds (85% total), Binomial tests, both p -values $< .001$. The 3-year-olds (61% total) were not different from chance performance, Binomial tests, both p -values = .17.

Fifty-two percent of the children were able to activate the machine on the final trial by putting the appropriate number of objects in the container. A generalized linear model assuming a binomial logistic response was constructed with age and effective number (2 v. 3) as factors. The only significant factor was age, $\beta = 0.08$, $SE = 0.03$, Wald $\chi^2(1) = 7.63$, $p = .006$. However, number was a marginal trend in this analysis, with children more likely to activate the machine when it activated for 2 objects (61%) than three objects (43%), $\beta = 0.79$, $SE = 0.46$, Wald $\chi^2(1) = 2.95$, $p = .08$.

These data suggest that 4-year-olds, but not 3-year-olds can interpret a number of objects as a causal mechanism. However, there is some evidence that they are better at doing so for the number '2' at least when asked to construct that number to activate the machine. This suggests two questions. First, in Study 2, we replicated this procedure using larger numbers to consider the robustness of the effect. Second, in Study 3, we revisit performance in the '2' condition to examine whether possessing the cardinal principles serves as a potential mechanism for children's performance.

Study 2

Study 2 expanded on Study 1 to examine whether 4-year-olds could make similar inferences about larger numbers (4 and 6) instead of the smaller number used in Study 1. Only 4-year-olds were considered because of 3-year-olds chance performance on the generalization trials in Study 1. Because there was no effect of size in Study 1, all the objects were the same size in Study 2.

Participants

Twenty-four 4-year-olds ($M = 52.75$ months) participated. There were 14 girls and 10 boys. Nineteen children were Caucasian, one was Hispanic and four were of mixed descent.

Materials and Procedure

The same machine, containers, and similar blocks from Study 1 were used here, except that the containers contained either four or six blocks respectively. Because the size of the objects did not affect the results in Study 1, all the blocks were the same size. To ensure children could always detect the number of blocks in the containers, the containers were placed on their side, and the objects were spaced out, and glued to the bottom. Children responded correctly to the questions about which container in the demonstration activated the machine 100% of the time, so no corrective feedback was needed. The only other difference was that in the final trial, in which children had to construct the efficacious number, eight objects were presented to the children, instead of five.

Results and Discussion

Overall, children generalized the appropriate number as efficacious on 79% of the trials. When four objects (75%) or six objects (83%) were shown to be efficacious, children chose novel instances of that number to activate the machine greater than chance levels, both Binomial tests, $p = .02$ and $p = .002$ respectively. A GLMM analyzing age, number, and order of trials did not reveal any significant effects.

In contrast to Study 1, only 29% of children were able to reconstruct the efficacious number to activate the machine in the final trial. As in Study 1, we analyze these data with a generalized linear model assuming a binary response. No significant effects were found, but numeracy was a marginal trend, with children more likely to reconstruct four objects when 4 was efficacious (42%) than reconstruct six objects when 6 was efficacious (17%), $\beta = 0.99$, $SE = 0.58$, Wald $\chi^2(1) = 2.80$, $p = .09$.

It is important to note that the sample size in Study 2 was much smaller than in Study 1, making these findings preliminary. Combining the results of Study 1 and 2 together, we constructed a generalized linear model assuming a binary response with age and numeracy (both treated continuously) as independent variables. Age was a significant predictor, $\beta = 0.09$, $SE = 0.03$, Wald $\chi^2(1) = 9.50$, $p = .002$, as was numeracy, with performance dropping as the efficacious number was higher, $\beta = -0.64$, $SE = 0.16$, Wald $\chi^2(1) = 15.83$, $p < .001$.

Taken with the results of Study 1, these data show that 4-year-olds can identify which of the two numbers presented to children is efficacious from a small set of examples. However, as the number got larger, their ability to construct a novel example of that numeracy decreased.

I also analyzed the errors children made when they responded incorrectly on the final trial (i.e., when they failed to generate a particular efficacious number). Errors fell into four types looking across Studies 1-2. Children sometimes placed the container on the machine with no objects inside it (3%). They were more likely to use one unique object (17%) or all the objects (70%) across the attempts. Only 10% of the time did children construct a numeracy close to the efficacious numeracy, but not the exact one (e.g., 2 or 4 when the effective numeracy was 3). When children erred on the final trial, they mostly seemed to use a hypothesis other than number. Given the high frequency of children using all objects or one individual object, I speculate that weight or individual objects having efficacy were obvious alternates.

What I have not considered is whether children's specific numerical knowledge (as opposed to age, which has served as a stand-in) related to their ability to infer that the number of objects on the machine was efficacious. In Study 3, I investigated a sample of children slightly younger than in Studies 1-2 to consider whether knowing the cardinal principles (as opposed to age alone) predicted their inferences. I used '2' as the efficacious number because in Study 1, children seemed most capable of responding correctly for this number.

Study 3

Study 3 reproduced the procedure used in Study 1, examining children's ability to identify and reconstruct '2' as an efficacious number. In addition to reproducing these data, children were given the Give-N task (Wynn, 1992) to determine whether they were a subset-knower or a CP-knower. The prediction was that CP knowers, regardless of age, would be more likely to construct the efficacious number (i.e., 2 objects) when asked to activate the machine.

Because the hypothesis was that number knowledge (and not age or general cognition) was underlying children's ability to infer that the number of objects on the machine resulted in its efficacy, I wanted to present children with another measure of cognition that would not be thought relevant to this relation, but that does show clear development during this age. Children thus also received an unexpected content false belief measure, which clearly develops during this age range (Wellman et al., 2001).

Participants

Fifty-one children between the ages of 32-54 months ($M = 43.16$ months, 22 girls and 29 boys) participated in this study. Forty-two children were Caucasian, four was Hispanic, two were Black, one was Asian, and two were from the Pacific Islands. Eleven additional children were tested, but not included in the final analysis. Eight failed an inclusion measure, one was excluded because of parent interference and two because of experimental error.

Materials and Procedure

The materials and procedure from Study 1 were used here. Children also saw 3 sets of 8-10 plastic animal figurines (bears, horses, and fish) in the Give-N task. Ten white plastic chips were used in the counting task, and a *Crayola* crayon box that contained birthday candles was used in the unexpected contents task.

Children were introduced to two experimenters. One of them administered a procedure similar to Study 1 in which children first observed that containers with two objects inside activated the machine and containers with three objects did not. They were then given three generalization trials and then the same trial in which they were given three attempts to activate the machine with any two objects.

Children participated in three other measures administered by the second experimenter. First, ten chips were placed on the table and children were asked to count them. Children who could not generate all numerals between 1-10 were not included in the final analysis. Eight children were excluded for this reason. Second, children were given an unexpected contents false belief task, modeled after Gopnik and Astington (1988). Children were shown a crayon box and asked what they believed was inside. All children responded "crayons" (or a synonym). The experimenter then revealed that candles were actually in the box. The candles were put back in the box. Children were asked what another person

would think is in the box, what they themselves previously thought was in the box, and what was really in the box.

Finally, children were administered the "Give a Number" task, based on Wynn (1992). Three sets of toy animals (bears, horses, and fish – the latter referred to as "fishes" when a plural form was necessary) were placed on the table. Children were asked to give the experimenter a particular number between one and six of one kind of toy (e.g., "give me two horses"). If children successfully gave N toys on a trial, the experimenter requested $N + 1$ toys (of a different category) on the next trial. If they failed to give N toys, the next trial asked for $N - 1$. Children were tested up to the smallest number they responded incorrectly to on at least two out of three tries or six, whichever came first.

The order in which children received the sets of tasks from the experimenters was randomized. Each experimenter was ignorant of the results of the other's session.

Results and Discussion

The machine procedure was scored similarly to Study 1. Children received a point if they stated that another person would have a false belief that the box contained crayons and if they themselves falsely believed that the box contained crayons, resulting in a score between 0-2. However, if children failed to say that candles were really in the box, they were assigned a score of 0 on the unexpected contents task. Following Wynn's (1992) coding scheme, children were divided into subset-knowers and CP-knowers. A subset knower for a particular value (N) had to (1) give N objects on at least two out of three trials when asked for N objects and (2) not give N objects when another number was requested on at least half the trials. Thus, if a child always gave one and two objects appropriately, but then always gave three objects for any other number they would not be credited with knowing three, and would be categorized as a two-knower. Children who responded correctly on at least two of three trials for all numbers were categorized as CP-Knowers.

There were three trials in which children were asked to generalize whether a container with two or three objects would activate the machine. Children chose the container with two objects on 59% of the trials, not significantly different from chance values, $\chi^2(2) = 6.78, p = .08$. These data were analyzed with a generalized linear model, assuming an ordinal response, with age and whether children were a subset- or CP-knower as independent variables. This model revealed a significant effect only of age, with children more likely to choose the container with two blocks in it as they got older, $\beta = 0.12, SE = 0.06, \text{Wald } \chi^2(1) = 4.31, p = .04$. Subset knowers ($N=33$) did not significantly differ from CP-knowers ($N=18$), 54% vs. 70%. However, although the subset knowers did not significantly differ from chance performance, $\chi^2(2) = 1.34, p = .72$, the CP-knowers did exceed chance performance, $\chi^2(2) = 18.15, p < .001$.

Finally, we considered whether children could construct the efficacious number when asked to activate the machine. Overall, 45% of children could do so within the three allotted attempts. We analyzed these data with a generalized linear

model assuming a binary response, with age, performance on the unexpected contents task, and number-knower level (subset-knower vs. CP-knower) as factors. Age and false belief scores were not significant. Number-knower-level was a marginal trend, $\beta = 1.14$, $SE = 0.67$, Wald $\chi^2(1) = 2.84$, $p = .09$, with CP-knowers (66%) more likely to activate the machine than subset-knowers (33%).

I ran another analysis, breaking down the subset knowers into two separate groups: one-knowers vs. two/three-knowers (based on Wynn's scoring system). When the same generalized linear model was constructed with this multinomial factor (one-knowers vs. other subset-knowers vs. CP-knowers), the effects of age and false belief were again not significant, but one-knowers were significantly less likely to activate the machine (17%) than CP-knowers, $\beta = -2.02$, $SE = 0.99$, Wald $\chi^2(1) = 4.14$, $p = .04$. Other subset-knowers (43%) did not differ from CP-knowers, $\beta = -0.82$, $SE = 0.71$, Wald $\chi^2(1) = 1.33$, $p = .25$, but also did not significantly differ from one-knowers, $\beta = 1.20$, $SE = 0.92$, Wald $\chi^2(1) = 1.68$, $p = .20$.

When shown that a container with two objects, but not three objects would activate a novel machine, subset knowers and CP-knowers interpreted the mechanism that governed the machine's activation somewhat differently. CP-knowers were more likely to generalize that efficacy to novel instances of '2' than chance expectations, however, they did not differ in overall level of performance from that of the subset knowers. Subset knowers, however, responded no differently from chance. When children were asked to construct a container that activated the machine, the difference between subset- and CP-knowers was only a marginal trend. However, when CP-knowers were broken down into one-knowers vs. other subset knowers vs. CP knowers, one-knowers were less likely to generate the effective number than CP-knowers.

Although one-knowers presumably could tell that the instances of '2' and '3' were different numbers, they did not seem able to register that number was potentially responsible for causing the machine to activate. A potential difference between the one-knowers and the other subset-knowers might stem from the latter's ability to not only label two objects verbally as "two", but also notice that other representations of '2' share that same verbal label, consistent with 'data-driven' inferential mechanisms in which labels act similarly to perceptual features as the basis for generalization (e.g., Sloutsky & Fisher, 2004). These subset-knowers, however, did not often construct a representation of '2' in order to activate the machine, even with several attempts. They could not register that number could be the potential mechanism that activates the machine. CP-knowers, in contrast, might understand that numbers are interrelated concepts (e.g., Sarnecka & Carey, 2008). They registered the effective number of objects as what would activate the machine above chance levels. They were also more likely to construct that exact number to activate the machine given a set of objects. This suggests that they can hypothesize that a specific number was caused the machine to activate.

These data suggest that the observed differences between 3- and 4-year-olds in Study 1 were not the result of simple age-related change. Instead, the results were potentially due to more 4-year-olds being CP-knowers and more 3-year-olds being subset knowers (and some, presumably, one-knowers). This suggests a longitudinal follow-up experiment: As children's knower-level changes, their ability to hypothesize number as a causal mechanism should change as well.

General Discussion

Three- and four-year-olds were presented with a causal system in which placing a container with a particular number of objects inside it would activate a novel machine. Study 1 and 2 showed that children could identify particular numbers as efficacious above chance levels when the number was between 2 and 6. As the number got higher, however, children's ability to construct a container that activated the machine diminished. Study 3 examined children's ability to construct the smallest number under consideration – '2' – contrasting performance among children with different number-knower types. Controlling for age, one-knowers were poor at reconstructing the number '2' as efficacious, and were significantly worse than CP-knowers.

Interestingly, children could make these inferences based on a relatively short demonstration and small amount of data. Because the way number was represented in these experiments correlated with the weight, size, and volume of the objects, any one of these features could be thought of as valid mechanism for the machine's activation. However, the 4-year-olds in Studies 1-2 (who presumably were more likely to be CP-knowers given the results of Study 3) as well as the CP-knowers in Study 3 seemed to extract *the number of objects on the machine* as opposed to these other features from the demonstration phase of the procedure. The best evidence for this comes from Study 1, in which there was no difference between performance when the size of the objects was varied and Study 3 in which children's number-knower-level related to their responses. Further studies are needed to consider the different hypotheses children consider and how they update those hypotheses given the data they observe.

Fundamentally, however, number-knower-level is thought to relate to how children count objects and understand the meaning of number words. The finding here suggests that CP-knowers recognize something more abstract about the concept of *number* itself, and that it can be a potential mechanism for a causal system. Performance on the Give-N task correlates with some measures of Spontaneous Focus on Number (*SFoN*; Savelkoul et al., 2020). It might be that the present study indicates children's emerging ability to focus on numbers more broadly than previously thought and form hypotheses about them. Given that the extent to which children engage in *SFoN* has relations to their academic achievement in mathematics (Hannula et al., 2010), appreciating numbers as causal might have relations to other facets of children's academic achievement or engagement. Future studies on this topic are necessary to isolate the exact

role of number-knower-level or other cognitive capacity on children's causal inferences.

There are also significant limitations to the present study. The most important one is that the samples used here were samples of convenience, collected without considering power. This makes all studies presented here preliminary and all analyses presented here exploratory. Moving forward, this project needs to be reproduced and preregistered. Attention to power, for example, will discern whether the findings in Study 3 generalize between subset- and CP-knowers or between the more fine-grained analysis of one-knowers, other subset-knowers and CP-knowers. Extending Study 3 to a case where the number '3' is efficacious in Study 3 would hypothesize that both one and two-knowers should perform poorly. Adding SFoN measures and refining how number of tested as a causal mechanism would further the understanding of potential relations between children's numerical knowledge and causal inference capacities.

There are also limits to the extent to which these studies investigated children's ability to reason abstractly. For instance, one could imagine scenarios in which quantities had different causal efficacy – such as that particular quantities predicted the loudness of the machine, or the duration the machine activated. Understanding such abstraction potentially has its origin in infancy (e.g., deHevia & Spelke, 2010), and some relational inferences might be available to children in infancy (e.g., Walker & Gopnik, 2014). Mapping out the exact relation between children's developing domain-specific knowledge and their use of that knowledge for more domain-general causal inferences is an open question for further studies.

Acknowledgments

This study was supported by NSF (0518161 and 2300459). I would like to thank Hannah Boettcher, David Buchanan, Karina Ikesoe, Caroline Kleeman, Deanna Macris, Rachel Shelley-Abrahamson, Cesalie Stepney, and Katherine Williams for help with participant recruitment, data collection, and analysis.

References

- Antell, S., & Keating, D. P. (1983). Perception of numerical invariance in neonates. *Child Development*, 54, 695-701.
- Bullock, M., Gelman, R., & Baillargeon, R. (1982). The development of causal reasoning. In W. J. Friedman (Ed.), *The developmental psychology of time* (pp. 209-254). New York: Academic Press.
- Carey, S. (2009). *The origins of concepts*. New York: Oxford.
- Condry, K. F., & Spelke, E. S. (2008). The development of language and abstract concepts: The case of natural numbers. *Journal of Experimental Psychology: General*, 137, 22-38.
- de Hevia, M. D., & Spelke, E. S. (2010). Number-space mapping in human infants. *Psychological science*, 21(5), 653-660.
- Gelman, R., & Gallistel, C. R. (1978). *The child's understanding of number*. Cambridge, MA: Harvard University Press.
- Gopnik, A., & Astington, J. W. (1988). Children's understanding of representational change and its relation to the understanding of false belief and the appearance-reality distinction. *Child Development*, 59, 26-37.
- Gopnik A., & Sobel, D. M. (2000). Detecting blickets: How young children use information about causal powers in categorization and induction. *Child Development*, 71, 1205-1222.
- Hannula, M. M., Lepola, J., & Lehtinen, E. (2010). Spontaneous focusing on numerosity as a domain-specific predictor of arithmetical skills. *Journal of Experimental Child Psychology*, 107, 394-406.
- Harris, P. L., German, T., & Mills, P. (1996). Children's use of counterfactual thinking in causal reasoning. *Cognition*, 61, 233-259.
- Leslie, A. M., & Keeble, S. (1987). Do six-month-old infants perceive causality? *Cognition*, 25, 265-288.
- Negen, J., Sarnecka, B. W., & Lee, M. D. (2012). An Excel sheet for inferring children's number-knower levels from give-N data. *Behavior Research Methods*, 44, 57-66.
- Sarnecka, B. W., & Carey, S. (2008). How counting represents number: What children must learn and when they learn it. *Cognition*, 108, 662-674.
- Savelkoul, S., Hurst, M. A., & Cordes, S. (2020). Preschoolers' number knowledge relates to spontaneous focusing on number for small, but not large, sets. *Developmental Psychology*, 56, 1879-1893.
- Schult, C. A., & Wellman, H. M. (1997). Explaining human movements and actions: Children's understanding of the limits of explanation. *Cognition*, 62, 291-324.
- Sloutsky, V. M., & Fisher, A. V. (2004). Induction and categorization in young children: A similarity-based model. *Journal of Experimental Psychology: General*, 133, 166-188.
- Slusser, E. & Sarnecka, B.W. (2011). Find the picture of eight turtles: A link between children's counting and their knowledge of number-word semantics. *Journal of Experimental Child Psychology*, 110, 38-51.
- Sobel, D. M., & Kirkham, N. Z. (2006). Blickets and babies: The development of causal reasoning in toddlers and infants. *Developmental Psychology*, 42, 1103-1115.
- Walker, C. M., & Gopnik, A. (2014). Toddlers infer higher-order relational principles in causal learning. *Psychological Science*, 25, 161-169.
- Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-analysis of theory-of-mind development: The truth about false belief. *Child Development*, 72, 655-684.
- Wynn, K. (1992). Children's acquisition of the number words and the counting system. *Cognitive Psychology*, 24, 220-251.
- Xu, F., & Spelke, E. S. (2000). Large number discrimination in 6-month-old infants. *Cognition*, 74, B1-B11.