A developmental psychology perspective on preschool science learning: Children's exploratory play, naïve theories, and causal learning

van Schijndel, T.J.P.

Publication date
2012

CHAPTER 6

Young children perform better experiments after observing evidence conflicting with their prior knowledge

Van Schijndel, T. J. P., Visser, I., Van Bers, B. M. C. W., & Raijmakers, M. E. J. *Preschoolers perform better experiments after observing evidence conflicting with their prior knowledge.* Manuscript in revision.

ABSTRACT

This study investigates the role of young children's prior knowledge in guiding their exploration and learning. In the preschooler age group, we examine the effect of evidence conflicting with children's naïve theory on the quality of their exploratory play. In 4- to 9-year-olds, we study, in a more explorative manner, how children's naïve theories and quality of play are related to their causal learning. Children's naïve theories on shadow size are assessed by using a latent variable technique. Quality of play is determined by the number of unconfounded, informative experiments children perform during play. Results show an effect of conflicting evidence on the quality of play, and specific relationships between children's prior knowledge, quality of play, and learning.

INTRODUCTION

"Young children construct knowledge by active exploration." Though this Piagetian claim is widely accepted, there is little evidence demonstrating how this process takes place (Schulz & Bonawitz, 2007). 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: more general abilities that allow for translating hypotheses into effective experiments and drawing conclusions from those experiments (Gopnik, Sobel, Schulz & Glymour, 2001; Gopnik et al., 2004). Several studies on the primary school age group investigated how children's prior knowledge (substantive knowledge) and exploration (formal knowledge) interact during discovery learning (e.g. Dunbar & Klahr, 1989; Schauble, 1990). However, few studies on the preschooler age group took into account both these types of knowledge in investigating children's causal learning.

Studies on preschoolers' exploration have focused on effects of characteristics of evidence on the duration and quality of children's play. It was shown that children's exploratory play is affected by the novelty of evidence (e.g. Berlyne, 1960; Henderson & Moore, 1980) and the ambiguity of evidence (Gweon & Schulz, 2008; Schulz & Bonawitz, 2007). The interaction between children's prior knowledge and the evidence they observe has also been shown to influence their exploratory play (Bonawitz, Van Schijndel, Friel & Schulz, 2012; Cook, Goodman & Schulz, 2011; Legare, Gelman & Wellman, 2010; Legare, 2012). For example, Legare et al. (2010) and Legare (2012) demonstrated that evidence conflicting with 3- to 5-year-olds' prior knowledge affects their explanatory reasoning, which in turn guides their exploration.

The majority of studies on causal learning in preschoolers, however, have not focused on children's learning from their own play, but on children's learning from evidence supplied by adults (Bonawitz et al., 2012). One line of work investigates children's ability to draw causal inferences from patterns of evidence (e.g. Gopnik et al., 2001, 2004; Sobel & Kirkham, 2006;

Sobel, Tenenbaum & Gopnik, 2004), while another line of work investigates how children's causal learning is affected by the interaction between their prior knowledge and the evidence they observe (Schulz, Bonawitz, & Griffiths, 2007; Sobel et al., 2004). For example, Schulz, Bonawitz and Griffiths (2007) demonstrated that 5-year-olds, in contrast to 3- and 4-year-olds, are capable of learning from evidence conflicting with their prior knowledge.

The few studies that have investigated preschoolers' causal learning from exploratory play, demonstrated that this learning is possible, though not self-evident: learning only took place in a limited subset of children (Gweon & Schulz, 2008; Schulz, Gopnik & Glymour, 2007). Bonawitz et al. (2012) took into account children's prior knowledge in investigating their learning from play. The study was performed with an age group that was slightly older than the preschool age group: 6- and 7-year-olds. Children's prior knowledge in the domain of balance was assessed and children were classified 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 theory. Bonawitz et al. (2012) found that children who were confronted with evidence conflicting with their balancing theory played longer with the balancing toy than children who were confronted with evidence that confirmed their theory. After play children were asked a final question to see whether their beliefs had changed. The authors found that children who were confronted with conflicting evidence were more likely to change their beliefs than children who were confronted with confirming evidence. In addition, as in Gweon and Schulz's (2008) study, children's learning was dependent on the actions they performed during play. When an explanatory auxiliary variable (magnet) was present that could explain the evidence, the minority (44%) of children who initially believed that an object balances on its geometrical center learned that it actually balances on its center of mass after seeing conflicting evidence. This learning was related to generating evidence during play that could not be explained by the magnet. When no magnet was present, the majority (63%) of the center-theorists changed their beliefs after seeing conflicting evidence.

The present study further investigates the role of young children's prior knowledge in guiding their exploratory play and learning. Previous work demonstrated that evidence conflicting with young children's prior knowledge affects the duration of their exploratory play (Bonawitz et al., 2012). This study examines whether evidence conflicting with children's prior knowledge also affects the quality of children's play. Primarily, this is investigated in the preschooler age group. As some studies have demonstrated effects of prior knowledge (e.g. Schulz, Bonawitz & Griffiths, 2007; Sobel et al, 2004) and the quality of play (Bonawitz et al., 2012; Gweon and Schulz, 2008) on young children's learning, a second goal of the present study is to examine in more detail the relationship between these factors and children's learning. For example, we distinguish multiple types of prior knowledge and investigate, in an explorative manner, how these types relate to children's knowledge acquisition through play. To make sure there is sufficient variation in prior knowledge, we answer this research question by looking at a broad age group: 4- to 9-year-olds.

Some studies on young children's exploration and causal learning have been carried out in ecologically valid contexts, in which children presumably already have some prior knowledge (e.g. Bonawitz et al., 2012; Schulz, Bonawitz & Griffiths, 2007). These studies rely on a large body of research on young children's knowledge in a variety of areas, such as astronomy, biology, physics, and psychology (e.g. Flavell, Green & Flavell, 1995; Gelman & Wellman, 1991; Gopnik & Meltzoff, 1997; Hatano & Inagaki, 1994; Inagaki & Hatano, 1993; Jansen & Van der Maas, 2002; Kalish, 1996; Siegler, 1981; Straatemeier, Van der Maas & Jansen, 2008; Wellman & Gelman, 1992). Other studies have used an artificial novel context in order to experimentally control children's prior knowledge (e.g. Cook et al., 2011; Legare, 2012; Legare et al., 2010; Sobel et al., 2004). However, as the representation of knowledge that children acquire in an artificial context over a brief time span is expected to differ from that of the knowledge children have acquired throughout their daily lives, in the current study we choose to investigate young children's exploration and learning in an ecologically valid context: the domain of shadow size (e.g. Chen 2009; Ebersbach & Resing, 2007; Feher & Rice, 1988; Fleer, 1996; Howe, Tolmie, Duchak-Tanner & Rattray, 2000; Inhelder & Piaget, 1958; Segal & Cosgrove, 1993; Siegler, 1978, 1981). There is discussion of whether children's knowledge in ecologically valid contexts is coherent, theory like, or fragmented (see DiSessa, Gillespie, & Esterly, 2004 for a review). We deal with this issue in a methodological way by using an advanced statistical technique to show that children's knowledge on shadow size is coherent (see Van der Maas & Straatemeier, 2008). Therefore, in the remainder of this paper we will refer to this (prior) knowledge as children's naïve theories, or rules.

The shadow task was developed to investigate children's naïve theories on shadow size (Inhelder & Piaget, 1958; Siegler, 1978, 1981). The set up of the task, the shadow machine (see Figure 1), consists of two light sources, a screen placed at a fixed distance of these light sources, and objects of different sizes that can be placed at different locations in between the light sources and the screen. For each item the experimenter puts two objects in place and asks the child which of two shadows would be the biggest. Siegler (1981) assessed the naive theories, or rules, 3- to 12-year-olds use on this task. Besides a group of children, mostly 3- and 4-year-olds, that did not use a rule, Siegler distinguished four different rule groups. Children using Rule 1, mostly 4- and 5-year-olds, took into account the size dimension, but not the distance dimension in determining shadow size. Children using Rule 2, mostly 8-year-olds, based their judgments on the size dimension, but in addition considered the distance dimension if the object sizes were equal. Children using Rule 3 or 4, mostly 12-year-olds, always considered both dimensions in determining shadow size. When both dimensions differed, children using Rule 4, in contrast to children using rule 3, knew how to integrate both dimensions. Several follow-up studies on children's knowledge on shadow size have confirmed in 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).

There is considerable debate on the optimal method to assess children's rules on proportional reasoning tasks, such as the shadow task (Van der Maas & Straatemeier, 2008). Siegler (1976, 1981) used the Rule Assessment Methodology, which makes use of pattern matching (see Siegler, 1981). A line of studies on the balance scale task, another proportional reasoning task (Boom, Hoijtink, & Kunnen, 2001; Jansen & Van der Maas 1997, 2001, 2002), have proposed an alternative for pattern matching: Latent Class Analysis (LCA; see Method section for a brief introduction, see e.g. Bouwmeester & Sijtsma, 2007; McCutcheon, 1987; Rindskopf, 1987 for more information). One of the advantages of LCA over matching observed to expected response patterns is that the technique makes it possible to detect unanticipated response patterns, or rules (Van der Maas & Straatemeier, 2008). This advantage was shown in research with the balance scale task: by applying LCA to children's response patterns additional rules to those proposed by Siegler (1981) were found (Boom, Hoijtink & Kunnen, 2001; Jansen & Van der Maas, 1997). Moreover, LCA results in a more reliable classification of response patterns than pattern matching techniques (Van der Maas & Straatemeier, 2008). Hence, we start the current study by re-investigating children's naïve theories on shadow size by means of LCA.

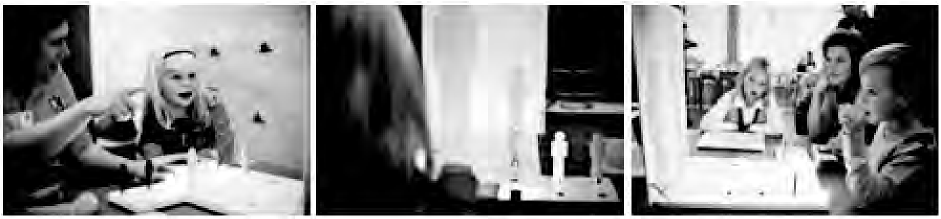

FIGURE 1. The shadow machine. During test administration for the study described in this paper, no people but the child and the experimenter were present. Photography: Hanne Nijhuis.

In the present study, 4- to 9-year-olds participated in an experiment with the shadow machine, aimed at answering three research questions: 1) What naïve theories do children have on shadow size? 2) Does evidence conflicting with children's naïve theory affect the quality of their exploratory play? 3) Do children learn from exploratory play? And is this learning related to a) children's naïve theory and b) the quality of their exploratory play? The experiment consisted of four phases: a pre-task, evidence exposure, free play episode, and post-task.

The pre-task was used to determine children's naïve theories on shadow size. In line with Siegler's (1976, 1981) Rule Assessment Methodology, children were administered a series of different item types: size items in which the size dimension is varied but the distance dimension is kept constant (see Figure 2A), and distance items in which the distance dimension is varied but the size dimension is kept constant (see Figure 2B). These items were selected to differentiate between a number of possible rules children could use. For example, children using Rule 1 were expected to answer the size items correctly, but to say “the same” on the distance items. Children using Rule 2 were expected to answer both the size- and distance items correctly. Latent Class Analysis was used to determine the different types of response patterns children demonstrated on the series. A priori we expected to find three groups: a group without systematic responses (Guess), a group only taking into account the size dimension (Rule 1), and a group taking into account both dimensions (Rule 2+).

The evidence exposure confronted children with evidence that was either conflicting with or confirming a specific naïve theory: Rule 1. In the conflicting condition children saw a small object having a bigger shadow than a large object. In the confirming condition children saw a small object having a smaller shadow than a large object. We decided to design the evidence for the Rule 1 group, because compared to the Rule 2+ group, which was expected to contain children using Rule 2 but also children using more advanced rules (see Siegler, 1981), we expected the Rule 1 group to be a relatively homogeneous group. This made it possible to construct evidence that would be conflicting with or confirming the predictions of the large majority of children classified as having Rule 1.

During the free play episode, children engaged in free exploratory play with the shadow machine. The quality of children’s play was determined by assessing how often during play they performed unconfounded experiments: experiments in which only one variable was varied. This type of experiments makes it possible to draw valid causal inferences. Previous work in an artificial context demonstrated that, in situations where information is to be gained, preschoolers selectively perform unconfounded actions (Cook et al., 2011). The current study was set up to investigate whether this is also the case in an ecologically valid context. Primarily, we investigated this in the Rule 1 group. Rule 1 children in the conflicting condition were expected to perform more unconfounded experiments during free play than Rule 1 children in the confirming condition. This prediction was driven by the fact that a child applying Rule 1 could logically explain conflicting evidence in one of two ways: 1) by assuming to be wrong about the direction of the effect of object size or 2) by assuming that, besides object size, there was another factor determining shadow size, distance being the most plausible one. Effectively testing either one of these hypotheses would imply the design of unconfounded experiments.

The post-task was used to assess children’s learning. Children’s naïve theories on shadow size were determined in a similar manner as with the pre-task. Learning was investigated on the individual level by inspecting whether children’s naïve theories on shadow size had changed over play, but also on the group level by testing whether the average proportions correct on the size- and distance items had improved over play.

METHOD

Participants

The final sample consisted of 102 children: 43 4-year-olds (*M*=53.81 months, *SD*=5.06), 32 5-year-olds (*M*=64.84 months, *SD*=3.87) and 27 6- to 9-year-olds (*M*=87.04 months, *SD*=13.21) that were recruited from two primary schools.[1] Twelve other children were recruited but not included in the analyses: 8 children were excluded because an error was made in administering the pre- or post-task (7 children pushed the light button during the pre- or post-task and got feedback) and 4 children were excluded because no complete video-recordings of the free play episode were available. Although most children were from White, middle-class backgrounds, a range of ethnicities reflecting the diversity of the population was represented.

Materials

The shadow machine, the set-up 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 Figure 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. Puppets of two sizes were used: two small ones measuring 7.5 x 2.25 cm and two large ones measuring 10 x 3 cm. The puppets 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 source (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. Subsequently the child participated in a pre-task, evidence exposure, free play episode, and post-task. The experiment took approximately 20 minutes per child.

Introduction to the shadow machine. The experimenter introduced the shadow machine by pointing out the light sources, the puppets of different sizes and by demonstrating how the puppets could be placed close to or further 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 left shadow) is equally big as this one (pointing to 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 left side of the screen), the shadow on this side will be the biggest (pointing to 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 Figure 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 Figure 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".

To enable random assignment of children over the evidence conditions per rule group, the pre-task scores were used to temporarily classify each child as using a specific rule. The temporary classification was based on previous literature (Siegler, 1981) and a pilot study. For example, children were assigned to the Rule 1 group when they had answered more than 4 size items correctly and more than 4 distance items incorrectly with "the same". To answer the research questions, a more advanced classification based on the results of the Latent Class Analysis was used.

Evidence exposure. Children per rule group were randomly assigned to the evidence conditions stratified on the basis of age and sex. In the conflicting condition children were shown evidence that conflicted with Rule 1, in the confirming condition children were shown evidence that confirmed Rule 1. It is important to notice that not only children using Rule 1, but all children that participated in the experiment were exposed to the evidence that was designed for the Rule 1 group, implying that the conflicting evidence was not necessarily conflicting and the confirming evidence was not necessarily confirming for children using |rules other than Rule 1. In the Results section we go further into this.

In the conflicting condition the experimenter placed a small puppet close to the light source (10 cm) and a large puppet further away from the light source (20 cm; see Figure 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?" (evidence-item prediction). 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 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?” (evidence-item observation). The confirming condition was similar to the conflicting condition, except that the experimenter placed a small puppet further away from the light source (30 cm) and a large puppet close to the light source (20 cm; see Figure 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 five minutes. 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 of the child’s play and all experiments children performed were scored. An experiment was defined as putting one or more puppets in place and pushing the light button. Different types of experiments were distinguished: unconfounded experiments in which one dimension was varied (see Figure 2A-B), confounded experiments in which both dimensions were varied (see Figure 2C-D) and other experiments, in which either no dimensions were varied (see Figure 2E), or an irrelevant comparison was made (see Figure 2F). Irrelevant comparisons included putting one puppet in place, putting two puppets in place at the same side of the machine, or putting 3 or 4 puppets in place. Double coding of the experiments performed by 19 children (19%) rendered a percentage agreement of 96%, corresponding with a Kappa of .95.

Post-task. The post-task consisted of four items: two size items (see Figure 2A) and two distance items (see Figure 2B). As 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”, incorrect not “the same”.

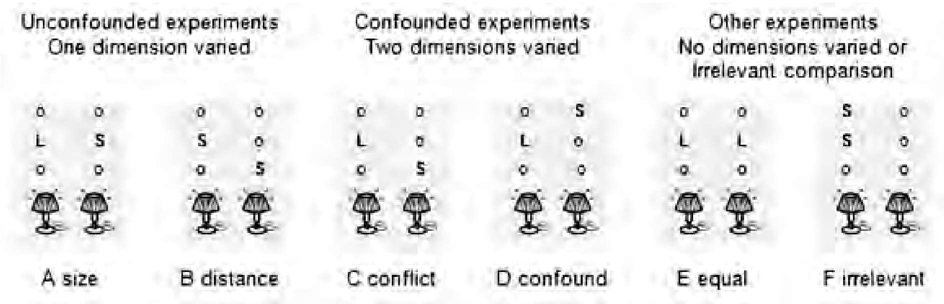

FIGURE 2. Examples of different types of items and experiments on the shadow machine. For the pre- and post-task size items (A) and distance items (B) were used (without feedback/ shadows). For the evidence exposure, a conflict item (C) was used for the conflicting condition and a confound item (D) for the confirming condition (with feedback/ shadows). For the free play episode, the experiments children performed were categorized in three categories (A-B, C-D, E-F, with feedback/ shadows). These figures present examples of the different types of experiments, though other variations are possible. Irrelevant experiments also included experiments with 1, 3 or 4 puppets.

Statistical approach

To detect children's naïve theories on shadow size, we used Latent Class Analysis (LCA): a statistical technique for models with categorical, manifest variables and a categorical latent variable. To this end, Latent Class Models were fit to the data. Latent Class Models describe categorically different response patterns in terms of latent classes. In this study these classes can be interpreted as children's rules or naïve theories on shadow size.

LCA provides a statistically more reliable technique to detect different types of response patterns than pattern matching techniques, which are used in Siegler's (1976, 1981) Rule Assessment Methodology (see Van der Maas & Straatemeier, 2008 for a more extended discussion). For example, in LCA no arbitrary criterion for correspondence between observed and expected responses is used to classify children, instead classification is based on the characteristics of the data. As mentioned before, LCA makes it possible to detect unanticipated response patterns, or classes. In addition, LCA uses model selection techniques that allow for an optimal decision between goodness-of-fit and parsimony of the model. Therefore, the technique can be used to test whether knowledge is coherent, theory-like, or fragmented. Previous studies on children's naïve theories on science subjects have used LCA to investigate rule-use on the balance scale task (Boom, Hoijtink, & Kunnen, 2001; Jansen & Van der Maas 1997, 2001, 2002), children's mental models of the earth (Straatemeier, Van der Maas & Jansen, 2008) and children's mental models of fetal development (Van Es, Van Schijndel, Franse & Raijmakers, 2009). Several introductions to LCA exist (e.g. Bouwmeester & Sijtsma, 2007; McCutcheon, 1987; Rindskopf, 1987), so in this paper only a brief description of the technique is given.

In the present study, exploratory LCA was used to determine the number of latent classes needed to describe children's response patterns to the pre- and post-task items (separately) in the best and most parsimonious manner. Latent Class Models were defined by fixing the number of latent classes and subsequently estimating the model parameters: the unconditional and conditional probabilities. Unconditional probabilities define the class sizes. Conditional probabilities of a class indicate probabilities of responses to specific items given membership of the class. Latent Class Models were fit to the data by calculating Log Likelihood estimates of the model parameters with the package depmixS4 (Visser & Speekenbrink, 2010) for the R statistical programming environment (R Development Core Team, 2011). In LCA the optimal number of latent classes cannot be estimated or tested, hence model selection was based on the Bayesian Information Criterion (BIC, Schwartz, 1978): the model with the lowest BIC was considered to be the most parsimonious, best fitting model. The best fitting model was then interpreted on the basis of previous work on children's naïve theories on shadow size (Chen, 2009; Ebersbach & Resing, 2007; Siegler, 1978, 1981).

RESULTS

Naïve theories on shadow size

First, we investigated what naïve theories young children have about shadow size. To this end, Latent Class Models (LCM; see Method section) with 1, 2, 3, 4 and 5 classes were fit to the pre-task data: children's trichotomous response patterns on 12 items. The pre-task items had been administered in one of two semi-random orders. As 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. Table 1 shows the goodness-of-fit measures of the different Latent Class Models. Based on the BIC values, it was found that a 4-class-model fit the data in the best and most parsimonious manner, indicating four different theories on shadow size. We put equality constraints on the response probabilities of the six size items in all classes and the same was done for the response probabilities of the six distance items in all classes. There was no significant difference in goodness-of-fit between the constrained and the unconstrained model (Log likelihood ratio (82)=85.62, p=.37). Therefore, the more parsimonious, constrained model was selected for interpretation and further analyses.

TABLE 1. Naïve theories on shadow size: goodness-of-fit measures for Latent Class Models of responses to pre- and post-task items.

Pretest				Posttest			
Number of classes	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 (calculated by the number of freely estimated parameters minus the number of parameters estimated at the boundary), BIC=Bayesian Information Criterion. The shaded rows show the goodness-of-fit measures of the constrained 4-class-model (pre-task) and constrained 3-class-model (post-task) that were selected for interpretation.

Figure 3 shows the parameter estimates of the selected 4-class-model. The first class was characterized as applying Rule 1 (N=39). Children in this class had high probabilities of giving correct responses on size items and answering “the same” on distance items. The second class was characterized as applying Rule 2-reversed (N=20). As children in the Rule 1 class, children in this class had high probabilities of giving correct responses on size items and incorrect responses on distance items. However, this group tended to answer distance items by claiming that a puppet closer to the light source would give a smaller shadow than a puppet further away from the light source. The third class was characterized as applying Rule 2+ (N=27). Children in this class had high probabilities of giving correct responses on both the size items and the distance items. As on the basis of the sole use of size- and distance items children applying Rule 2 could not be distinguished from children applying Rule 3 or 4, the plus sign (Rule 2+) was used to indicate the possible use of a more advanced rule. The fourth class was characterized as applying a Guess strategy (N=16). Children in this class did not show systematic responses to the size and distance items.

The most likely class membership was calculated for each child separately based on the posterior probabilities given the selected 4-classs-model (e.g. McCutcheon, 1985). It was demonstrated that rule-use was age-related ($\chi(6)=29.96$, $p<.001$). As shown in Figure 4, the use of Rule 1 decreased with age, the use of Rule 2-reversed and Rule 2+ increased with age and the Guess strategy was applied in all age categories.

To summarize, consistent with our expectations, two age-related theories, Rule 1 and Rule 2+, and a Guess strategy were found. In addition, the data also revealed children’s use of an unexpected theory: Rule 2-reversed.

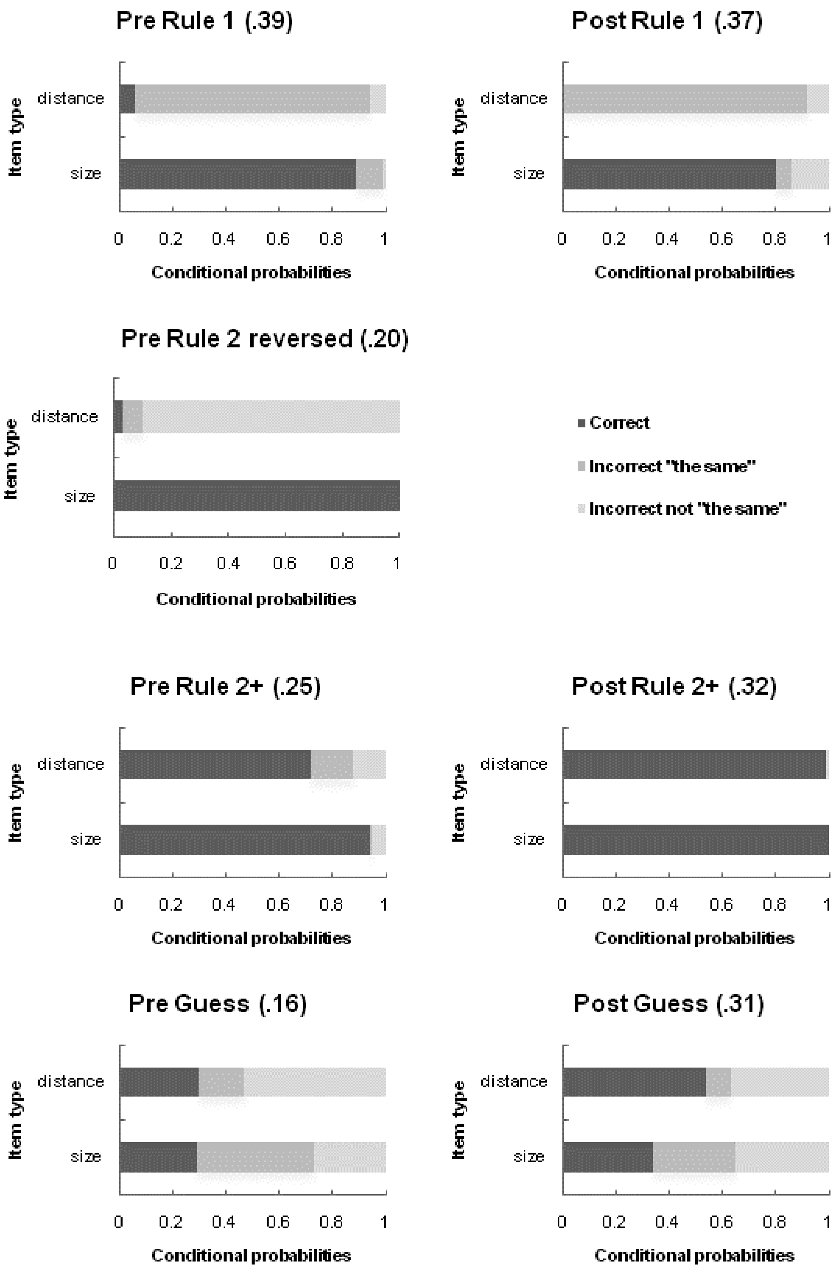

FIGURE 3. Naïve theories on shadow size: model parameters for selected pre- and post-models. Unconditional probabilities of the different classes are shown between brackets. These probabilities are not necessarily exact copies of percentages of children assigned to a class.

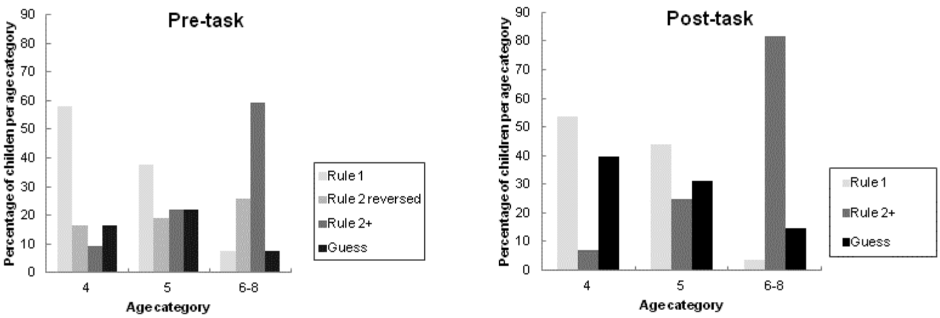

FIGURE 4. Naïve theories on shadow size: percentages of children per age category using the different rules on the pre- and post-task.

Conflicting evidence and play

Second, we investigated whether the quality of children's free play was affected by evidence conflicting with their naïve theory. As the evidence was designed for children using Rule 1, this question was first answered by focusing on this rule group's play behavior. Out of the 39 children that applied Rule 1 during the pre-task, 16 were assigned to the conflicting condition (*M*=56.13 months, *SD*=7.42) and 23 to the confirming condition (*M*=57.30 months, *SD*=7.33).[2] The evidence-item prediction (see Method section) was made correctly by 6% of the children in the conflicting condition, compared to 70% of the children in the confirming condition ($\chi(1)=15.38$, $p<.001$). The evidence-item observation (see Method section) was done correctly by 93% of the children in the conflicting condition and 100% of the children in the confirming condition.

A multivariate analysis of variance was conducted with condition (conflicting, confirming) as between-subjects factor on the following four free-play outcome measures: number of unconfounded experiments, number of confounded experiments, number of other experiments, and total number of experiments (see Method section for a description of the experiments in each of these categories). Effects of condition were found on the number of unconfounded experiments ($F(1,38)=5.70$, $p<.05$) and the total number of experiments $F(1,38)=4.33$, $p<.05$). Children in the conflicting condition performed more unconfounded experiments, and demonstrated a higher total number of experiments during free play than children in the confirming condition (see Table 2).

TABLE 2. Conflicting evidence and play: Mean numbers of unconfounded, confounded, and other experiments (and standard deviations) that children in both conditions performed during free play

		Unconfounded experiments	Confounded experiments	Other experiments	Total experiments
Rule 1 group	Conflicting condition	4.69 (2.73)	3.31 (4.48)	9.06 (7.23)	17.06 (8.20)
N=39	Confirming condition	2.61 (2.64)*	2.39 (2.66)	7.13 (5.03)	12.13 (6.58)*
Total sample	Conflicting condition	5.14 (3.34)	3.56 (3.63)	8.86 (6.77)	17.56 (8.19)
N=102	Confirming condition	3.58 (2.89)*	3.25 (2.84)	8.52 (6.82)	15.34 (8.71)

Note. * Indicates a significant difference.

Next, the play analysis was repeated for the total sample. Fifty children were assigned to the conflicting condition (*M*=65.86 months, *SD*=16.09; 9 Guess, 16 Rule 1, 10 Rule 2-reversed, 15 Rule 2+) and 52 children were assigned to the confirming condition (*M*=66.27 months, *SD*=15.19; 7 Guess, 23 Rule 1, 10 Rule 2-reversed, 12 Rule 2+).[2] Even though the evidence distinguishing the conditions was specifically designed for the Rule 1 group, we expected the conflicting evidence to be considered conflicting and the confirming evidence to be considered confirming by children applying several other rules too. For example, children applying Rule 2-reversed would predict shadows to be bigger when puppets are large and placed far away from the light source. The evidence in the conflicting condition would contradict these predictions. Children applying Rule 2 would, as the object sizes on the evidence items differed, base their judgments on the size dimension and predict shadows to be bigger when puppets are large. The evidence in the consistent condition would not contradict these predictions, while the evidence in the conflicting condition would. The difference between conditions in the percentages of children making correct evidence-item predictions (see Method section), confirmed our expectations: 14% of the children in the conflicting condition made a correct prediction, compared to 56% of the children in the confirming condition ($\chi(1)=19.47$, $p<.001$). The evidence-item observation (see Method section) was done correctly by 96% of the children in the conflicting condition and 100% of the children in the confirming condition.

A multivariate analysis of variance was conducted with condition (conflicting, confirming) as between-subjects factor on the following four free-play outcome measures: number of unconfounded experiments, number of confounded experiments, number of other experiments, and total number of experiments. As performing the analysis on the total sample made it possible to look at the relation between prior knowledge and play, theory group (Guess, Rule 1, Rule 2-reversed, Rule 2+) was added as a between-subjects factor and age in months as a covariate. Effects of age were found on the number of unconfounded ($F(1,93)=11.34$, $p<.01$) and confounded experiments ($F(1,93)=5.06$, $p<.05$), and the total number of experiments ($F(1,93)=8.47$, $p<.01$): older children performed more unconfounded and confounded experiments, and demonstrated a higher total number of experiments during free play than younger children. An effect of condition was found on the number of unconfounded experiments ($F(1,93)=7.21$, $p<.01$): children in the conflicting condition performed more unconfounded experiments during free play than children in the confirming condition (see Table 2). Effects of theory group were found on the number of other experiments ($F(3,93)=2.79$, $p<.05$), and the total number of experiments. ($F(3,93)=3.56$, $p<.05$). Post hoc tests with Bonferroni corrections revealed that the Rule 2-reversed group performed more other experiments than the Guess group ($t(34)=2.77$, $p<.05$), and that the Rule 2-reversed and Rule 2+ performed a higher total number of experiments than both the Guess group and Rule 1 group (Guess-Rule 2-reversed, $t(34)=3.59$, $p<.01$; Guess-Rule 2+, $t(41)=2.99$, $p<.05$; Rule 1-Rule 2-reversed, $t(57)=3.24$, $p<.05$; Rule 1-Rule 2+, $t(64)=2.64$, $p<.05$).

To summarize, both the analyses on the Rule 1 group and the analyses on the total sample demonstrated that children who were confronted with conflicting evidence showed an increase in quality of free play, reflected by the higher number of unconfounded, informative experiments they conducted. This was not a sole reflection of an overall increase in the quantity of play, as they did not perform more uninformative, confounded or other, experiments.[3] In addition, controlling for age, children having more prior knowledge (children with more advanced theories on shadow size) performed more experiments during free play, but did not demonstrate a higher quality of play than children having less prior knowledge.

Learning from play – analyses on the individual level

Third, we investigated whether children learned from play. We first analyzed learning on the individual level. We fit Latent Class Models (LCM; see Method section) to the post-task data, children's trichotomous response patterns on 4 shadow task items, in the same way as was done for the pre-task data (see Table 1 for the goodness-of-fit measures of the different models). It was found that a constrained 3-class-model fit the data in the best and most parsimonious manner (Log likelihood ratio constrained versus unconstrained 3-class-model (12)=20.86, $p=.05$)[4], and therefore this model was selected for interpretation and further analyses (see Figure 3 for the parameter estimates of the model).As in the pre-task-model, one class could be characterized as applying Rule 1 (N=38), one as applying Rule 2+ (N=33) and one as applying the Guess 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 Figure 4, the use of Rule 1 and the Guess strategy decreased with age and the use of Rule 2+ increased with age.

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 3). Overall children showed consistency in rule-use: the large majority of children who used a specific rule on the pre-task also used this rule on the post-task (77% for Rule 1, 74% for Rule 2+ and 75% for the Guess strategy). However, as no Rule 2-reversed group was found in the post-task data, the children who had used this rule on the pre-task either used Rule 1 (15%), Rule 2+ (40%) or the Guess strategy (45%) on the post-task. A trend approaching significance demonstrated that few children (10%) applying Rule 1 learned from play, but about half of the children applying Rule 2-reversed did (McNemar $\chi(1)=7.22$, $p<.1$).

TABLE 3. Learning on the individual level: numbers (and percentages) of children using specific combinations of rules on the pre- and posttest.

		Posttest			
		Rule 1	Rule 2+	Guess strategy	Total
Pretest	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)
	Guess strategy	3 (19)	1 (6)	12 (75)	16 (100)
	Total	38 (37)	33 (32)	31 (31)	102 (100)

Learning from play – analyses on the group level

Next, we analyzed learning on the group level. When investigating children's learning on the individual level, we could not investigate learning for the different conditions separately as the number of participants per cell was too small. Therefore, the whole group was included for analysis. Two repeated measures analyses of variance were conducted: one on proportion size items correct and one on proportion distance items correct. In both analyses time (pre-task, post-task) was included as within-subjects factor, condition (conflicting, confirming) and theory on pre-task (Rule 1, Rule 2-reversed, Rule 2+, Guess strategy) as between-subjects factors, and age in months and play as covariates. Play was operationalized as the number of unconfounded size experiments (see Figure 2A) the child performed during free play in the analysis on proportion size items correct, and as the number of unconfounded distance experiments (see Figure 2B) the child performed during free play in the analysis on proportion distance items correct.

In the analysis on proportion size items correct, effects of condition ($F(1,92)=10.59$, $p<.01$), theory ($F(1,92)=27.12$, $p<.001$), and time ($F(1,92)=4.10$, $p<.05$) were found. An interaction between time and condition was found ($F(1,92)=19.98$, $p<.001$): over time the conflicting condition's performance on size items deteriorated ($t(49)=4.38$, $p<.001$), while the confirming condition's performance did not change (see Figure 5A). An interaction between time and theory was found ($F(1,92)=6.93$, $p<.001$): the difference scores (proportion size items correct on post-task - proportion size items correct on pre-task) of the Rule 2-reversed group differed from the difference scores of the other theory groups (Guess group $t(33.92)=3.24$, *p*<.01, Rule 1 group *t*(32.23)=2.21, *p*<.05, Rule 2+ group *t*(25.03)=2.91, *p*<.01). Over time the Rule 2-reversed group's performance on size items deteriorated (*t*(19)=3.47, *p*<.01), while performance in the other groups did not change (see Figure 5B).[5]

In the analysis on proportion distance items correct, effects of theory (*F*(1,92)=18.37, *p*<.001) and age in months (*F*(1,92)=10.93, *p*<.01) were found. An interaction between time and play was found (*F*(1,92)=4.49, *p*<.05): over time the performance on distance items of the group that conducted more distance experiments during free play improved (*t*(50)=5.65, *p*<.001), while the performance of the group that conducted less distance experiments during free play did not change (see Figure 6A). An interaction between time and theory was found (*F*(1,92)=6.66, *p*<.001): the difference scores (proportion distance items correct on post-task - proportion distance items correct on pre-task) of the Rule 2-reversed group differed from the difference scores of the other theory groups (Guess group *t*(34)=3.01, *p*<.01, Rule 1 group *t*(31.56)=3.63, *p*<.01, Rule 2+ group *t*(36.39)=3.67, *p*<.01). Over time the Rule 1 group's (*t*(38)=2.53, *p*<.05) and Rule 2-reversed group's (*t*(19)=5.60, *p*<.01) performance on distance items improved, while performance in the other groups did not change (see Figure 6B).[6]

To summarize, over time children's performance on size items was negatively affected by evidence conflicting with their naïve theory. Over time children's performance on distance items improved more for the group that conducted more distance experiments during free play than for the group that conducted less distance experiments during free play. In addition, children who had a Rule 2-reversed theory at the start of the experiment showed a larger negative change in performance on size items and a larger positive change in performance on the distance items over time than children in the other theory groups.

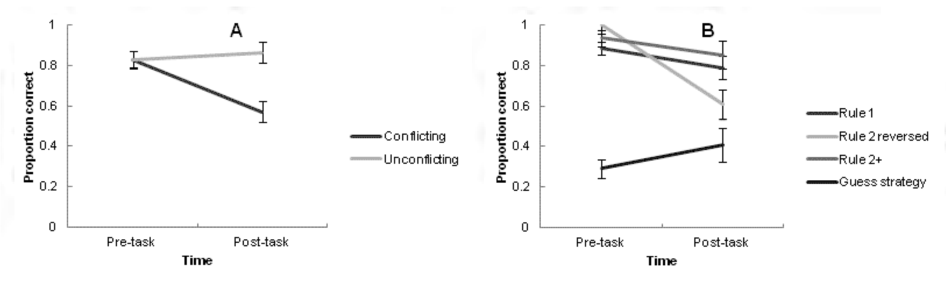

FIGURE 5. Learning on the group level: average proportion correct on the size items on the pre- and post-task per condition (A) and per theory group (B). Error bars represent standard errors.

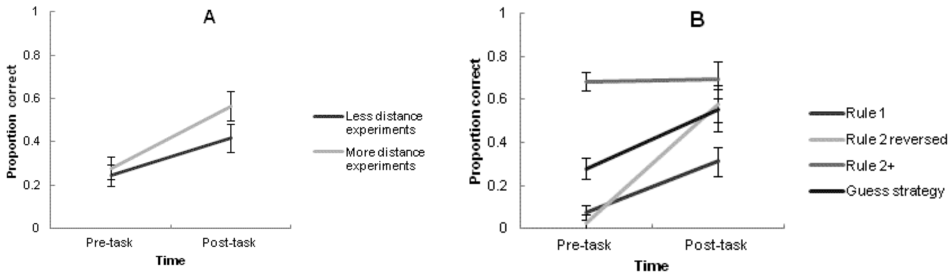

FIGURE 6. Learning on the group level: average proportion correct on the distance items on the pre- and post-task per play group (A) and per theory group (B). Error bars represent standard errors. A) The play variable was taken along in the repeated measures analysis as a covariate. To explain and plot the interaction between time and play on proportion distance items correct, two groups were created by splitting the play variable at the median.

DISCUSSION

This study investigated the role of young children's prior knowledge in guiding their exploratory play and learning. We first assessed 4- to 9-year-olds' naïve theories on shadow size by using a categorical latent variable technique: Latent Class Analysis (LCA). In line with previous literature (Chen 2009; Ebersbach & Resing, 2007; Siegler, 1981), we found a group taking into account the size dimension (Rule 1), a group taking into account both dimensions (Rule 2+), and a Guess group. Children's naïve theories were found to be age-related: the majority of the 4- and 5-year-olds used Rule 1, while the majority of older children used Rule 2+. In addition to the expected groups, a fourth group that took into account the size dimension in the right direction, but the distance dimension in the wrong direction, was distinguished (Rule 2-reversed). The existence of this Rule 2-reversed group is in line with Ebersbach and Resing's (2007) results. The authors assessed implicit beliefs by asking participants to estimate shadow sizes of different objects closer to and further away from a light source. They found, depending on age, that participants estimated 63 to 93% of shadows of objects in the closest position to the light source as being longer than shadows of objects in the furthest position from the light source. The authors assessed explicit beliefs by asking participants whether a shadow would become longer, shorter or would stay the same when an object approaches the light. They found that approximately half of the 5-year-olds and the majority of older participants answered this question correctly. Interestingly, 20% of the adults gave an incorrect answer. However, even though Ebersbach and Resing's (2007) results suggest the existence of a Rule 2-reversed theory, no systematic assessment of children's naïve theories was presented in their paper. For example, for the implicit results, the authors looked at averages over participants compared to individual performance. For the explicit results, the authors did not distinguish between the incorrect answers "longer" and "the same". Therefore, to the best of our knowledge, the present study is the first to systematically demonstrate the existence of Rule 2-reversed group. This finding illustrates the advantage of using LCA over inspecting averages, but also over pattern matching, which is done in Siegler's (1976, 1981) Rule Assessment Methodology. When matching observed to expected response patterns, other than by visual inspection of the data, no theories can be detected that are not defined beforehand. When using LCA, different models are fit to the data and only after selection of the best fitting model, the model's classes are interpreted. Analyzing the data in this way, other than by selecting specific item types, the results are not influenced by prior hypotheses about existing theories, enabling the finding of unanticipated groups, such as the Rule 2-reversed group.

Second, we investigated the effect of evidence conflicting with young children's naïve theories on the quality of their free play. Both in the preschooler age group (Rule 1 group) and the total sample, it was shown that children who were confronted with conflicting evidence conducted a higher number of unconfounded, informative experiments during free play than children who were confronted with consistent evidence. Importantly, this was not due to an overall increase in quantity of experiments, as children did not perform more uninformative experiments. The result complements previous work on the effects of different types of uncertainty on young children's exploratory play. Bonawitz et al. (2012) demonstrated that uncertainty generated by conflicting information increases the duration of young children's play. This study extends those results by showing that this type of uncertainty also affects the quality of play. Also, we show this effect in a younger age group: preschoolers. Cook et al. (2011) demonstrated that uncertainty generated by ambiguous evidence affects the quality of young children's play in an artificial context (Cook et al., 2011). The current results extend this finding by showing that uncertainty can also affect the quality of play in ecologically valid contexts involving multiple causal factors (size & distance).

The finding that children in the preschool age-range, without any form of training, apply the principle of controlling variables, contrasts with research on older children's science learning. In science education research, the skill of designing unconfounded experiments from which valid causal inferences can be drawn is called the Control of Variables Strategy (CVS; e.g. Chen & Klahr, 1999; Klahr, 2005; Klahr & Nigam, 2004; Kuhn & Dean, 2005). The acquisition of CVS is considered essential in development of scientific reasoning, but several studies demonstrated primary school-aged children's difficulty with the use and transfer of the strategy (e.g. Chen & Klahr, 1999; Klahr & Nigam, 2004; Kuhn, Garcia-Mila, Zohar & Andersen, 1995; Schauble, 1996). For example, Chen and Klahr (1999) found that the 7- to 10-year-olds in their study, when asked to investigate the effect of a specific variable, only used CVS in 26 to 48% of the experiments they designed. Instruction improved the use of the strategy, but discovery learning did not. There are several explanations for the discrepancy between older children's difficulty with acquiring CVS and preschoolers' selective application of the strategy in uncertain situations (see Bonawitz et al., 2012 and Cook et al., 2011 for a more extended discussion). One explanation is that the contexts used for older children's inquiry learning are generally far more complex than the contexts used in research on preschoolers' exploratory play. Additionally, the tasks administered to older children often ask for a meta-cognitive understanding of CVS, while the tasks administered to preschoolers do not.

Third, in a more explorative manner, we investigated children's learning. In line with previous work on young children's learning from play (Gweon & Schulz, 2008; Schulz, Gopnik & Glymour, 2007), only a small percentage (21%) of 4- to 9-year-olds that did not take into account both dimensions in the correct manner at the start of the experiment had a more advanced theory after free play than before free play. As predicted, children's prior knowledge was found to affect learning: children applying Rule 2-reversed were affected differently by evidence and free play than children applying other rules. The data on the individual level revealed a trend showing a higher percentage Rule 2-reversed children learning from play than Rule 1 children. The analyses on the group level demonstrated that over time the Rule 2-reversed group showed a larger negative change in performance on size items and a larger positive change in performance on the distance items than children in the other theory groups. The fact that children having this quite advanced theory showed a negative change in performance on the size items contrasts with the developmental pattern outlined by Siegler (1981). A possible explanation for these findings is that children using Rule 2-reversed are a transitional group: they are aware of both the size and distance dimension having a role in determining shadow size, but are still unsure about the specific relations. This group might be more susceptible to evidence than children in non-transitional groups. Conflicting evidence shown by the experimenter or encountered during free play might have either led them to hypothesize that they were wrong about the direction of the size effect, or that they were wrong about the direction of the distance effect. The former option would lead them to give incorrect responses on both the size and distance items, leading them to end up in the Guess group (45%). The latter option would lead them to give correct responses on both item types, leading them to end up in the Rule 2+ group (40%). Partial support for this hypothesis comes from the analyses on the group level showing an effect of condition on children's performance on size items over time: children in the conflicting condition showed a decrease, while children in the confirming condition did not.

Besides children's prior knowledge, the quality of children's free play was also predicted to be related to their learning. This prediction was confirmed with regards to the distance items: children who conducted more unconfounded distance experiments during free play, showed a larger increase in performance over time than children who conducted less of these experiments during free play. However, as 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.

Future work could investigate the unique contributions of different factors affecting young children's causal learning from play in this study's knowledge domain. An important question is how prior knowledge affects learning. Descriptive work on older children's and adults' causal learning from experimentation suggests that this is through the ability to formulate hypotheses and draw valid conclusions from experiments (e.g. Dunbar & Klahr, 1989; Schauble, 1990). For example, Dunbar and Klahr (1989) compared 8- to 11-year-old's performance to adults' performance on a robot programming task. They found that children's learning was limited (9%) compared to that of adults (95%). Both groups performed similar experiments, but children came up with different, less efficient, hypotheses than adults due to a lack of prior knowledge. Children also evaluated the evidence of experiments differently, often ignoring negative evidence, and only interpreting the results of their most recent experiment. Parallels could be drawn to the present study: for example, younger children applying Rule 1 might not have been able to formulate hypotheses including the distance factor, affecting their ability to learn from distance experiments (Jansen & Van der Maas, 2001). A second line of questions for future research is related to factors influencing the quality of children's free play. As in Dunbar and Klahr's (1989) study, in the present study participants with more prior knowledge did not demonstrate a higher quality of play than participants with less prior knowledge. Conflicting evidence shown by the experimenter, however, did affect the quality of children's free play. An interesting question concerns how much of this effect was caused by the pedagogical context in which the conflicting evidence was shown (e.g. Bonawitz et al., 2011). In other words, does children's quality of play also increase when they themselves encounter conflicting evidence during free play? Answering these questions will further unravel the mechanisms through which young children learn from experimentation.

ACKNOWLEDGEMENTS

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 a NWO-VIDI grant. We would like to thank the primary schools for their cooperation. We thank Marjolein Hof and Juri Peters for their help in collecting the data, Wouter Weeda and Han van der Maas for giving statistical advice, Coos Hakvoort for providing technical support and Elise Herrig performing a language check.

NOTES

1. In this study children were not equally divided over the different age groups: less 6- to9-year-olds participated than 4- and 5-year-olds. This difference was caused by the researchers aiming at knowledge-homogeneous groups of comparable sizes instead of age-homogeneous groups of comparable sizes. This study investigated the relationship between prior knowledge and learning, and no added effect of age on learning was expected. Therefore, striving for knowledge groups of comparable numbers of participants was considered more important than striving for age groups of comparable numbers of participants. Final classification and therefore final sizes of the different knowledge groups were based on the results of the Latent Class Analysis.
2. The number of participants in both conditions was not the same. During task administration a shortcut was used to assign children to theory groups (see method section). Half of the children that was assigned to a specific theory group during task administration was shown conflicting evidence and the other half was shown confirming evidence. However, the final assignment to the theory groups was based on the Latent Class Analysis. This final classification differed from the temporary classification and this caused the unequal number of participants in both conditions.
3. On the basis of significant differences between conditions in number of unconfounded experiments and non-significant differences between conditions in confounded or other experiments, the conclusion was drawn that the conflicting condition shows a higher quality of play than the confirming condition. However, the reported analyses did not test whether the difference between the conditions in number unconfounded experiments was larger than the difference in confounded or other experiments. To preempt this point we also performed analyses on percentages of unconfounded experiments. In the total sample (N=97, 5 participants did not perform experiments, so no percentages could be calculated) a significant difference between conditions was found ($t(95)=2.04$, $p<.05$): children in the conflicting condition demonstrated a higher percentage of unconfounded comparisons than children in the confirming condition. In the Rule 1 group (N=37, 2 participants did not perform experiments, so no percentages could be calculated) the difference did not reach significance due to the small number of participants. However, the results on percentages of unconfounded experiments support the conclusion that children in the conflicting condition demonstrated a higher quality of play than children in the confirming condition.
4. The item-heterogeneity was just sufficient to reject the unconstrained model. This might be caused by that fact that children paid less attention at the first post-task item. However, the interpretation of the classes in the unconstrained model was the same as the interpretation of the classes in the constrained model, so the rejection of the unconstrained model did not change the results.
5. The interaction between time and theory group on proportion correct on the size items was not solely caused by regression to the mean. The mean proportion correct for the

Rule 2-reversed and Rule 2+ group differed on the pre-task ($t(26)=2.08$, $p<.05$). However, this difference was caused by the lack of variance of the Rule 2-reversed group: all children demonstrated a proportion correct of 1. Therefore, we constructed matched samples by removing the five participants from the Rule 2+ group that did not demonstrate a proportion correct of 1, leaving a matched Rule 2+ group of 22 participants. The Rule 2-reversed group showed a larger decrease in performance from pre- to post-task than the matched Rule 2+ group ($t(26.85)=2.58$, $p<.05$).

6. The interaction between time and theory group on proportion correct on the distance items was not solely caused by regression to the mean: on the pre-task the mean proportion correct did not differ for the Rule 2-reversed and Rule 1group ($t(57)=1.12$, $p=.27$), but the Rule 2-reversed group showed a larger increase in performance from pre- to post-task than the Rule 1 group ($t(31.56)=3.63$, $p<.01$).