



## UvA-DARE (Digital Academic Repository)

### Preschoolers perform more informative experiments after observing theory-violating evidence

van Schijndel, T.J.P.; Visser, I.; van Bers, B.M.C.W.; Raijmakers, M.E.J.

**DOI**

[10.1016/j.jecp.2014.11.008](https://doi.org/10.1016/j.jecp.2014.11.008)

**Publication date**

2015

**Document Version**

Final published version

**Published in**

Journal of Experimental Child Psychology

**License**

Article 25fa Dutch Copyright Act (<https://www.openaccess.nl/en/policies/open-access-in-dutch-copyright-law-taverne-amendment>)

[Link to publication](#)

**Citation for published version (APA):**

van Schijndel, T. J. P., Visser, I., van Bers, B. M. C. W., & Raijmakers, M. E. J. (2015). Preschoolers perform more informative experiments after observing theory-violating evidence. *Journal of Experimental Child Psychology*, 131, 104-119. <https://doi.org/10.1016/j.jecp.2014.11.008>

**General rights**

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

**Disclaimer/Complaints regulations**

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, P.O. Box 19185, 1000 GD Amsterdam, The Netherlands. You will be contacted as soon as possible.

*UvA-DARE is a service provided by the library of the University of Amsterdam (<https://dare.uva.nl>)*



Contents lists available at ScienceDirect

# Journal of Experimental Child Psychology

journal homepage: [www.elsevier.com/locate/jecp](http://www.elsevier.com/locate/jecp)



## Preschoolers perform more informative experiments after observing theory-violating evidence



Tessa J.P. van Schijndel\*, Ingmar Visser, Bianca M.C.W. van Bers, Maartje E.J. Raijmakers

Department of Psychology, University of Amsterdam, 1018 XA Amsterdam, The Netherlands

### ARTICLE INFO

#### Article history:

Received 22 July 2014

Revised 24 November 2014

Available online 24 December 2014

#### Keywords:

Naive theories

Theory-violating evidence

Exploratory play

Ecologically valid domain

Preschoolers

Latent variable technique

### ABSTRACT

This study investigated the effect of evidence conflicting with preschoolers' naive theory on the patterns of their free exploratory play. The domain of shadow size was used—a relatively complex, ecologically valid domain that allows for reliable assessment of children's knowledge. Results showed that all children who observed conflicting evidence performed an unconfounded informative experiment in the beginning of their play, compared with half of the children who observed confirming evidence. Mainly, these experiments were directed at investigating a dimension that was at the core of children's initial theory. Thus, preschoolers were flexible in the type of experiments they performed, but they were less flexible in the content of their investigations.

© 2014 Elsevier Inc. All rights reserved.

### Introduction

The Piagetian claim that young children construct knowledge by active exploration has been accepted widely (e.g., Singer, Golinkoff, & Hirsh-Pasek, 2006). The claim implies that young children are capable of integrating observed evidence with prior knowledge to formulate hypotheses, designing experiments, and drawing conclusions that enable learning. This process requires the use of substantive domain-specific knowledge as well as formal knowledge—general abilities that allow

\* Corresponding author.

E-mail address: [t.j.p.vanschijndel@uva.nl](mailto:t.j.p.vanschijndel@uva.nl) (T.J.P. van Schijndel).

for translating hypotheses into effective experiments and drawing conclusions from these experiments (Gopnik, Sobel, Schulz, & Glymour, 2001; Gopnik et al., 2004).

Recently, several studies have provided empirical evidence for preschoolers' possession of such formal knowledge (see Schulz, 2012, for a review). These studies demonstrate rationality and systematicity in preschoolers' exploration. Researchers have looked at how characteristics of evidence affect children's exploratory play, showing that uncertainty about the causal structure of an event promotes preschoolers' exploration (Bonawitz, Van Schijndel, Friel, & Schulz, 2012; Cook, Goodman, & Schulz, 2011; Gweon & Schulz, 2008; Legare, 2012; Legare, Gelman, & Wellman, 2010; Schulz & Bonawitz, 2007). Researchers have proposed that such findings are consistent with a Bayesian inference framework (Bonawitz et al., 2012; Cook et al., 2011; Schulz, 2012).<sup>1</sup> A specific case of uncertainty arises when children's theories conflict with the evidence children observe (e.g., Berlyne, 1960; Chinn & Brewer, 1993). Legare et al. (2010) and Legare (2012) showed that this type of evidence affects preschoolers' explanatory reasoning, which in turn was shown to be related to their exploratory play. Bonawitz et al. (2012) demonstrated that this type of evidence affects the duration of young children's exploratory play. They assessed 6- and 7-year-olds' prior knowledge in the domain of balance and classified children as having a center theory (objects balance on their geometrical center) or a mass theory (objects balance on their center of mass). Children were then confronted with evidence that either confirmed or conflicted with their balancing theory, and those who observed conflicting evidence played longer with a balancing toy than children who observed confirming evidence.

Several studies have shown that it is the patterns of children's exploration, rather than the time spent exploring, that determine opportunities for learning (e.g., Bonawitz et al., 2012; Gweon & Schulz, 2008; Schulz, Gopnik, & Glymour, 2007). Testing children's ability to use exploratory play for learning, therefore, implies not only a demonstration of children selectively exploring after observing conflicting evidence (Bonawitz et al., 2012) but also a demonstration of children selectively performing specific patterns of exploration after observing this type of evidence. Cook and colleagues (2011) investigated these patterns in the situation where children observe ambiguous evidence, that is, evidence that is ambiguous with respect to which variable controls the effect. However, to our knowledge, these patterns have not been investigated in the situation where children observe conflicting evidence. The goal of the current study, therefore, was to investigate the effect of evidence conflicting with preschoolers' naive theory on the patterns of their free exploratory play.

Preschoolers' patterns of exploration have been quantified in different ways such as by looking at the variability or objectives of children's actions (e.g., Legare, 2012; Sobel & Sommerville, 2010). For example, Legare (2012) used the blicket detector paradigm—a machine that activates (lights up and plays music) when some objects (blickets), but not others, are placed on it (Gopnik & Sobel, 2000; Nazzi & Gopnik, 2000). Legare then coded the objectives of actions by looking at hypothesis-testing strategies that children used to investigate conflicting evidence. Two such strategies are switching locations of object pairs on the machine and trying to open an object. Variability in actions was then coded by looking at how many of these different strategies children employed. In the current study, we focused on children's use of unconfounded informative experiments, that is, experiments from which valid causal conclusions can be drawn. Performing unconfounded experiments is a domain-general skill that is at the core of scientific practice; therefore, the learning of the skill is considered to be of importance in the development of scientific reasoning (e.g., Chen & Klahr, 1999; Klahr & Nigam, 2004). In science education research, the skill of designing unconfounded experiments is called the *control of variables strategy*, and several studies have demonstrated that primary school-aged children have difficulty with the use and transfer of the strategy (e.g., Chen & Klahr, 1999; Klahr & Nigam, 2004; Kuhn, Garcia-Mila, Zohar, & Andersen, 1995). These findings stand in contrast to studies in the field of developmental psychology showing young children's ability to perform unconfounded experiments (e.g., Cook et al., 2011; Schulz et al., 2007; Sobel & Sommerville, 2010). The current study investigated whether theory-violating evidence leads preschoolers to selectively perform

<sup>1</sup> The Bayesian inference framework provides a formal account of how children's prior theories interact with observed evidence to affect exploratory play. Specifically, Bayesian inference specifies how children update their beliefs on a hypothesis given the observed data. An explanation of the account is beyond the scope of this article. See Bonawitz and colleagues (2012), Cook and colleagues (2011), and Schulz (2012) for more information.

unconfounded experiments during the course of free play. The results could possibly provide evidence not only for young children's ability to use unconfounded experiments but also for young children to do so in a situation where knowledge is to be gained because of theory-violating evidence.

The majority of studies investigating preschoolers' exploratory play have been carried out in novel artificial domains (e.g., Cook et al., 2011; Gweon & Schulz, 2008; Legare, 2012; Schulz & Bonawitz, 2007; Sobel & Sommerville, 2010), for example, using the blicket detector paradigm (Gopnik & Sobel, 2000; Nazzi & Gopnik, 2000). These domains enable researchers to control children's prior knowledge of the causal relations between events. However, the representation of knowledge that children acquire in an artificial domain over a brief time span is expected to differ from the representation of knowledge that children have acquired throughout their daily lives. Because children's theories were at the core of the current study, we chose to use an ecologically valid domain guaranteeing children's theories being real preexisting ideas that they acquired prior to the experiment. We used the domain of shadow size (Chen, 2009; Ebersbach & Resing, 2007; Feher & Rice, 1988; Fleeer, 1996; Howe, Tolmie, Duchak-Tanner, & Rattray, 2000; Inhelder & Piaget, 1958; Segal & Cosgrove, 1993; Siegler, 1978, 1981). Compared with many of the above-mentioned artificial domains, this is a relatively complex domain; it entails two interacting causal factors. Shadow size is proportional to the size of an object (size dimension) and inversely proportional to the distance of an object to the light source (distance dimension). Performing unconfounded experiments in this domain, therefore, implies keeping one of the variables constant while applying variation to the other one.

A second reason for selecting this domain was that previous studies consistently demonstrated the existence among preschoolers of a specific naive theory in this domain—children taking into account the size dimension, but not the distance dimension, in determining shadow size. Siegler (1981) first distinguished this theory group and described these children as applying "Rule 1." Several follow-up studies on children's knowledge on shadow size confirmed an increase with age in children's tendency to take into account the subordinate distance dimension in determining shadow size (e.g., Chen, 2009; Ebersbach & Resing, 2007).

The relative complexity of the selected domain and the results of previous work in this domain (Chen, 2009; Ebersbach & Resing, 2007; Siegler, 1981) made it possible to assess children's naive theories by applying a combination of Siegler's (1976, 1981) *rule assessment methodology* and a latent variable technique (e.g., McCutcheon, 1987; Rindskopf, 1987). This approach is considered to lead to a reliable assessment because it relies on children's nonverbal responses, allows detection of both anticipated and unanticipated theories, and does not require the researcher to set an arbitrary criterion for correspondence between observed and expected responses to classify children into theory groups (Van der Maas & Straatemeier, 2008). Together, these characteristics ensure that the theory assessment resulting in the approach is data driven and less influenced by the researcher's anticipations regarding the existence of certain naive theories or the researcher's choice regarding correspondence criteria. The approach is described in the Appendix. For discussion, see Van der Maas and Straatemeier (2008).

To summarize, what is new in this study is that we examined children's patterns of exploratory play in the situation where they observe conflicting evidence. This is a key factor in understanding the process through which children acquire knowledge from exploratory play. A second defining characteristic of the study is the use of an ecologically valid domain in combination with a reliable assessment of children's naive theories. In short, we hypothesized that children who are confronted with conflicting evidence will perform more unconfounded informative experiments during free play than children who are confronted with confirming evidence. That is, in the conflicting condition more often than in the confirming condition, children who believe that only the size of an object affects the size of a shadow (Rule 1 children) will perform experiments in which they vary one of the two relevant dimensions—size or distance. This prediction was driven by the fact that a child who believes that only the size of an object affects shadow size could logically explain theory-violating evidence in one of two ways. First, the child could hypothesize that small objects give larger shadows than large objects instead of the other way around. Second, the child could hypothesize that not only size but also the distance of objects to the light source affects shadow size. Effectively testing either one of these hypotheses necessarily involves the design of unconfounded experiments. In addition to studying children's patterns of play, we investigated children's learning from evidence and play. Because we did not

have clear hypotheses on children's knowledge acquisition, we did this in an exploratory manner as opposed to a confirmatory one.

## Method

### Participants

The total sample before classification into theory groups consisted of 102 4- to 9-year-olds (45 boys and 57 girls,  $M_{\text{age}} = 66.07$  months,  $SD = 15.56$ ) who were recruited from two primary schools. An additional 12 children were recruited but not included in the analyses; of these, eight children were excluded because an error was made in administering the pre- or post-task (seven children pushed the light button during the pre- or post-task and got feedback, and for one child the test leader did not time the free play episode) and four children were excluded because no complete video-recordings of the free play episode were available. The sample characteristics were chosen to allow for an optimal discrimination of a Rule 1 group. Even though we expected to find Rule 1 children mainly in the preschool age range, the sample's age range was taken wider to guarantee sufficient power for reliably classifying children into theory groups. When using a latent variable technique to detect subgroups of children, the subgroups need to be large enough to be separated from each other. In the preschool age range, we expected to find only a small group of children using a more advanced rule; therefore, we included older children in the sample to ensure that we could detect this advanced rule group (see [Appendix](#)). This way, we avoided preschoolers having an advanced rule being incorrectly assigned to the Rule 1 group. In Results, we describe the Rule 1 sample after classification into theory groups.

### Materials

The shadow machine, the setup of the shadow task ([Inhelder & Piaget, 1958](#); [Siegler, 1978, 1981](#)), was used for all four phases of the experiment. The machine consisted of two light sources, a screen placed 50 cm from the light sources, and puppets that could be placed between the light sources and screen (see [Fig. 1](#)). When a button was pressed, the lights were activated (they stayed lit as long as the button was held) and shadows of the puppets were portrayed on the screen. There were two small puppets ( $7.5 \times 2.25$  cm) and two large puppets ( $10 \times 3$  cm) that could be placed at three distances from the light sources (10, 20, and 30 cm). Relative shadow size depended on both the size of the object and the distance from the object to the light sources (the distance from the light sources to the screen was kept constant).

### Procedure

Children were tested individually by one of two experimenters in a private room at their school. The child and experimenter sat at the same side of a table facing the shadow machine. The child was first introduced to the machine and then participated in four experimental phases: pre-task, evidence exposure, free play episode, and post-task. The total experiment took approximately 20 min.



**Fig. 1.** The shadow machine. During test administration for the study described in this article, no people but the child and the experimenter were present. Photography: Hanne Nijhuis.

The experiment was designed for Rule 1 children; the evidence either conflicted with or confirmed the Rule 1 group's theory. The responses on the pre-task were used to select the Rule 1 group. However, because the final classification into theory groups was performed after data collection on the basis of the results of the latent variable technique (see [Appendix](#)), all children were administered all phases of the experiment. Importantly, we report only the results of the Rule 1 group in this article because no clear hypotheses could be formulated about the effect of evidence on the other theory group's play and learning.

### Introduction to the shadow machine

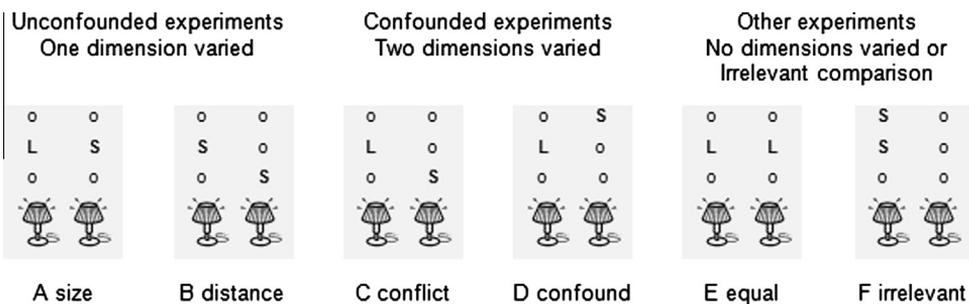
The experimenter introduced the shadow machine by pointing out the light sources and the puppets of different sizes and by demonstrating how the puppets could be placed close to or farther away from the light sources. She then demonstrated how to make the shadows. She placed two equally sized puppets at equal distances from the light sources, pushed the light button, and said, "Do you see the shadows? This one [pointing to the left shadow] is equally big as this one [pointing to the right shadow]. They are the same."

### Pre-task

The experimenter introduced the pre-task by saying, "Now we are going to play a game. Each time I will put puppets in place. You then say whether you think that the shadow on this side will be the biggest [pointing to the left side of the screen], the shadow on this side will be the biggest [pointing to the right side of the screen], or that they will be the same." She then administered 12 items to the child: six size items, in which the size of the puppets was varied but the distance from the puppets to the light sources was kept constant (see [Fig. 2A](#)), and six distance items, in which the distance from the puppets to the light sources was varied but the size of the puppets was kept constant (see [Fig. 2B](#)). The items were administered in one of two fixed semi-random orders. For each item, the experimenter put two puppets in place and said, "I put this puppet here and this puppet here. When I make the shadows, which one will be the biggest? This one [pointing to the left side of the screen], this one [pointing to the right side of the screen], or will they be the same?" Importantly, during the pre-task, the child did not see shadows and, therefore, did not get any feedback. Responses on the pre-task were scored trichotomously: correct, incorrect "the same," or incorrect not "the same." A latent variable technique was used to determine children's naive theories on the basis of these trichotomous responses. This procedure is explained in the [Appendix](#).

### Evidence exposure

To enable random assignment (stratified by age and sex) of Rule 1 children to the evidence conditions (conflicting and confirming condition), the pre-task scores were used to perform an



**Fig. 2.** Examples of different types of items and experiments on the shadow machine. For the pre- and post-tasks, size items (A) and distance items (B) were used (without feedback/shadows). For the evidence exposure, a conflict item (C) was used for the conflicting condition and a confound item (D) was used for the confirming condition (with feedback/shadows). For the free play episode, the experiments that children performed were assigned to the same categories or in the categories of equal items (E) or irrelevant items (F). These figures present examples of the different types of experiments (L = large puppet, S = small puppet), although other variations are possible. Irrelevant experiments also included experiments with one, three, or four puppets.

ad hoc classification into a group possibly having Rule 1 and a group not having Rule 1. This ad hoc classification was based on previous literature (Siegler, 1981); children were assigned Rule 1 if they had answered more than 4 size items correctly and more than four distance items incorrectly with “the same.”

In the conflicting condition, the experimenter placed a small puppet close to the light source (10 cm) and a large puppet farther away from the light source (20 cm) (see Fig. 2C). As in the pre-task, she asked, “I put this puppet here and this puppet here. When I make the shadows, which one will be the biggest? This one [pointing to the left side of the screen], this one [pointing to the right side of the screen], or will they be the same?” (prediction evidence item). A child using Rule 1 was expected to predict that the large puppet would have the biggest shadow, which was not the case in this condition. Next, the experimenter showed the child the shadows by pushing the light button. To make sure that the child paid attention, she asked, “Do you see the shadows? Which one is the biggest? This one [pointing to the left side of the screen], this one [pointing to the right side of the screen], or are they the same?” (observation evidence item). The confirming condition was similar to the conflicting condition except that the experimenter placed a small puppet farther away from the light source (30 cm) and a large puppet close to the light source (20 cm) (see Fig. 2D). A child using Rule 1 was expected to predict that the large puppet would have the biggest shadow, which was also the case in this condition.

#### *Free play episode*

During the free play episode, the child was encouraged to engage in free play with the shadow machine for 5 min. The experimenter sat in a corner of the room out of the child's sight so that she did not influence or disturb the child. Video-recordings were made, and all experiments that children performed were scored by a coder who was blind to the conditions. An experiment was defined as putting one or more puppets in place and pushing the light button. For each performed experiment, a code was noted corresponding to the experiment's unique combination of puppet(s) and location(s). Experiments were coded as unconfounded when one dimension (size or distance) was varied and the other was kept constant (see Fig. 2A and B: size and distance experiments, 18 variations including left–right reversals). Other experiments included those in which both dimensions were varied (see Fig. 2C and D: conflict and confound experiments, 12 variations including left–right reversals), no dimensions were varied (see Fig. 2E: equal experiments, six variations), and an irrelevant comparison was made such as by putting two puppets in place at the same side of the machine or by using one, three, or four puppets (see Fig. 2F: irrelevant experiments). Ignoring the irrelevant experiments, the number of possible experiments in which one dimension was varied (unconfounded experiments) was equal to the number of possible experiments in which two or no dimensions were varied (confounded plus equal experiments); therefore, a child would have a .50 chance of performing an unconfounded experiment for each given trial. A second coder, also blind to the conditions, coded the experiments performed by 19 children (19%) again, and this double-coding rendered a percentage agreement of 96%, corresponding to a kappa of .95.

Because children continuously generated evidence during free play, the effect of the observed evidence was expected to be most prominent in the beginning of the play episode. Therefore, the play analyses were performed on the first five trials, that is, the first five experiments children performed after evidence exposure.

#### *Post-task*

The post-task consisted of four items: two size items (see Fig. 2A) and two distance items (see Fig. 2B). Because a maximum of six size items could be constructed with this version of the shadow task, the size items in the post-task were repetitions of items that had been administered in the pre-task. The distance items had not been used in the pre-task. The items were administered in a fixed semi-random order. As during the pre-task, the child did not see shadows and, therefore, did not get any feedback. The items were scored trichotomously: correct, incorrect “the same,” or incorrect not “the same.” Again, a latent variable technique was used to determine children's naive theories on the basis of these trichotomous responses (see Appendix).

## Results

### Assessment of naive theories

To assess children's naive theories on shadow size, latent class analysis was performed on the pre-task data (see Appendix for the procedure and detailed results of the analysis). As expected, one of the distinguished classes was interpreted as applying Rule 1 ( $n = 39$  4- to 6-year-olds,  $M_{\text{age}} = 56.82$  - months,  $SD = 7.29$ ). The children in this Rule 1 group took into account the size dimension but not the distance dimension; they had high probabilities of giving correct responses on size items [ $p(\text{correct}) = .89$ ] and answering "the same" on distance items [ $p(\text{"the same"}) = .88$ ].

### Effect of theory-violating evidence on patterns of play

Of the 39 children in the Rule 1 group, four were removed from the play analyses because they performed fewer than five experiments during free play (see Method), leaving 15 children in the conflicting condition ( $M_{\text{age}} = 56.53$  months,  $SD = 7.49$ ) and 20 children in the confirming condition ( $M_{\text{age}} = 56.40$  months,  $SD = 6.86$ ). The unequal number of participants in the conditions was caused by the fact that the final classification based on the latent class analysis differed from the ad hoc classification made during task administration (see Method). A small proportion of children in the conflicting condition (7%) correctly predicted the evidence item (see Method), whereas a large proportion of children in the confirming condition did (70%, Fisher exact,  $n = 24$ ,  $p < .0001$ ). Children in both conditions tended to report correctly about the observed evidence (conflicting condition 92%, confirming condition 100%; see Method).

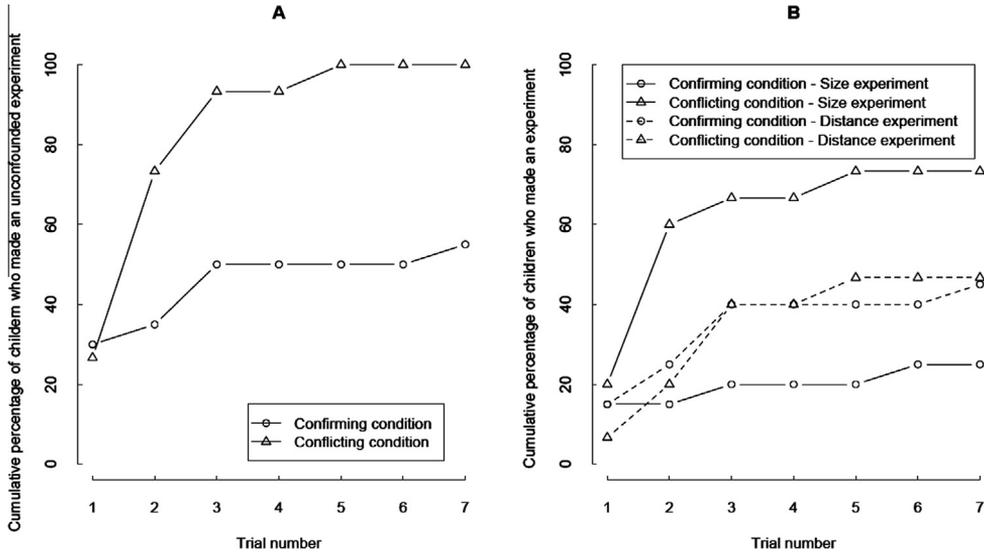
In line with our hypothesis, a difference between the conditions was found in patterns of play; fully 100% of children in the conflicting condition performed at least one unconfounded experiment within the first five trials (first five experiments that children performed after evidence exposure), compared with 50% of children in the confirming condition (Fisher exact,  $n = 35$ ,  $p < .01$ ) (see Table 1). The odds of children performing an unconfounded experiment within the first five trials was 31 times higher in the conflicting condition than in the confirming condition (calculated with a zero-cell correction of adding 0.5 to each cell). This result held up when performing the analysis over the first four trials (Fisher exact,  $n = 35$ ,  $p < .01$ ) or first six trials (Fisher exact,  $n = 35$ ,  $p < .01$ ), implying that the rather arbitrary choice to focus on the first five trials did not influence the results (see Fig. 3A). In addition, the result also held up when removing from the analysis children who did not predict the evidence item in line with their theory (children in the conflicting condition who did not predict the evidence item incorrectly and children in the confirming condition who did not predict the evidence item correctly) and children who reported incorrectly about the observed evidence (Fisher exact,  $n = 24$ ,  $p < .01$ ).

Unconfounded experiments can be aimed either at investigating the size dimension (size experiments: varying the size dimension while keeping the distance dimension constant; see Fig. 2A) or at investigating the distance dimension (distance experiments: varying the distance dimension while keeping the size dimension constant; see Fig. 2B). A difference between the conditions was found in children's tendency to perform size experiments; nearly three quarters (73%) of the children in the

**Table 1**

Percentages of Rule 1 children ( $n = 35$ ) in the different conditions performing at least one experiment of Types A to F (see Fig. 2) during the first five trials.

	Unconfounded experiments			Confounded experiments			Other experiments	
	A size	B distance	Total unconfounded	C conflict	D confound	Total confounded	E equal	F irrelevant
Confirming condition	20	40	50	59	30	70	80	50
Conflicting condition	73.3	46.7	100	46.7	20.0	66.7	80.0	73.3



**Fig. 3.** Cumulative percentages of children who made at least one unconfounded experiment (A) or at least one size or distance experiment (B) during the first seven trials (first seven experiments that children performed after evidence exposure).

conflicting condition performed at least one size experiment within their first five trials, compared with 20% of the children in the confirming condition (Fisher exact,  $n = 35$ ,  $p < .01$ ) (see Table 1). This result held up when performing the analysis over the first four trials (Fisher exact,  $n = 35$ ,  $p < .05$ ) or first six trials (Fisher exact,  $n = 35$ ,  $p < .01$ ) (see Fig. 3B), implying that the differences between conditions emerged relatively early after evidence observation. No difference between the conditions was found in children's tendency to perform distance experiments or in their tendencies to perform conflict, confound, equal, or irrelevant experiments.

### Learning from evidence and play

To assess children's naive theories after play, latent class analysis was also performed on the post-task data (see Appendix for the procedure and detailed results of the analysis). Comparing children's classifications pre- and post-play, it was found that 77% ( $n = 30$ ) of the children who applied Rule 1 on the pre-task applied the same rule on the post-task, 13% ( $n = 5$ ) reverted to guessing, and 10% ( $n = 4$ ) applied a more advanced rule in which they took into account both dimensions.

First, we analyzed the Rule 1 group's learning at the individual level by distinguishing between children who did and did not have a more advanced theory post versus pre. Learning was found to be unrelated to condition, and to children's tendency to perform unconfounded experiments during play. When looking at size and distance experiments separately, we found learning to be unrelated to children's tendency to perform size experiments during play, but it was related to children's tendency to perform distance experiments during play; more than one quarter (27%) of the children who made a distance experiment within the first five trials learned, whereas none of the children who did not make a distance experiment within the first five trials did (Fisher exact,  $n = 35$ ,  $p < .05$ ). Again, this result held up when performing the analysis over the first four trials (Fisher exact,  $n = 35$ ,  $p < .05$ ) or first six trials (Fisher exact,  $n = 35$ ,  $p < .01$ ).

However, because in the learning analyses on the individual level learning was defined in terms of going to a more advanced class, the analyses did not allow us to determine how the accuracy within item types (size and distance items) changed from pre- to post-task for the different conditions and whether this was related to children's play. Therefore, we also analyzed the Rule 1 group's learning at the group level. An analysis of variance (ANOVA) was conducted on the difference scores

(post–pre) of the proportion size items correct with condition (conflicting or confirming) and children's tendency to perform a size experiment (child did or did not perform a size experiment within the first five trials) as between-participants factors. One significant effect was found, that of condition,  $F(1, 31) = 5.17, p < .05$ , partial  $\eta^2 = .14$ ); on average, the performance on size items of children in the conflicting condition deteriorated over play, whereas the performance of children in the confirming condition improved over play (conflicting condition  $M = -.31, SD = .38$ ; confirming condition  $M = .07, SD = .36$ ). Next, an ANOVA was conducted on the difference scores (post–pre) of the proportion distance items correct with condition (conflicting or confirming) and children's tendency to perform a distance experiment (child did or did not perform a distance experiment within the first five trials) as between-participants factors. No effects were found.

## Discussion

This study investigated the effect of evidence conflicting with preschoolers' naive theory on the patterns of their free exploratory play. In line with our expectations, we found that *all* children who were confronted with conflicting evidence performed an unconfounded informative experiment in the beginning of their play, whereas only half of the children who were confronted with confirming evidence did so. This result connects previous findings on young children's exploratory play. Bonawitz and colleagues (2012) showed that uncertainty generated by theory-violating evidence increases the duration of young children's play. It was also shown that young children are capable of performing informative experiments (e.g., Cook et al., 2011; Schulz et al., 2007; Sobel & Sommerville, 2010). The current study demonstrates that uncertainty generated by theory-violating evidence leads preschoolers to selectively generate these informative interventions, thereby providing an important explanatory demonstration of how children's behavior might support learning.

A defining characteristic of the current study is the domain in which the study was performed. First, in contrast to the majority of studies on preschoolers' exploratory play (e.g., Cook et al., 2011; Gweon & Schulz, 2008; Legare, 2012; Schulz & Bonawitz, 2007; Sobel & Sommerville, 2010), an ecologically valid domain was used. In artificial domains, causes and effects are constructed by the experimenter, but the used domain of shadow size has real causal properties. For example, when children would explore effects of object size on shadow size in the classroom, the effects would hold. Exploratory behavior is a central component in preschool science programs (e.g., French, 2004; Gelman & Brennenman, 2004), and the choice for an ecologically valid domain increases the study's relevance for the practice of science education. Second, the domain of shadow size is relatively complex; it entails two interacting causal factors (size and distance dimension). The complexity of the domain is not an advantage in itself, but it does imply that performing unconfounded informative experiments in this domain requires a more complex compound of behaviors than performing such experiments in a less complex domain. A child performing an unconfounded experiment in this study needed to apply variation to one variable while actively keeping the other variable constant. Refraining from performing the latter action, keeping the other variable constant, would result in random variation, that is, confounded experiments. However, in studies with less complex domains, unconfounded experiments could be performed by solely performing the former action, applying variation to one variable. For example, such actions could consist of first pressing Button A and then pressing Button B or first putting a gear on Peg A and then putting a gear on Peg B (e.g., Cook et al., 2011; Schulz et al., 2007; Sobel & Sommerville, 2010). The relative complexity of performing informative experiments in the current study strengthens its results; the study not only shows that one instance of theory-violating evidence evokes children's curiosity and motivates them to explore but also shows that, despite its relative complexity, this motivation is translated into efficient behavior. Third, the choice for the domain of shadow size made it possible to reliably assess children's naive theories with a combination of Siegler's (1976, 1981) rule assessment methodology and a latent variable technique (e.g., McCutcheon, 1987; Rindskopf, 1987; see Appendix). As mentioned before, one of the advantages of this approach is that it makes it possible to detect unanticipated theories. This also happened in the current study; besides the expected groups, such as the Rule 1 group (children who take into account only the size dimension in determining shadow size), a group of children using Rule 2–reversed

(children who take into account the size dimension in the right direction but the distance dimension in the wrong direction in determining shadow size; see [Appendix](#)) was detected. Even though [Ebersbach and Resing's \(2007\)](#) results point in the direction of the possible existence of this group, to our knowledge the current study is the first to show the existence of this theory in a reliable manner. Hence, this finding validates the usefulness of the applied approach because rule assessment methodology in combination with pattern matching ([Siegler, 1981](#)) would not have provided this result without prior specification of this particular rule.

As mentioned in the Introduction, the finding that preschoolers, without any form of training, perform unconfounded informative experiments contrasts with research demonstrating older children's difficulty with acquiring the control of variables strategy (e.g., [Chen & Klahr, 1999](#); [Klahr & Nigam, 2004](#); [Kuhn et al., 1995](#)). There are several explanations for this discrepancy (see [Bonawitz et al., 2012](#), and [Cook et al., 2011](#), for more extended discussions). One explanation is that the domains used for older children's inquiry learning are generally far more complex than the domains used in research on preschoolers' exploratory play. In addition, the tasks administered to older children often ask for a metacognitive understanding of controlling variables, whereas the tasks administered to preschoolers do not. Regardless, the findings of the current study provide evidence not only for preschoolers' ability to use a skill that is at the core of science but also for them to do this in response to theory-violating evidence.

In the current study, unconfounded experiments could be aimed either at investigating the dominant size dimension or at investigating the subordinate distance dimension. Results showed that children in the conflicting condition performed more size experiments than children in the confirming condition. This finding suggests that children who believe that only the size of an object affects the size of a shadow (Rule 1 children) who were confronted with theory-violating evidence designed experiments to verify their theory instead of experiments to falsify their theory, as was also found by [Kuhn \(1989\)](#) and [Schauble \(1990\)](#). Evidently, observing conflicting information was not a direct incentive for children to test the hypothesis that the distance of the object to the light source is also a factor determining shadow size. It is possible that children intuitively controlled for other variables when investigating the size dimension even though they did not expect distance to affect shadow size. Thus, preschoolers proved to be flexible in the type of experiments they performed, but they were less flexible in the topic of their investigations. The results of children's learning follow from these play results; it was found that confronting children with conflicting evidence did not support their learning, and the group-level analysis even showed that children's performance on size items in the conflicting condition deteriorated over play. These findings tentatively suggest that some formal knowledge abilities that are necessary for the process of constructing knowledge by active exploration might develop earlier than others; in this study, preschoolers were capable of designing unconfounded informative experiments. However, they lacked the cognitive flexibility to change the topic of their investigations and did not always draw correct conclusions from the experiments they performed. One factor was found to be positively related to children's learning—their tendency to perform distance experiments. However, because children's play was not experimentally controlled in this study, no causal conclusions can be drawn about the relation between play and learning on the basis of this result.

Studies on young children's learning in other knowledge domains have also demonstrated children's difficulty with taking into account a second subordinate dimension. Training studies on the balance scale task demonstrated that for the Rule 1 group (children who take into account only the weight dimension in determining which side will go down), feedback alone was insufficient for theory revision, but increasing awareness of the relevance of the distance dimension led some children to change their theories ([Jansen, Raijmakers, & Visser, 2007](#); [Jansen & Van der Maas, 2001](#); [Siegler, 1976](#)). The literature on cognitive flexibility offers different theoretical explanations for young children's difficulty with switching attention between dimensions. Some of these explanations focus on children's failure to suppress attention to the initial dimension (e.g., [Kirkham, Cruess, & Diamond, 2003](#)), whereas other explanations focus on children's failure to activate the previously inhibited dimension (e.g., [Chevalier & Blaye, 2008](#); [Müller, Dick, Gela, Overton, & Zelazo, 2006](#)). Future work might be aimed at investigating the exact nature of children's lack of cognitive flexibility on tasks such

as the shadow task. This knowledge could be useful in developing educational approaches for encouraging children to widen the scope of their exploratory behavior.

## Acknowledgments

The research of Tessa van Schijndel is sponsored by an NWO–Aspasia grant and the Dutch Curious Minds program, which is supported by the Dutch Ministry of Education, Culture, and Science and the Platform Bèta Techniek. The research of Maartje Raijmakers is sponsored by an NWO–VIDI grant. We thank the primary schools for their cooperation. We thank Marjolein Hof and Juri Peters for providing their help in collecting the data, Wouter Weeda and Han van der Maas for giving statistical advice, Dorothy Mandell for making comments on earlier drafts on the article, Coos Hakvoort for providing technical support, and Elise Herrig for performing a language check.

## Appendix. Approach and results theory assessment

### *Approach theory assessment*

To assess children's naive theories in the domain of shadow size, we used [Siegler's \(1976, 1981\)](#) rule assessment methodology (RAM) combined with latent class analysis (LCA; e.g., [McCutcheon, 1987](#); [Rindskopf, 1987](#)). Previously, this approach has been used to assess children's naive theories on balance ([Boom, Hoijsink, & Kunnen, 2001](#); [Jansen & Van der Maas, 1997, 2001, 2002](#)), the earth ([Straatemeier, Van der Maas, & Jansen, 2008](#)), and fetal development ([Van Es, Van Schijndel, Franse, & Raijmakers, 2009](#)). Combining RAM with LCA has several advantages over combining RAM with pattern-matching techniques (e.g., [Siegler, 1981](#)). For example, the approach allows for the detection of both anticipated and unanticipated theories, it does not require the researcher to set an arbitrary criterion for correspondence between observed and expected responses to classify children in theory groups, and it uses model selection techniques that allow for an optimal decision between goodness-of-fit and parsimony of the model. See [Van der Maas and Straatemeier \(2008\)](#) for an elaborate discussion on the advantages of using LCA over pattern-matching techniques.

The core idea of Siegler's RAM is to select different item types in such a way that each naive theory corresponds to a qualitatively different pattern of responses on the item types. In the current study, we aimed at discriminating between children who did and did not use Rule 1, and for that purpose the use of two item types (size and distance items; see Method) was sufficient. Because children applying Rule 1 take into account the size dimension, but not the distance dimension, in determining shadow size, they were expected to answer the size items correctly and to answer the distance items with "the same." Besides children applying Rule 1, based on [Siegler's \(1981\)](#) results, we also expected to find a group applying an advanced rule taking into account both object size and the distance of an object to the light source in determining shadow size and a group responding by making guesses. The advanced group was expected to answer both the size and distance items correctly, whereas the guessing group was expected to show unsystematic responses to the size and distance items.

The statistical technique LCA can be used to describe categorical manifest data in terms of categorical latent classes. In the current study, the technique was used to describe children's response patterns to the pre- and post-task items (separately) in terms of latent classes, and these classes were subsequently interpreted as different rules or naive theories on shadow size. LCA was performed by fitting latent class models (LCMs) to the data by calculating maximum likelihood estimates of the model parameters with the package `depmixS4` ([Visser & Speekenbrink, 2010](#)) for the R statistical programming environment ([R Development Core Team, 2011](#)). An LCM is defined by the number of latent classes and two sets of parameters: the unconditional probabilities and the conditional probabilities. The number of latent classes indicates the number of distinguished rule or theory groups. Unconditional probabilities indicate the probability of belonging to a class; that is, they define the class sizes. Conditional probabilities indicate the probability of a specific response (correct, incorrect "the same," or incorrect not "the same") given membership of a specific class. These probabilities allow for interpreting the rule or theory that is represented by a specific class. In the current study, latent class mod-

els with an increasing number of classes (1, 2, 3, 4, and 5) were fit to the data of the pre- and post-task separately. Subsequently, the optimal LCM was selected, that is, the model with a relatively good fit but the fewest number of classes. To this end, the Bayesian information criterion (BIC; Schwarz, 1978) was used; the LCM with the lowest BIC was considered to be the best-fitting, most parsimonious model. After model selection, the optimal model was interpreted. The selected model was used to assign individual children to a latent class by calculating the posterior probabilities of class membership given their responses. Each child was assigned to the class that appeared to be most probable, that is, for which the child had the highest posterior probability.

### Results theory assessment pre-task

The pre-task items had been administered in one of two semi-random orders. Because no differences in number of size items correct, number of distance items correct, or total number of items correct were found between the orders, they were aggregated for further analyses. LCMs with 1, 2, 3, 4, and 5 classes were fit to the pre-task data—children’s trichotomous response patterns on 12 items. Table A1 shows the fit statistics for the different LCMs. Based on the BIC values, it was found that a 4-class model was the most parsimonious, best-fitting model for the pre-task data, indicating four different theories on shadow size or classes showing unsystematic responses to the items. Because reliability analyses confirmed our expectation that the six size items consistently measured the same construct (Cronbach’s alpha = .87), as did the six distance items (Cronbach’s alpha = .88), we put equality constraints on the response probabilities of the six size items, and similarly on the response probabilities of the six distance items, in all classes. There was no significant difference in goodness-of-fit between the constrained and unconstrained models [log likelihood ratio:  $\chi(82) = 85.62, p = .37$ ]. Therefore, the more parsimonious constrained model was selected for interpretation and further analyses.

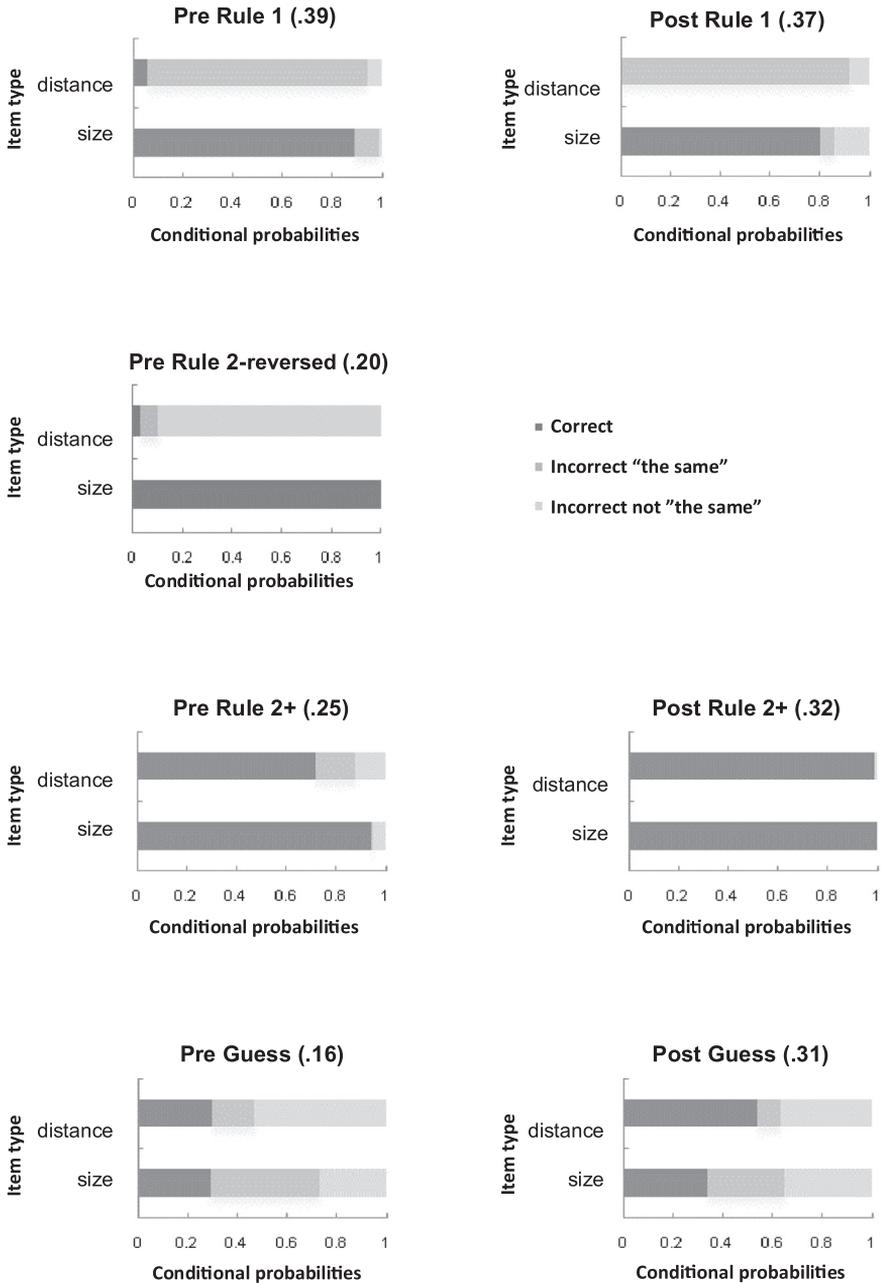
Fig. A1 shows the parameter estimates of the selected 4-class model. The first class was interpreted as applying Rule 1 ( $n = 39$ ). Children in this class took into account only the size dimension; they had a high probability of giving correct responses on size items and answering “the same” on distance items. The second class was interpreted as applying Rule 2–reversed ( $n = 20$ ). Children in this class took into account the size dimension in the right direction but the distance dimension in the wrong direction; they had a high probability of giving correct responses on size items and incorrect responses on distance items. The group tended to answer distance items by claiming that a puppet closer to the light source would give a smaller shadow than a puppet farther away from the light source. The third class was interpreted as applying Rule 2+ ( $n = 27$ ). Children in this class took into account both dimensions in the right direction; they had a high probability of giving correct responses on both size items and distance items. Because on the basis of the sole use of size and distance items children applying Rule 2 could not be distinguished from children applying a more advanced rule, the plus sign (Rule 2+) was used to indicate the possible use of a more advanced rule (see Siegler, 1981; Rule 2: the child makes a

**Table A1**

Fit statistics for different LCMs that were fit on the pre- and post-task data.

Number of classes	Pre-task			Post-task			
	L	df	BIC	Number of classes	L	df	BIC
1	-1001.86	24	2114.72	1	-359.98	8	756.95
2	-850.88	49	1928.39	2	-334.48	17	747.58
3	-761.60	74	1865.45	3	-288.16	26	696.56
4	-690.59	99	1839.05	3b	-298.59	14	661.92
4b	-733.40	17	1545.43	4	-288.12	35	724.31
5	-662.83	124	1899.15	5	-270.91	44	745.31

Note. L, log likelihood; df, degrees of freedom; BIC, Bayesian information criterion. The shaded rows show the fit statistics of the constrained 4-class model (pre-task) and constrained 3-class model (post-task) that were selected for interpretation.



**Fig. A1.** Parameter estimates for selected pre- and post-task models. Unconditional probabilities of the different classes are shown in parentheses. These probabilities are not necessarily exact copies of percentages of children assigned to a class.

decision based on size dimension when the puppets are unequal in size but decides on the basis on the distance dimension when the puppets are equal in size; Rule 3: the child takes into account both dimensions, but when one puppet is larger and the other puppet is closer to the light source, the child muddles through or guesses; Rule 4: the child takes into account both dimensions and computes

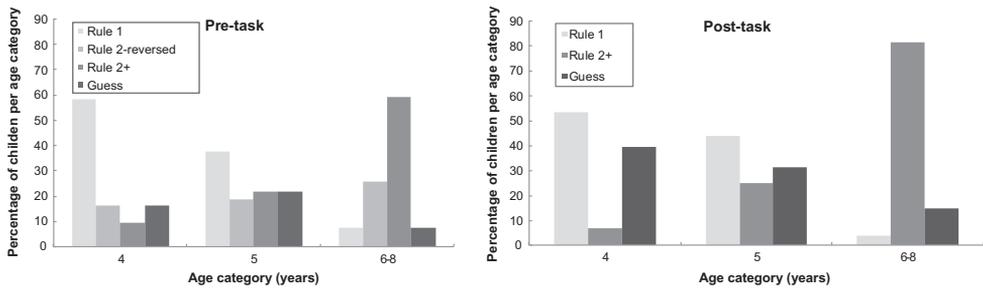


Fig. A2. Percentages of children per age group using the different rules on the pre- and post-tasks.

shadow sizes if necessary). The fourth class was interpreted as applying a guessing strategy ( $n = 16$ ). Children in this class did not show systematic responses to the size and distance items.

The detection of the Rule 1, Rule 2+, and guessing groups is in line with Siegler's (1981) results. However, by using a combination of RAM with LCA, we also detected an unanticipated group: the Rule 2-reversed group. Importantly, the approach used was not abundant for detecting the Rule 1 group as well. Using RAM and pattern matching (criterion 83.3%), only 62% of the children in this study would have been classified as having a (Rule 1 or Rule 2+) theory, leaving a rather substantial "guessing group"; specifically, seven children who were assigned to the Rule 1 group in the current study would have been assigned to the guessing group when using pattern matching.

Lastly, the most likely class membership was calculated for each child separately based on the posterior probabilities of his or her data given the selected 4-class model (e.g., McCutcheon, 1987; see above). It was demonstrated that rule use was age related,  $\chi(6) = 29.96$ ,  $p < .001$ . As shown in Fig. A2, the use of Rule 1 decreased with age, the use of Rule 2-reversed and Rule 2+ increased with age, and the guessing strategy was applied in all age categories. That is, with age, more children took into account the distance of an object to the light source in determining shadow size.

### Results theory assessment post-task

LCMs with 1, 2, 3, 4, and 5 classes were fit to the post-task data—children's trichotomous response patterns on four items. Table A1 shows the fit statistics for the different LCMs. Based on the BIC values, it was found that a constrained 3-class model was the most parsimonious, best-fitting model for the post-task data [log likelihood ratio constrained vs. unconstrained 3-class model:  $\chi(12) = 20.86$ ,  $p = .05$ ]; therefore, this model was selected for interpretation and further analyses. Fig. A1 shows the parameter estimates of the selected 3-class model. As in the pre-task model, one class could be interpreted as applying Rule 1 ( $n = 38$ ), one as applying Rule 2+ ( $n = 33$ ), and one as applying the guessing strategy ( $n = 31$ ). No Rule 2-reversed class was found in the post-task data. It was demonstrated that rule use was age related,  $\chi(4) = 44.15$ ,  $p < .001$ . As shown in Fig. A2, the use of Rule 1 and the guessing strategy decreased with age and the use of Rule 2+ increased with age. That is, with

Table A2

Learning: Numbers (and percentages) of children using specific combinations of rules on the pre- and post-tasks.

		Post-task			
		Rule 1	Rule 2+	Guessing strategy	Total
Pre-task	Rule 1	30 (77)	4 (10)	5 (13)	39 (100)
	Rule 2-reversed	3 (15)	8 (40)	9 (45)	20 (100)
	Rule 2+	2 (7)	20 (74)	5 (19)	27 (100)
	Guessing strategy	3 (19)	1 (6)	12 (75)	16 (100)
	Total	38 (37)	33 (32)	31 (31)	102 (100)

age, more children made systematic responses and more children took into account the distance of an object to the light source in determining shadow size.

To see whether children changed their theories over free play, we looked at the crosstabs of children's rule use on the pre- and post-task (see Table A2). Overall, children showed consistency in rule use (77% for Rule 1, 74% for Rule 2+, and 75% for the guessing strategy). However, because no Rule 2–reversed group was found in the post-task data, the children who had used this rule on the pre-task used either Rule 1 (15%), Rule 2+ (40%), or the guessing strategy (45%) on the post-task.

## References

- Berlyne, D. E. (1960). *Conflict, arousal, and curiosity*. New York: McGraw-Hill.
- Bonawitz, E. B., Van Schijndel, T. J. P., Friel, D., & Schulz, L. E. (2012). Children balance theories and evidence in exploration, explanation, and learning. *Cognitive Psychology*, *64*, 215–234.
- Boom, J., Hooijink, H., & Kunnen, S. (2001). Rules in the balance scale: Classes, strategies, or rules for the balance scale task?. *Cognitive Development*, *16*, 717–735.
- Chen, S. (2009). Shadows: Young Taiwanese children's views and understanding. *International Journal of Science Education*, *31*, 59–79.
- Chen, Z., & Klahr, D. (1999). All other things being equal: Acquisition and transfer of the control of variables strategy. *Child Development*, *70*, 1098–1120.
- Chevalier, N., & Blaye, A. (2008). Cognitive flexibility in preschoolers: The role of representation activation and maintenance. *Developmental Science*, *11*, 339–353.
- Chinn, C. A., & Brewer, W. F. (1993). The role of anomalous data in knowledge acquisition: A theoretical framework and implications for science instruction. *Review of Educational Research*, *63*, 1–49.
- Cook, C., Goodman, N. D., & Schulz, L. E. (2011). Where science starts: Spontaneous experiments in preschoolers' exploratory play. *Cognition*, *120*, 341–349.
- Ebersbach, M., & Resing, W. C. M. (2007). Shedding new light on an old problem: The estimation of shadow sizes in children and adults. *Journal of Experimental Child Psychology*, *97*, 265–285.
- Feher, E., & Rice, K. (1988). Shadows and anti-images: Children's conceptions of light and vision (part II). *Science Education*, *72*, 637–647.
- Fleer, M. (1996). Early learning about light: Mapping preschool children's thinking about light before, during, and after involvement in a two week teaching program. *International Journal of Science Education*, *18*, 819–836.
- French, L. (2004). Science as the center of a coherent, integrated early childhood curriculum. *Early Childhood Research Quarterly*, *19*, 150–158.
- Gelman, R., & Brenneman, K. (2004). Science learning pathways for young children. *Early Childhood Research Quarterly*, *19*, 138–149.
- Gopnik, A., Glymour, C., Sobel, D. M., Schulz, L. E., Kushnir, T., & Danks, D. (2004). A theory of causal learning in children: Causal maps and Bayes nets. *Psychological Review*, *111*, 3–32.
- Gopnik, A., & Sobel, D. M. (2000). Detecting blickets: How young children use information about novel causal powers in categorization and induction. *Child Development*, *71*, 1205–1222.
- Gopnik, A., Sobel, D. M., Schulz, L. E., & Glymour, C. (2001). Causal learning mechanisms in very young children: Two-, three-, and four-year-olds infer causal relations from patterns of variation and covariation. *Developmental Psychology*, *37*, 620–629.
- Gweon, H., & Schulz, L. E. (2008). Stretching to learn: Ambiguous evidence and variability in preschoolers' exploratory play. In B. C. Love, K. McRae, & V. M. Sloutsky (Eds.), *Proceedings of the 30th annual conference of the Cognitive Science Society* (pp. 570–574). Austin, TX: Cognitive Science Society.
- Howe, C., Tolmie, A., Duchak-Tanner, V., & Rattray, C. (2000). Hypothesis testing in science: Group consensus and the acquisition of conceptual and procedural knowledge. *Learning and Instruction*, *10*, 361–391.
- Inhelder, B., & Piaget, J. (1958). *The growth of logical thinking from childhood to adolescence*. New York: Basic Books.
- Jansen, B. R. J., Raijmakers, M. E. J., & Visser, I. (2007). Rule transition on the balance scale task: A case study in belief change. *Synthese*, *155*, 211–236.
- Jansen, B. R. J., & Van der Maas, H. L. J. (1997). Statistical test of the rule assessment methodology by latent class analysis. *Developmental Review*, *17*, 321–357.
- Jansen, B. R. J., & Van der Maas, H. L. J. (2001). Evidence for the phase transition from Rule I to Rule II on the balance scale task. *Developmental Review*, *21*, 450–494.
- Jansen, B. R. J., & Van der Maas, H. L. J. (2002). The development of children's rule use on the balance scale task. *Journal of Experimental Child Psychology*, *81*, 383–416.
- Kirkham, N. Z., Cruess, L., & Diamond, A. (2003). Helping children apply their knowledge to their behavior on a dimension-switching task. *Developmental Science*, *6*, 449–476.
- Klahr, D., & Nigam, M. (2004). The equivalence of learning paths in early science instruction: Effects of direct instruction and discovery learning. *Psychological Science*, *15*, 661–667.
- Kuhn, D. (1989). Children and adults as intuitive scientists. *Psychological Review*, *96*, 674–689.
- Kuhn, D., Garcia-Mila, M., Zohar, A., & Andersen, C. (1995). Strategies of knowledge acquisition. *Monographs of the Society for Research in Child Development*, *60* (4, Serial No. 245).
- Legare, C. H. (2012). Exploring explanation: Explaining inconsistent evidence informs exploratory, hypothesis-testing behavior in young children. *Child Development*, *83*, 173–185.
- Legare, C. H., Gelman, S. A., & Wellman, H. M. (2010). Inconsistency with prior knowledge triggers children's causal explanatory reasoning. *Child Development*, *81*, 929–944.
- McCutcheon, A. L. (1987). *Latent class analysis*. Newbury Park, CA: Sage.

- Müller, U., Dick, A. S., Gela, K., Overton, W. F., & Zelazo, P. D. (2006). The role of negative priming in preschoolers' flexible rule use on the Dimensional Change Card Sort task. *Child Development, 77*, 395–412.
- Nazzi, T., & Gopnik, A. (2000). A shift in children's use of perceptual and causal clues to categorization. *Developmental Science, 3*, 389–396.
- R Development Core Team, R. (2011). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Rindskopf, D. (1987). Using latent class analysis to test developmental models. *Developmental Review, 7*, 66–85.
- Schauble, L. (1990). Belief revision in children: The role of prior knowledge and strategies for generating evidence. *Journal of Experimental Child Psychology, 49*, 31–57.
- Schulz, L. E. (2012). The origins of inquiry: Inductive inference and exploration in early childhood. *Trends in Cognitive Sciences, 16*, 382–389.
- Schulz, L. E., & Bonawitz, E. B. (2007). Serious fun: Preschoolers engage in more exploratory play when evidence is confounded. *Developmental Psychology, 43*, 1045–1050.
- Schulz, L. E., Gopnik, A., & Glymour, C. (2007). Preschool children learn about causal structure from conditional interventions. *Developmental Science, 10*, 322–332.
- Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics, 6*, 461–464.
- Segal, G., & Cosgrove, M. (1993). "The sun is sleeping now": Early learning about light and shadows. *Research in Science Education, 23*, 176–285.
- Siegler, R. S. (1976). Three aspects of cognitive development. *Cognitive Psychology, 8*, 481–520.
- Siegler, R. S. (1981). Developmental sequences within and between concepts. *Monographs of the Society for Research in Child Development, 46* (2, Serial No. 189).
- Siegler, R. S. (1978). The origins of scientific reasoning. In R. S. Siegler (Ed.), *Children's thinking: What develops?* (pp. 109–149). Hillsdale, NJ: Lawrence Erlbaum.
- Singer, D. G., Golinkoff, M. R., & Hirsh-Pasek, K. (2006). *Play = learning: How play motivates and enhances children's cognitive and social-emotional growth*. New York: Oxford University Press.
- Sobel, D. M., & Sommerville, J. A. (2010). The importance of discovery in children's causal learning from interventions. *Frontiers in Developmental Psychology, 1*, 176–183.
- Straatemeier, M., Van der Maas, H. L. J., & Jansen, B. R. J. (2008). Children's knowledge of the earth: A new methodological and statistical approach. *Journal of Experimental Child Psychology, 100*, 276–296.
- Van der Maas, H. L. J., & Straatemeier, M. (2008). How to detect cognitive strategies: Commentary on "Differentiation and integration: Guiding principles for analyzing cognitive change". *Developmental Science, 11*, 449–453.
- Van Es, S. E., Van Schijndel, T. J. P., Franse, R. K., & Raijmakers, M. E. J. (2009). Children's thoughts on unborn babies, Representational redescription in preconceptions of children on fetal development. In N. A. Taatgen & H. van Rijn (Eds.), *Proceedings of the 31st annual conference of the Cognitive Science Society* (pp. 112–117). Austin, TX: Cognitive Science Society.
- Visser, I., & Speekenbrink, M. (2010). DepmixS4: An R-package for hidden Markov models. *Journal of Statistical Software, 36*(7), 1–21.