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 5

Development of causal learning in 2- to 5-year-olds: Multiple age-related types of causal inference

Van Schijndel, T. J. P., Huijpen, K., & Raijmakers, M. E. J. *Development of causal learning in 2- to 5-year-olds: Multiple age-related types of causal inference*. Manuscript in revision.

ABSTRACT

Children's capabilities for causal inference are usually studied on the basis of the average behavior of children within age groups. However, children within age groups show considerable variance in their responses to causal inference trials. The present study aimed to describe the development of young children's causal inference by investigating individual differences in the type of inference they use. To this end, a series of carefully selected causal inference trials was administered to children of a relatively broad age range: 2- to 5-year-olds. Instead of analyzing the group's responses on each trial separately, individuals' response patterns over trials were analyzed with a categorical latent variable technique. Different sources of variance were distinguished: children using different types of causal inferences and children making errors in applying these types of inferences. The developmental pattern was described as the application of multiple types of causal inferences: an older group responding consistent with the Bayes net account (Gopnik et al., 2004), a group imitating the experimenter, and a younger group responding consistent with causal inference on the basis of associative models. The existence of this third group suggests a developmental progression from inferring associations to the construction of causal maps. The results demonstrate the importance of accounting for individual differences when studying the development of causal inference.

INTRODUCTION

Young children have causal knowledge in a variety of domains (e.g. Bullock, Gelman & Baillargeon, 1982 in Gopnik et al., 2004; Gelman & Wellman, 1991; Wellman, 1990). The question that has dominated the literature on children's causality for the last decade is how new causal relations are learned and represented (Gopnik et al., 2004; Sobel, Tenenbaum & Gopnik, 2004). Several studies have shown that children are capable of making causal inferences in an unfamiliar domain from a very young age (Gopnik, Sobel, Schulz & Glymour, 2001; Schulz & Gopnik, 2004; Sobel & Kirkham, 2006; Sobel et al., 2004). Many of these studies investigated whether young children's inferences are in line with the Bayes net account, a computational theory of children's causal learning (see Gopnik et al., 2004 for an outline of the account). Researchers in these studies presented children with interventions[1] and correlations among events and then tested whether children's responses were consistent with the construction of causal maps, which are abstract, coherent, learned representations of causal relations (Gopnik et al., 2004). Excitingly, very young children appeared to construct causal maps to represent observed evidence (e.g. Gopnik et al., 2001; Sobel & Kirkham, 2006).

A paradigm that has often been used in these studies is the blicket detector (Gopnik & Sobel, 2000; Nazzi & Gopnik, 2000). This is a machine that activates (lights up and plays music) when some objects (blickets), but not others, are placed on it. The paradigm allows for testing how new causal relations are learned, because one can largely control the evidence learners are presented with. Each blicket detector trial is designed to measure a learner's capability to make a specific causal inference. In many studies, inferences and corresponding trials have been selected that enable researchers to distinguish between responses consistent with the Bayes net account and responses consistent with other types of causal inferences, such as causal inference on the basis of associative models (see Gopnik et al., 2004 and Sobel et al., 2004 for an overview of learning mechanisms; De Houwer & Beckers, 2002). Below we will briefly discuss a couple of trials that are relevant for the present study: *screening-off*, *indirect screening-off* and *backwards blocking*.

Several studies have demonstrated that the majority of 19-, 24- and 30-month-olds (Gopnik et al., 2001; Sobel & Kirkham, 2006) and 3- to 5-year-olds (Gopnik et al., 2001; Schulz & Gopnik, 2004) make inferences consistent with the Bayes net account on *screening-off* trials (Reichenbach, 1956). Children did not consider a target object that did not activate the machine by itself, but activated the machine in the presence of an established blicket, to be a blicket.

Other studies have investigated young children's ability to make retrospective inferences (Sobel et al., 2004). It has been shown that the majority of 24-month-olds (Sobel & Kirkham, 2006) and 3- and 4-year-olds (Sobel et al., 2004), but not the majority of 19-month-olds (Sobel & Kirkham, 2006) make inferences consistent with the Bayes net account on *indirect screening-off* trials.[2] When a target object activated the machine in the presence of a second object and the second object was subsequently shown not to activate the machine by itself, children considered the target object to be a blicket. A second retrospective inference is *backwards blocking*. It has been shown that the majority of 4-year-olds, but not the majority of 3-year-olds (Sobel et al., 2004) make inferences consistent with the Bayes net account on *backwards blocking* trials. When a target object activated the machine in the presence of a second object and the second object was subsequently shown to activate the machine by itself, children, most of the time, did not consider the target object to be a blicket. As the associative strength of the target object is the same in *indirect screening-off* and *backwards blocking* trials, the asymmetry in children's classification of this object on the two retrospective trials has been used to rule out causal inference based on associative models (e.g. Dickinson, 2001; Rescorla & Wagner, 1972; Shanks & Dickinson, 1987) as a possible mechanism accounting for children's causal learning (Sobel & Kirkham, 2006; Sobel et al., 2004).

Development & Individual differences

Some developmental differences have been uncovered in the studies described above (Sobel & Kirkham, 2006; Sobel et al., 2004). For example, 4-year-olds were found to give responses that were in line with the Bayes net account on *backwards blocking* trials, while 3-year-olds were not (Sobel et al., 2004). The conclusions with regard to these developmental differences were based on the average behavior of children within age groups. However, these studies showed sizable variance within age groups. Within age groups with a majority of children responding consistent with the Bayes net account, percentages ranged from 72 to 100 for *screening-off inferences* (Gopnik et al., 2001; Schulz & Gopnik, 2004; Sobel & Kirkham, 2006) and from 76 to 100 for *indirect screening-off* and *backwards blocking* inferences (Sobel et al., 2004; Sobel & Kirkham, 2006). The aim of the present study was to further investigate the development of young children's ability to use causal inference by studying these individual differences. In particular, it was investigated whether individual differences exist in the type of causal inferences children make and whether these differences are age-related. The study uniquely contributes to the literature by administering a series of carefully selected causal inference trials to children of a relatively broad age range and analyzing individuals' response patterns over trials with a categorical latent variable technique.

Individual differences in children's causal inference could be caused by different sources of variance. A first possible source is the existence of individual differences in the type of causal inferences children make, which could be related to age. Besides a group of children using Bayesian inference,[3] a subgroup of children might use an alternative learning mechanism resulting in a different type of causal inference (see Gopnik et al., 2004 and Sobel et al., 2004 for an overview of learning mechanisms). Individual differences in type of causal inference could also result from a subgroup of children not making causal inferences, but applying a task related strategy, such as imitating the experimenter (e.g. Sobel & Kirkham, 2006). Another source of variance could result from children making errors in causal inference. For example, children may use a specific type of causal inference, but may make mistakes in applying this mechanism. Both abovementioned sources of variance could also hold simultaneously. Therefore, in this study we aimed to answer three questions: (1) Do children show different types of causal inferences? (2) Can these types of causal inferences be characterized in terms of learning mechanisms or task-related strategies? (3) Are these different types of causal inference age-related?

The present study

For the purpose of answering the research questions, Siegler's (1976, 1981) Rule-Assessment Methodology was applied in the present study. That is, a series of carefully selected causal inference trials was administered to participants from a broad age range: 2-to 5-year-olds. Instead of analyzing the group's responses on each trial separately, individuals' response patterns over trials were analyzed. Trials were selected because of their frequent use in previous work and because children had shown considerable variance in their responses on them. Another reason for selection was that these trials are suitable for distinguishing between different types of causal inferences. In the next paragraph the purpose of each of the trials will be discussed briefly, in the Method section the trials are described in detail.

The first two test trials were used to discriminate between children giving responses consistent with the Bayes net account and children giving responses consistent with a task-related strategy: imitation of the experimenter. The first trial was a *screening-off* trial (e.g. Gopnik et al., 2001) and the second trial was a variant of Schulz and Gopnik's (2004) *indirect screening-off trial in which children are not shown information about single candidate causes*. On both of these trials, children responding consistent with the Bayes net account were expected to respond to the question "Can you make the machine go?" by putting the causally efficacious block on the detector, while children responding consistent with the imitation strategy were expected to imitate the action of the experimenter that most frequently led to the activation of the machine.[4] The remaining three test trials were used to discriminate between children giving responses consistent with the Bayes net account and children giving responses consistent with causal inference based on associative models (e.g. Dickinson, 2001; Rescorla & Wagner, 1972; Shanks & Dickinson, 1987; Sobel et al., 2004). The third trial was an *indirect screening-off* trial and the fourth trial was a *backwards blocking* trial (e.g. Sobel et al., 2004). Work on adult reasoning showed that participants classified a target object that caused an effect in the presence of a second object differently, dependent on whether participants subsequently did or did not see the second object cause the effect by itself (Shanks, 1985 in Sobel et al., 2004; Shanks & Dickinson, 1987). This performance can be explained by the Bayes net account, but not by associative models (Sobel et al., 2004). Therefore, children responding consistent with the Bayes net account were expected to mimic adult behavior and show an asymmetry in the classification of the target object on both trials, while children responding consistent with causal inference based on associative models were expect to classify the target object equally often as a blicket on both trials (Sobel et al., 2004). The fifth trial was a variant of Gopnik and Sobel's (2000) and Nazzi and Gopnik's (2003) *non-causal association* trials. On this trial, children responding consistent with the Bayes net account were expected to respond to the question "Can you make the machine go?" by using the object that was in physical contact with the machine during activation, while children responding consistent with causal inference based on associative models were expect to use object(s) that were present during activation, but not in contact with the machine (see Method section).

By means of the statistical technique Latent Class Analysis (McCutcheon, 1987) categorical differences in response patterns were modeled (see Method section for an introduction to this technique). Finally, the resulting model (see Results section) was interpreted in the light of possible sources of variance in children's causal inference (see Discussion section).

METHOD

Participants

The final sample consisted of seventy-eight children (35 boys and 43 girls): 13 2-year-olds (*M*=29.38 months, *SD*=3.71), 21 3-year-olds (*M*=40.71 months, *SD*=3.52), 23 4-year-olds (*M*=54.39 months, *SD*=2.98) and 21 5-year-olds (*M*=67.33 months, *SD*=2.44) recruited from two daycare centers and a primary school. Sixteen other children were recruited but not included in the analyses: 2 refused to participate, 1 could not be tested due to failure of the blicket detector, 11 (8 2-year-olds and 3 3-year-olds) were excluded for failing the training trials (see Procedure section) and 2 (a 2-year-old and a 5-year-old) were excluded due to a missing value on one of the test trials. Although most children were from White, middle-class backgrounds, a range of ethnicities reflecting the diversity of the population was represented.

Materials

The blicket detector (Gopnik & Sobel, 2000; Nazzi & Gopnik, 2000) was used. The detector was made of grey plastic with an orange top and measured 7.4 x 4.3 x 2.8 in. (18.8 x 10.8 x 7 cm). A hand held remote control, hidden from the child's view, was used to control whether objects activated the detector. When the remote control was pressed the detector activated (it lit up and played music) when an object with a minimum weight of 50 grams was placed on the orange top. When the remote control was pressed again the detector did not activate when objects were placed on the orange top.[5] Objects that activated the detector did so as soon as they made contact with it and deactivation immediately followed the removal of the objects from the detector. Ten sets of unique objects, 22 in total, were used: 5 sets of weighted wooden blocks and 5 sets of weighted plastic toys of different shapes and colors (2 sets of fruits, 1 set of ducks and 2 sets of bowling pins). The distribution of the objects over the trials was counterbalanced (2 versions).[6]

Procedure

Children were tested individually by one of three female experimenters in a private room at their daycare center or at their school. The child sat facing the experimenter at a table. The blicket detector was on the experimenter's side of the table sitting on a grey tray. The experimenter introduced the blicket detector to the children by saying (in Dutch): "We are going to play with this machine. The machine goes on if some things are placed on top of it. If other things are placed on top of the machine it does not go on. You need to help me figure out which things make the machine go." Children were then shown a fixed order of ten trials: three training trials (A, B, C), two test trials (1, 2), two more training trials (D, E) and three more test trials (3, 4, 5).[7] The side of the detector where the causally efficacious object was placed on each of the trials was counterbalanced (2 versions with alternating sides).

Training trials A, B and C. The training trials were used to familiarize participants with the two types of responses that were required for the different test trials: an intervention to make the machine go (test trial 1, 2 and 5) and a verbal response to the question "Does this one make the machine go?" (test trial 3 and 4). The training trials also served as control trials to ensure that children were on-task and understood the nature of the task.[8] In order to be included in the analyses, participants had to answer a minimum of four out of five training trials correct. In trial A children were shown two new objects: X and Y. The experimenter placed object X on the machine by itself, the machine activated and she said: "See, it makes the machine go." The experimenter then placed object Y on the machine by itself, the machine did not activate and she said: "See, it does not make the machine go." Subsequently the experimenter slid the tray with the detector and two objects alongside it towards the child and asked: "Can you make the machine go?"[9] The child was allowed to make one response after which the experimenter slid the tray back to her side of the table. If the child placed the causally efficacious object (A) on the machine, it activated. Training trial B and C were similar to trial A, except that in trial B the second object was the causally efficacious one and in trial C the experimenter did not state anymore that objects did or did not make the machine go.

Test trial 1. Trial 1 was a *screening-off* trial. Children were shown two new objects: X and Y. The experimenter placed object X on the machine and the machine activated. The experimenter placed object Y on the machine and the machine did not activate. She then placed both objects on the machine together and the machine activated. This action was repeated. Subsequently the experimenter followed the procedure of training trial A and asked the child to make the machine go. On this trial, the response consistent with the Bayes net account is to put the causally efficacious object on the machine (object X). Causal inference based on associative models generates the same response.[10] However, the use of the imitation strategy predicts another response: imitation of the experimenter's action that most frequently led to the activation of the machine (objects X & Y).

Test trial 2. Trial 2 was an *indirect screening-off trial in which children were not shown information about single candidate causes.* Children were shown three new objects: X, Y and Z. The experimenter placed object X and Z on the machine together and the machine activated. She placed object Y and Z on the machine together and the machine activated. She then placed object X and Y on the machine together and the machine did not activate. Subsequently the experimenter followed the procedure of training trial A and asked the child to make the machine go. On this trial, the response consistent with the Bayes net account is to put the causally efficacious object on the machine (object Z). Causal inference based on associative models generates the same response.[9] However, the use of the imitation strategy predicts another response: imitation of the experimenter's action that most frequently led to the activation of the machine (objects X & Z or Y & Z).

Training trials D and E. In trial D children were shown two new objects: X and Y. The experimenter placed each object on the machine by itself. Object X did not activate the machine, object Y did. Subsequently the experimenter showed the child object X and asked "Does this one make the machine go?" She then showed the child object Y and asked "Does this one make the machine go?" Trial E was similar to trial D, except that in this trial the first object was the causally efficacious one.

Test trials 3 and 4. Trial 3 was an *indirect screening-off* trial. Children were shown two new objects: X and Y. The experimenter placed both objects on the machine together and the machine activated. This was demonstrated twice. The experimenter then placed object X on the machine by itself and the machine did not activate. Subsequently the experimenter used the procedure of training trial D and asked the child if each object made the machine go. Trial 4 was a *backwards blocking* trial. It was similar to trial 3, except this time when object X was placed on the machine by itself, it did activate. On these trials, the Bayes nets account predicts that on the question of whether target object Y makes the machine go, the proportion of affirmative responses on trial 4, the *backwards blocking* trial, will be smaller than the proportion of affirmative responses on trial 3, the *indirect screening-off* trial. However, causal inference based on associative models predicts an equal proportion of affirmative responses on both trials: participants would indicate that target object Y makes the machine go twice. As participants were asked to answer a question on these trials, but not to make the machine go, no prediction could be formulated for the imitation strategy.

Test trial 5. Trial 5 was a *non-causal association* trial. Children were shown three new objects: X, Y and Z. Object X was held slightly above the machine, the experimenter pressed the top of the machine with her finger and the machine activated. The same was done for object Y. Then object Z was held slightly above the machine, the experimenter placed her finger near the machine but would not press it and the machine did not activate. Subsequently the experimenter followed the procedure of training trial A and asked the child to make the machine go. On this trial, the response consistent with the Bayes net account is to press the top of the machine with a finger. However, the response consistent with causal inference based on associative models is to use an object that was associated with the effect or both objects, as they have the same associative strength (object X, object Y or objects X & Y). [9] Last, the response consistent with the imitation strategy is to hold an associated object above the machine in combination with the use of a finger (object X & finger or object Y & finger).

Statistical approach

To analyze the response patterns on the test trials Latent Class Analysis (LCA; McCutcheon, 1987) was used. It has been shown that LCA provides a statistically more reliable method to detect different types of response patterns than techniques based on matching observed response patterns with expected response patterns (Van der Maas & Straatemeier, 2008). LCA is a standard statistical technique for models with categorical, manifest variables and a categorical latent variable. Latent Class Models (LCM) describe categorically different response patterns in terms of latent classes. In the present study the classes can be interpreted as different types of causal inferences. In previous research in the field of child development LCA has been used to investigate children's 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 free classification (Raijmakers, Jansen & Van der Maas, 2004). Because several introductions to LCA exist (e.g. McCutcheon, 1987; Rindskopf, 1987), in this paper only a brief description of the technique is given.

In this study exploratory LCA was used to determine the number of latent classes needed to model the data in the best and most parsimonious manner. The modeled data consisted of children's response patterns to the test trials. To define an LCM one has to fix the number of latent classes, and subsequently estimate the unconditional probabilities and the conditional probabilities (the parameters). Unconditional probabilities define the class sizes (the frequency distribution of the different types of causal inferences). Conditional probabilities of a class indicate probabilities of responses to specific trials given membership of the class. To fit an LCM to the data, Log Likelihood estimates of the parameters are calculated with PANMARK (Van de Pol, Langeheine & De Jong, 1996). In LCA the optimal number of latent classes cannot be estimated or tested, yet model selection is based on the Bayesian Information Criterion (BIC, Schwartz, 1978); the model with the lowest BIC is considered to be the most parsimonious, best fitting model. Response patterns of classes do not need to be formulated beforehand, so the classes that are found can also represent unanticipated response patterns. Because this study's data set was small compared to the number of possible response patterns, the parametric bootstrap method (Langeheine, Pannekoek & Van de Pol, 1995) was used to determine the absolute fit of the selected model. A large, nonsignificant, bootstrapped p-value indicates a good fit of the model to the data.

RESULTS

General

In order to analyze the same data for all test trials with all techniques (standard and latent variable techniques), responses were rescored in a binary way: responses consistent with and responses inconsistent with the Bayes net account. Test trial 5 was not included in the analyses, because there was not enough variation in children's responses on his trial. Only two children (3%) had responses in line with the Bayes net account whereas the other 76 did not.[11] Preliminary χ^2 tests showed that neither the distribution of the blocks over the trials, nor the side of the detector where the causally efficacious object was placed affected the proportion of responses consistent with the Bayes net account on each of the four remaining test trials.

Comparison results to previous studies

Test trial 1, the *screening-off* trial, test trial 3, the *indirect screening-off* trial and test trial 4, the *backwards blocking* trial, had been administered with the exact same procedures in previous studies. Hence, we could compare the responses of participants in this study to those of children of comparable ages in previous studies. For the *screening-off* trial, the proportion of responses consistent with the Bayes net account did not differ between the 2-year-olds in this study and the 2-year-olds in Sobel and Kirkham's (2006) study. For the *indirect screening-off* trial and the *backwards blocking* trial, the proportion of responses consistent with the Bayes net account did not differ between the 3-year-olds in this study and the 3-year-olds in Sobel et al.'s (2004) study. For the *indirect screening-off* trial, this was also the case for the 4-year-olds in this study and the 4-year-olds in Sobel et al.'s first two experiments. For the *backwards blocking* trial, the proportion of responses consistent with the Bayes net account did not differ between the 4-year-olds in this study and the 4-year-olds in Sobel et al.'s (2004) first experiment, but did differ between the 4-year-olds in this study and the 4-year-olds in Sobel et al.'s (2004) second experiment ($\chi(1)$=7.04, p=.01). Overall, our results were highly comparable to previous results (see Table 1 for an overview).

Individual differences, Latent Class Analyses

First, it was investigated whether children demonstrated different types of causal inferences. To this end, Latent Class Models (LCM; see Method section) with 1, 2, 3, and 4 classes were fit to the data: children's binary response patterns on four causal inference trials. Table 2 shows the goodness-of-fit measures of these models. Based on the BIC values, it was found that a 3-class-model fit the data in the best and most parsimonious manner (good absolute fit: bootstrapped p-value of the difference between data and the model=.40), indicating that children demonstrated three different types of causal inferences. Table 3 shows the parameter estimates for the selected 3-class-model. The unconditional probabilities define one larger class (62%) and two smaller classes (21% and 17%).

TABLE 1. Comparison of children's responses on causal inference trials in the present study with children's responses on the same trials in previous studies; numbers and proportions of participants giving responses consistent (B) with and inconsistent (NB) with the Bayes net account.

Trial type	Age (mean age in months)	Study (experiment, year)	B (%)	NB (%)	χ^2 (df, Fisher's exact *p*-value)
Screening-off	2-year-olds (29.38)	Present study test trial 1	8 (62)	5 (39)	
	2-year-olds (24.40)	Sobel & Kirkham (1, 2006)	18 (72)	7 (28)	0.43 (1, .71)
Indirect screening-off	3-year-olds (40.71)	Present study test trial 3	20 (95)	1 (5)	
	3-year-olds (44)	Sobel et al. (1, 2004)	16 (100)	0 (0)	0.78 (1, 1)
	4-year-olds (54.39)	Present study test trial 3	22 (96)	1 (4)	
	4-year-olds (55)	Sobel et al. (1, 2004)	16 (100)	0 (0)	0.71 (1, 1)
	4-year-olds (58)	Sobel et al. (2, 2004)	15 (94)	1 (6)	0.07 (1, 1)
Backwards blocking	3-year-olds (40.71)	Present study test trial 4	14 (67)	7 (33)	
	3-year-olds (44)	Sobel et al. (1, 2004)	8 (50)	8 (50)	1.05 (1, .34)
	4-year-olds (54.39)	Present study test trial 4	22 (96)	1 (4)	
	4-year-olds (55)	Sobel et al. (1, 2004)	14 (88)	2 (13)	0.88 (1, .56)
	4-year-olds (58)	Sobel et al. (2, 2004)	10 (63)	6 (38)	7.04 (1, .01)

Note. See Method section for the description of the trials and the specific responses on each trial consistent with the Bayes net account. Note well: Sobel et al. (2004) administered both the *indirect screening-off* and the *backwards blocking* trials twice. They did not report the responses on the first and second administration separately, but did mention that these did not differ significantly. Therefore, in this Table we reported their results as if both administrations rendered exactly the same responses.

TABLE 2. Goodness-of-fit measures for Latent Class Models of responses to four causal inference trials.

Number of classes	L	df	BIC
1	-154.25	4	325.93
2	-140.48	8	311.46
3	-138.00	8	306.51
4	-137.36	4	322.65

Note. L=Log Likelihood, df=degrees of freedom, BIC=Bayesian Information Criterion. The degrees of freedom are calculated by the number of freely estimated parameters minus the number of parameters estimated at the boundary. The 4-class model is poorly identified (McCutcheon, 1987).

TABLE 3. Parameter Estimates for the selected Latent Class Model of responses to four causal inference trials.

	Unconditional probabilities	Conditional probabilities			
		Test trial 1	Test trial 2	Test trial 3	Test trial 4
Class 1	.62	1	.65	.94	1
Class 2	.21	0	0	1	.73
Class 3	.17	.67	.44	1	0

Note. The unconditional probabilities define the class sizes. The conditional probabilities indicate the probabilities of giving an answer that is consistent with the Bayes net account on a specific trial given being assigned to a specific class.

Next, by looking at the estimated conditional probabilities, the demonstrated types of causal inferences were characterized in terms of learning mechanisms or task-related strategies. As can be seen in Table 3, children in class 1 (62%) had a high probability of giving a response consistent with the Bayes net account on all four test trials. However, there were some differences between trials: children in this group were less likely to give these responses on trial 2 than on the other trials. Children in class 2 (21%) tended to give responses consistent with the Bayes net account on the last two test trials, but gave responses inconsistent with the Bayes net account on the first two test trials. A closer look at this group's responses on trial 1 and 2, showed that the majority of children (74% on trial 1 and 63% on trial 2) responded by putting the causally effective block plus one or two ineffective blocks on the machine. This pattern may indicate that these children chose to avoid uncertainty by either imitating the action of the experimenter that most frequently led to the activation of the machine or by putting all available blocks on the machine. Children in class 3 (17%) showed a similar pattern of conditional probabilities on the first two test trials as children in class 1, but they had smaller probabilities of giving responses consistent with the Bayes net account. Contrary to class 1, this group showed a pattern of responses inconsistent with the Bayes net account on trial 3 and 4. A possible characterization of this group's behavior could be causal inference based on an associative model.

Finally, it was investigated whether the different types of causal inferences identified were age-related. Based on the posterior probabilities the most likely class membership was calculated for each participant separately. Age was related to type of causal inference ($\chi(6)=29.17$, $p<.001$). Figure 1 shows the percentages of children per latent class by age group. The majority of the 2-year-olds were assigned to class 3, while the majorities of the older age groups were assigned to class 1. In all age groups, a proportion of children were assigned to class 2. For a 2-year-old the odds of being assigned to class 3, as opposed to to another class, were 17.79 times higher than for an older child. For a 5-year-old the odds of being assigned to class 1, as opposed to to another class, were 9.17 times higher than for a younger child.

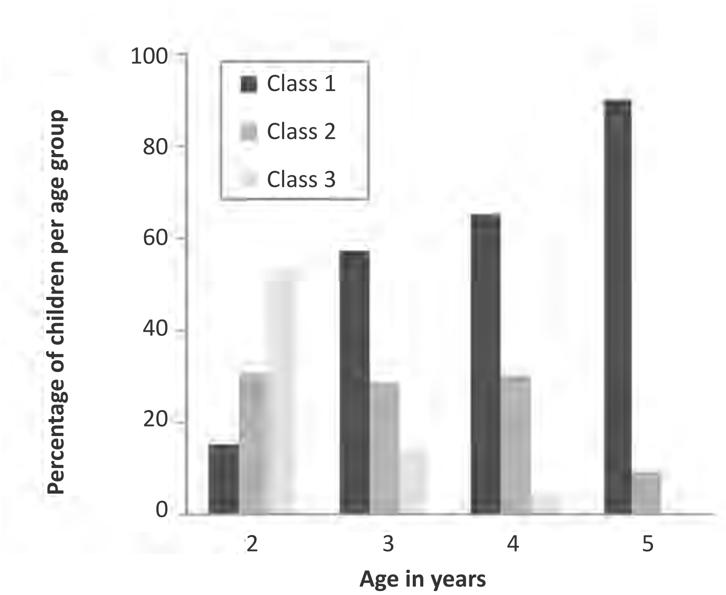

FIGURE 1. Percentages of children per latent class by age group.

DISCUSSION

The aim of the present study was to further investigate the development of children's capability for causal inference by focusing on individual differences. In particular, it was investigated whether individual differences exist in the type of causal inferences children make and whether these differences are age-related. By administering a series of carefully selected causal inference trials to children of a relatively broad age range and analyzing individuals' response patterns over trials with a categorical latent variable technique, this study has made a unique contribution to the literature. The results demonstrated the added value of the individual differences approach. Three types of causal inferences best described children's response patterns and these types were age-related. Even though the setup of this study does not allow for definite conclusions with regards to the learning mechanisms children used, examination of the literature did yield plausible explanations of the results.

The first group tended to give responses consistent with the Bayes net account (see Gopnik et al., 2004 for the outline of the account). This was a relatively large group mostly consisting of children in the older age-range. Several learning mechanisms have been proposed that might account for these responses. Gopnik et al. (2004) described two computational approaches, Bayesian methods (e.g. Glymour, 2001; Gopnik & Glymour, 2002 in Sobel et al., 2004; Steyvers, Tenebaum, Wagenmakers & Blum, 2003) and constraint-based methods (e.g. Scheines, Spirtes, Glymour & Meek, 1994 in Gopnik et al., 2004), and two psychological approaches, causal learning on the basis of associative models (e.g. Dickinson, 2001; Rescorla & Wagner, 1972; Shanks & Dickinson, 1987) and Cheng's power PC method (Cheng, 1997). Based on their review they concluded that children's learning mechanisms extended beyond the psychological methods in their current forms and were in line with the Bayesian method. Sobel et al. (2004) came to a similar conclusion. Based on children's performance on backwards blocking trials, they excluded causal learning based on associative models. To distinguish between parameter estimation models, such as Cheng's power PC method, and Bayesian methods, they designed a task in which the prior probability of an outcome had to be used to disambiguate observed data (Sobel et al., 2004). The authors considered 4-year-olds' performance on this task to go beyond the predictions of parameter estimation models. They proposed a Bayesian method making use of substantive prior knowledge as a mechanism for children's causal learning (Tenenbaum & Griffiths, 2003).

The second group's responses deviated from the first group's responses on the *screening-off* and *indirect screening-off* trials. This was a relatively small group consisting of children of all ages. The group's responses point towards the use of a task-related strategy. As children were asked to make the machine go without being told to use only one block, children in this group might have chosen to avoid uncertainty and respond in a manner that they had observed to work.

The third group mainly deviated from the first group in their pattern of responses on the *indirect screening-off* and *backwards blocking* trials. This was a relatively small group consisting of children in the younger age-range. This group's pattern of responses cannot be explained by a yes-bias, as the majority of this group (63.60 %) responded "no" when asked if block X made the machine go on trial 3. However, this group's responses are consistent with causal inference based on associative models (e.g. Dickinson, 2001; Rescorla & Wagner, 1972; Shanks & Dickinson, 1987; Sobel et al., 2004). The existence of this third group, contradicts the exclusion of causal inference based on associative models as a mechanism used for causal inference (e.g. Gopnik et al., 2004; Sobel et al., 2004). It suggests that in the development of children's causal inference, causal inference based on associative models might precede Bayesian inference. This developmental pattern is not unlikely as associative reasoning is a powerful learning mechanism in infancy (e.g. Kirkham, Slemmer & Johnson, 2002; Saffran, Aslin & Newport, 1996). Children using causal inference based on associative models either do not have a Bayesian inference system at their disposal or do not use it under certain task constraints (Sobel et al., 2004). The latter option could also be the case in the present study. As this study did not set out to investigate the minimum age at which a child is capable of using a specific mechanism for causal inference, children in the third group might be able to use Bayesian inference under different task constraints. This idea is consistent with Sobel and Kirkham's (2006) finding that 8-month-old infants' predictive inferences were consistent with the Bayes net account.

The method that was used in this study allows for disentangling different sources of variance. As described above, part of the variance in children's causal inference resulted from the different types of causal inference children applied. However, another part of the variance was caused by children making mistakes in applying these types of causal inferences. For example, this was shown by the first and third group's performance on trial 2. On the basis of the proposed learning mechanisms, children in these groups would be expected to give a response consistent with the Bayes net account. However, compared to other trials, children in both groups were less likely to do so (see Table 3). A possible explanation for these mistakes is that trial 2 places relatively high demands on a child's memory. In this trial no blocks were placed on the detector by itself, the causal structure had to be deduced from three consecutive pairings of blocks. A close look at the characteristics of the different groups also reveals a relatively high variance in the third group's responses on the first two trials (see Table 3). This group consisted of children in the younger age-range and possibly they suffered from information processing limitations or a lack of substantial prior knowledge (cf. Sobel et al., 2004).

This study introduced a new approach for investigating the development of children's causal inference: a focus on individual differences. Several recommendations for future research could be given. Trials could be chosen that enable distinguishing other possible mechanisms. For example, by using base rate trials, response patterns consistent with Bayesian inference could be distinguished from response patterns consistent with parameter estimation models (Sobel et al., 2004; Sobel & Munro, 2009). However, as argued by Steyvers et al. (2003), questions related to the mechanisms of causal inference are difficult to answer. Steyvers et al. (2003) designed a powerful paradigm to study adult causal inference processes and outcomes based observations and interventions. Taking into account psychologically reasonable representational assumptions and computationally reasonable processing constraints, they showed that adults can reason in line with the principles of optimal Bayesian decision-making (Steyvers et al., 2003). However, the individual differences within their group of adults were large. Despite the fact that their paradigm yielded a detailed insight in people's reasoning, they concluded that it was still difficult to reveal the mechanism of causal inference. The inferences they observed in their study could also be explained by relatively simple heuristics. However, the individual differences approach described in this paper is not only suited for studying the mechanisms of causal inference. For example, the approach could also be used to investigate how children's causal inference depends on the context, such as the task constraints (e.g. Lagnado & Sloman, 2004) or the domain (e.g. Schulz & Gopnik, 2004; Sobel & Munro, 2009). The approach could also be used to enable distinguishing groups that apply certain substantial prior knowledge from groups that do not apply this knowledge (e.g. Lagnado & Sloman, 2006). Moreover, children's cognitive capabilities, such as short-term memory capacity, could be related to the type of causal inference they use (cf. Schmittmann, Van der Maas & Raijmakers, in press). In short, the individual differences approach can be considered a valuable tool in future work investigating the development of causal inference.

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 daycare centers and primary school for their cooperation in the data collection. We thank research assistant Heleen Drapers for her help in collecting the data. We would like to express appreciation to Coos Hakvoort for the technical support. We also thank Ingmar Visser and Dorothy Mandell for statistical advice and comments on earlier drafts on the paper.

NOTES

1. Interventions are defined as actions that directly manipulate objects (Gopnik et al., 2004).
2. In a second experiment, Sobel & Kirkham (2006) used an eye-tracking paradigm to investigate 8-months-old's retrospective inferences. The results of this experiment are not discussed here, as they reflect, as indicated by the authors, predictive relations and not necessarily causal relations.
3. The term "Bayesian inference" was used here to refer to any mechanism that could account for responses consistent with the Bayes net account. See the Discussion section for a brief overview of the proposed mechanisms in the literature (Gopnik et al., 2004; Sobel et al., 2004).
4. In this study children using an imitation strategy were expected to imitate the action of the experimenter that most frequently led to the activation of the machine when asked to make the machine go. The results of this study partially validated this choice: a group was found that behaved relatively consistent with the predicted pattern of behavior for this imitation strategy (74% of the group on trial 1 and 63% on trial 2). Note that the used methodology for analyzing different types of causal inference ensured that our expectations about the characteristics of a possible imitation group did not influence the results (cf. Van der Maas & Straatemeier, 2008).
5. A small LED at the back of the detector hidden from the child's view indicated the state of the detector: allowing or not allowing activation.
6. Because a relatively long series of trials was administered in this study, a large number of unique objects was needed. To make sure objects were clearly distinguishable, different types of objects (wooden blocks, plastic, fruits, plastic ducks, etc.) were used. However, only one type of object was used per trial. To check for effects of specific object types, the distribution of the objects over the trials was counterbalanced.
7. In this study the trials were administered in a fixed order. First, this was done to make sure that training trials in which a response type was introduced preceded test trials in which this response type was required. For example, training trials A, B and C in which the child was asked to make the machine go preceded test trials 1 and 2 in which the same response type was required. Second, this was done to make sure that test trial 5 was always administered last. This way, the use of the finger at trial 5 could not influence children's responses on the other trials. In previous blicket detector studies order effects were investigated, but none were found (e.g. Sobel & Kirkham, 2006; Sobel et al., 2004).
8. In this study the training trials also served as control trials. A possible risk of this procedure could be that children would not be included in the analyses, based on their performance on the first training trials. Paired sample t-tests showed that this was not the case: children did not make more mistakes on the first training trial of every sort (A and D) than on the second and third training trials of every sort (B, C and E).
9. In trials A, B, C, 1 and 2 children were asked the question "What makes the machine go? Can you point to it?" before they were asked

to make the machine go. This was done to check whether responses on the go-questions were influenced by children's limited motor skills. If a child intended to respond to a go-question by putting multiple objects on the machine, but as a result of limited motor skills put the objects on the machine one by one, the experimenter could have slid the tray away after the first object and the child's reasoning would not have been captured by the trial. The pointing-questions served as a check whether this was the case. However, as a considerable part of the younger participants in this study did not understand the pointing-questions, the responses on this question were not used for further analyses.

10. In this study, children using causal inference on the basis of associative models were expected to put the object that was most associated with the effect on the machine when asked to make the machine go. On trial 1, this group was expected to use object X, as this object was associated with the effect 3/3 times, while this was 2/3 times for object Y. On trial 2, this group was expected to use object Z, as this object was associated with the effect 2/2 times, while this was 1/2 times for object X and Y. On trial 5, this group was expected to use object X and/ or Y, as these objects were associated with the effect 1/1 times, while this was 0/1 times for object Z and 2/3 times for the finger. The results of this study partially validate this choice: a group was found that behaved relatively consistent with the predicted pattern of behavior for causal inference on the basis of associative models (See Table 3). Note that the used methodology for analyzing different types of causal inference ensured that our expectations about the characteristics of a possible associative group did not influence the results (cf. Van der Maas & Straatemeier, 2008).

11. In this study, 2 (3%) children did and 76 (97%) children did not respond consistent with the Bayes nets account on the *non causal association* trial (trial 5). Of the group who did not respond consistent with the Bayes net account, 37 (49%) used one or two associated objects, 19 (25%) used the non-associated object or a combination of non-associated and associated objects, 11 (15%) used an associated object and a finger, 7 (9%) used a non-associated object and a finger and 2 (3%) gave a response that could not be scored. When administering a *non causal association* trial to 30-month-olds, Nazzi and Gopnik (2003) found a larger proportion of responses consistent with the Bayes net account (27%). These different findings can be explained by differences in procedures between both studies, such as the number of times the trial was administered and the number of objects children could choose from to make the machine go.