

Children's and Adults' Interpretation of Covariation Data: Does Symmetry of Variables Matter?

Andrea Saffran, Petra Barchfeld, and Beate Sodian
Ludwig-Maximilians-Universität München

Martha W. Alibali
University of Wisconsin-Madison

In a series of 3 experiments, the authors investigated the influence of symmetry of variables on children's and adults' data interpretation. They hypothesized that symmetrical (i.e., present/present) variables would support correct interpretations more than asymmetrical (i.e., present/absent) variables. Participants were asked to judge covariation in a series of data sets presented in contingency tables and to justify their judgments. Participants in Experiments 1 and 2 were elementary school children (Experiment 1: $n = 52$ second graders, $n = 44$ fourth graders; Experiment 2: $n = 50$ second graders). Participants in Experiment 3 were adults ($n = 62$). In Experiment 1, children in the symmetrical variables condition performed better than those in the asymmetrical variables condition. Children in the symmetrical variables condition judged more data patterns correctly and they more frequently justified their choices by referring to the complete table. Experiment 2 ruled out the possibility that this effect was caused by differences in question format. Even when question format was held constant, second graders performed better with symmetrical variables. Experiment 3 showed that adults' data interpretation is also affected by symmetry of variables. Collectively, these results indicate that symmetry of variables affects interpretation of covariation data. The authors argue that symmetrical variables provide a context for meaningful comparison. With asymmetrical variables, the importance of the comparison is less salient. Thus, the symmetry of variables should be considered by researchers as well as educators.

Keywords: scientific reasoning, evidence evaluation, statistical reasoning, development

Decision making in everyday life as well as in science depends heavily on our ability to draw appropriate causal inferences from information about covariation between events (Arkes & Rothbart, 1985; Collins & Shanks, 2002). Such intuitive statistical reasoning has been shown to be difficult even for adults (e.g., Shaklee, 1983). However, in apparent contradiction, young children and even infants have been shown to be remarkably sensitive to covariation information that they experience in the environment (e.g., Kushnir & Gopnik, 2007; Sobel & Kirkham, 2006). The task factors that contribute to the difficulty of explicit causal judgments based on covariation information are poorly understood.

One core demand of statistical reasoning tasks is comparison between conditions: Does an effect appear with equal/higher/lower frequency in the presence or absence of a critical factor? In the present article, we focus on elementary school children's and

adults' abilities to perform such comparisons in interpreting covariation data in contingency tables. We argue that the standard way of presenting the task of drawing a causal inference from a 2×2 contingency table may have obscured both children's and adults' competencies and that even young elementary school children may possess some basic competence in interpreting data in contingency tables.

Interpretation of Covariation Data

Processing of Covariation Information in Young Children

A growing body of developmental studies points to young children's sensitivity to covariation data (e.g., Gopnik, Sobel, Schulz, & Glymour, 2001; Kushnir & Gopnik, 2007; Sobel & Kirkham, 2006; Sobel, Tenenbaum, & Gopnik, 2004). For example, Gopnik and colleagues (2001) presented preschoolers with a machine that sometimes lit up and played music when novel objects ("blickets") were placed on it. They placed various combinations of objects on the machine, and children observed when the machine activated. In this paradigm, preschoolers display a rudimentary ability to use patterns of covariation data to make causal inferences. In a similar paradigm that used a looking measure, Sobel and Kirkham (2006) found that even 8-month-old infants can make inferences based on covariation data. Whereas young children use covariation information implicitly in everyday life, scientific or statistical reasoning tasks typically require an explicit judgment about a causal relation between variables.

Andrea Saffran, Petra Barchfeld, and Beate Sodian, Department of Psychology, Ludwig-Maximilians-Universität München; Martha W. Alibali, Department of Psychology, University of Wisconsin-Madison.

This research was supported by a grant from the German Research Council (DFG SO 213/31-3). The international collaboration was supported by a Friedrich Wilhelm Bessel Research Award from the Alexander von Humboldt Foundation to Martha W. Alibali. Some of these data were presented at the Biennial Meeting of the Society for Research in Child Development, Philadelphia, Pennsylvania.

Correspondence concerning this article should be addressed to Andrea Saffran, Department of Psychology, Ludwig-Maximilians-Universität München, Leopoldstrasse 13, 80802 Munich, Germany. E-mail: andrea.saffran@psy.lmu.de

Schulze and Hertwig (2016) have analyzed the difference between implicit and explicit tasks as one between experience-based and descriptive task formats. Experience-based tasks typically present information sequentially, whereas descriptive tasks present participants with summary data. Our focus in this work is on children's explicit judgments of summary data presented in contingency tables.

Children's Judgments of Covariation

Children's developing ability to interpret covariation data in descriptive task formats and to make causal judgments based on their interpretations is not well understood. Although older research (e.g., Shaklee & Paszek, 1985; Shaklee & Tucker, 1980) showed severe deficits in children's and adults' ability to make causal judgments based on data tables, recent studies show some competence in very simple tasks even in preschool children (e.g., Koerber, Sodian, Thoermer, & Nett, 2005).

Preschoolers display a basic competence in interpreting patterns of perfect or near perfect covariation, especially when each data instance is depicted in a separate picture. For example, Koerber and colleagues (2005) asked children to decide whether green or red chewing gum causes bad teeth based on 10 pictures showing green or red chewing gum and children with good or bad teeth. They found that 4-year-olds could interpret patterns of perfect and near perfect covariation quite well, but even 6-year-olds had problems interpreting noncovariation without support. These results are in line with other findings that 4- and 5-year-olds (Piekný, Grube, & Maehler, 2014; Ruffman, Perner, Olson, & Doherty, 1993) as well as 8- and 12-year-olds (Amsel & Brock, 1996) interpreted perfect covariation without problems, but had problems with non-covariation (Amsel & Brock, 1996; Piekný et al., 2014). On the continuum from experience-based to descriptive tasks (see Schulze & Hertwig, 2016), the individual-picture presentation format used in these studies seems to be closer to experience-based formats, in contrast to descriptive, summary table formats.

Children's interpretations of more complex patterns presented in summary format were investigated in two studies by Shaklee and her colleagues (Shaklee & Mims, 1981; Shaklee & Paszek, 1985). In both studies, data were presented in 2×2 contingency tables, with the cells labeled as A, B, C, and D (see Figure 1). Data were embedded into fictive contexts such as the possible relation between snowing and the happiness of space creatures.

Shaklee and Mims (1981) focused on the development of strategy use in fourth, seventh and tenth graders and college students.

They identified four possible strategies, which can be ordered hierarchically (see Table 1). The least sophisticated strategy was the *Cell A strategy* (i.e., only Cell A of the contingency table is considered). In the given example about space creatures, in the Cell A strategy, the covariation judgment is based only on the number of cases in which it is snowing and the space creatures are happy. The next most sophisticated strategy is called the *A versus B strategy* (i.e., Cell A and B of the contingency table are compared). In the space creature example, in the A versus B strategy, the covariation judgment is based on the comparison between the number of cases in which it is snowing and space creatures are happy and the number of cases in which it is snowing and space creatures are sad. In the *Sum of Diagonals* strategy, the sum of the confirming cases (i.e., one diagonal) is compared to the sum of the disconfirming cases (i.e., the other diagonal). In the Sum of Diagonals strategy, the solver compares the sum of cases in which it is snowing and space creatures are happy and in which it is not snowing and space creatures are sad (cases confirming a positive relationship) against the sum of cases in which it is snowing and space creatures are sad and in which it is not snowing and space creatures are happy (cases disconfirming a positive relationship). A relationship is assumed if the sum of confirming cases exceeds the sum of disconfirming cases. The most sophisticated strategy is the *Conditional Probability* strategy, in which two conditional probabilities are calculated and compared. In the Conditional Probability strategy, a solver compares the proportion of cases with happy space creatures under the condition of snowing to the proportion of cases with happy space creatures under the condition of not snowing. Shaklee and Mims structured their data sets so that some items could be correctly solved by all of the strategies, whereas others could be correctly solved only by more advanced strategies (i.e., Sum of Diagonals, Conditional Probability). Based on their solution patterns, participants were classified as using one of the four strategies.

Shaklee and Mims (1981) found that judgment accuracy depended on problem type and age group. The most common strategies were the A versus B strategy and the Sum of Diagonals strategy. Among fourth graders, more children were classified as A versus B strategy users than Sum of Diagonals strategy users. Among seventh and tenth graders, this pattern was reversed.

Shaklee and Paszek (1985) used the same paradigm but focused on younger children from Grades 2, 3, and 4. They found that even second graders could make judgments based on data summarized in contingency tables. The most common strategy in all three age groups was the A versus B strategy. Only 2.7% of the second graders, 5.5% of the third graders and 5.9% of the fourth graders were classified as Sum of Diagonals strategy users, and no child solved at least two of three Conditional Probability problems correctly.

In summary, children can draw conclusions from simple data patterns such as perfect covariation with data points depicted in individual pictures (e.g., Koerber et al., 2005), but they have difficulties when explicit judgments about more complex data patterns based on summary data are required. Children fail to consider all relevant information and tend to make their judgments based on simple comparisons (Shaklee & Mims, 1981; Shaklee & Paszek, 1985).

		Outcome	
		Level 1	Level 2
Factor	Level 1	A	B
	Level 2	C	D

Figure 1. Labeled 2×2 contingency table.

Table 1
Hierarchy of Strategies According to Shaklee and Mims (1981)

Strategy	Calculation	Proposed by ^a
Cell A	Focus on cell A	Inhelder and Piaget (1958); Smedslund (1963)
A vs. B	Comparison of cells A and B	Inhelder and Piaget (1958)
Sum of Diagonals	Comparison (A + D) and (B + C)	Inhelder and Piaget (1958)
Conditional Probability	Comparison $A/(A + B)$ and $C/(C + D)$	Jenkins and Ward (1965)

^a As cited in Shaklee & Mims (1981).

Adults' Judgments of Covariation

Whereas studies on children's data interpretation abilities are rare, adults' interpretations of covariation data, typically presented in summary form, have been studied extensively. Several weaknesses in adults' reasoning have been identified. Judgment accuracy is often poor (e.g., Arkes & Harkness, 1983; Batanero, Estepa, Godino, & Green, 1996; Shaklee & Elek, 1988; Smedslund, 1963; Ward & Jenkins, 1965) and inadequate strategies are common (e.g., Batanero et al., 1996; Mata, Garcia-Marques, Ferreira, & Mendonça, 2015; Shaklee & Mims, 1982; Shaklee & Tucker, 1980; Shaklee & Wasserman, 1986; Smedslund, 1963). For instance, Shaklee and Elek (1988) tested adolescents and college students with a covariation interpretation task. They used 12 problems presented in 2×2 contingency tables and embedded in everyday contexts. The contexts ranged from risen or fallen bakery products as related to the presence or absence of an ingredient, healthy or sick plants and the presence or absence of some kind of plant food, healthy or sick animals and the presence or absence of some kind of medicine, and happy or sad space creatures and the presence or absence of one of three weather conditions. To identify strategies, they used the same procedure as Shaklee and Mims (1981, described above). Their results showed low judgment accuracy for problems that required more advanced strategies (i.e., Sum of Diagonals, Conditional Probability). The most common strategies for adolescents were the Cell A and A versus B strategies, and for college students the A versus B strategy. This pattern of findings indicates that the majority of participants did not consider the information from all four cells. Klayman and Ha (1987) referred to this focus on the first row of a contingency table as "positive hypothesis testing."

This interpretation was supported by studies that used a set of covariation problems in which problems could be examined in pairs to assess the impact of each cell on participants' judgments (Kao & Wasserman, 1993; Levin, Wasserman, & Kao, 1993; Wasserman, Dorner, & Kao, 1990). These studies revealed that adults' judgments are not affected equally by each of the four cells in a contingency table, but rather they are weighted in the order A, B, C, D.

To sum up, the difficulties identified in children's interpretations of summary covariation data seem to persist until adulthood. Even adults do not integrate the information from the four cells of a contingency table in an adequate way. Like children, adults often draw conclusions from only one critical comparison of conditions and do not consider all four cells as equally important for their judgments.

Symmetry of Variables

The present research investigates the reasons why both children and adults find it so difficult to perform the critical comparisons when analyzing data presented in contingency tables. Our hypothesis is that the standard way of framing data interpretation tasks may obscure the necessary comparisons between conditions. In most of the studies mentioned above, participants had to decide whether a factor has an effect or not (e.g., "Is fertilizer good for plants?", with fertilizer present vs. fertilizer absent; Amsel & Brock, 1996; Arkes & Harkness, 1983; Kao & Wasserman, 1993; Levin et al., 1993; Shaklee & Wasserman, 1986; Smedslund, 1963; Ward & Jenkins, 1965; Wasserman et al., 1990; Wasserman & Shaklee, 1984). Only a few studies have used tasks that require a choice between two alternative levels of one factor (e.g., "Does Fertilizer A have a stronger effect than Fertilizer B?"; Batanero et al., 1996; Shaklee & Mims, 1982; Shaklee & Tucker, 1980), although such choices are frequently required in everyday life.

The difference between judging the effectiveness of a factor and comparing the effectiveness of two levels of a factor may influence covariation judgments. The comparison between the two rows of a contingency table, which is crucial for judging covariation adequately (i.e., applying the Conditional Probability strategy), seems to be more salient in the second case. It may be easier for participants to neglect the data that arise when a treatment is omitted (e.g., no fertilizer) than to neglect that data that arise when a treatment is applied (e.g., Fertilizer B). That is, participants may be less likely to attend to an omission than to an action.

To date, only one study has addressed the effects of the symmetry of variables. Beyth-Marom (1982) described present-absent variables as *asymmetrical* and present-present variables as *symmetrical*. An example of an asymmetrical variable is a fertilizer that is either used or not used, whereas an example of a symmetrical variable is two types of fertilizer that are both used. Beyth-Marom systematically investigated the effect of symmetry of variables on adults' concept of correlation, expecting more serious misinterpretations of this concept in the asymmetrical case than in the symmetrical one. She argued that in the asymmetrical case, the absent value has a lower "status" than the present value, whereas both values have a similar "status" in the symmetrical case. This difference is reflected in the labels of the two values of a variable. In the asymmetrical case, the present value is typically labeled with the same word as the variable itself, whereas the absent value is labeled with its negation. For example, the present value of the asymmetrical variable fertilizer is labeled with "fertilizer" and the absent value is labeled with "no fertilizer." In comparison, for symmetrical variables, either both values include the name of the

variable (e.g., variable fertilizer: fertilizer A vs. fertilizer B) or neither one of them does (e.g., variable gender: male vs. female).

To investigate the effects of symmetry of variables on understanding of correlation, *Beyth-Marom (1982)* presented adult participants with sentences about a strong relationship between two variables that were either asymmetrical or symmetrical. For example, the introductory sentence in the asymmetrical condition was, “A paper published in a major medical journal reported that for one species of animals a strong relationship was found between a specific symptom and a specific disease” (*Beyth-Marom, 1982, p. 514*). Participants had to choose one out of five possible interpretations, which can be characterized in terms of the cells of a traditionally labeled 2×2 contingency table (see *Table 2*).

A similar proportion of participants in both conditions interpreted the correlation correctly, in terms of conditional probabilities. However, there were different distributions of incorrect interpretations: In the asymmetrical case, the incorrect interpretations corresponded to the Cell A and A versus B strategies, whereas in the symmetrical case, the incorrect interpretations reflected the Sum of Diagonals strategy. Thus, in the asymmetrical case, participants interpreted the sentences about the relationship more narrowly than in the symmetrical case. Although this finding lends some support to our hypothesis that symmetrical tasks may facilitate contingency table analysis, it is unclear whether one can generalize from *Beyth-Marom’s (1982)* sentence-choice task to tasks requiring actual data interpretation.

The Present Study

In summary, the reviewed literature shows, on the one hand, that infants and young children are sensitive to experiencing covariation (e.g., *Gopnik et al., 2001; Sobel & Kirkham, 2006*) and that preschoolers are able to draw explicit causal inferences from simple covariation patterns presented via individual pictures representing each data point (e.g., *Koerber et al., 2005*). On the other hand, elementary school children as well as adolescents and adults have substantial difficulties interpreting data summarized in contingency tables (e.g., *Shaklee & Paszek, 1985; Shaklee & Tucker, 1980; Smedslund, 1963*). However, it is not clear how task characteristics contribute to these problems, or which factors might facilitate data interpretation. There is some support for the idea that the framing of the problem in terms of the asymmetry or symmetry of variables may affect comparisons between conditions in the interpretation of contingency tables. However, there have been no studies that have directly examined the influence of symmetry of variables on data interpretation in children or adults.

The present research addresses the following questions: First, does the symmetry of variables affect interpretations of covariation data in children and adults? We hypothesize that participants will display more accurate interpretations and use more sophisticated strategies when the data involve symmetrical variables. Second, does the magnitude of the symmetry of variables effect depend on age? One could hypothesize that the symmetry effect would decrease with age, as participants become more likely to successfully interpret contingency tables, and less susceptible to variations in the task conditions. However, given the extensive literature on adults’ difficulties in data interpretation tasks, it is unclear if this argument applies. Therefore, we consider our investigation of age-related differences in the symmetry of variables effect to be exploratory, and we do not advance a specific hypothesis regarding the interaction of symmetry of variables and age.

To address these research questions, we investigated the symmetry of variables effect in both children and adults using a data interpretation task with summary table format. Because no developmental studies have investigated symmetry of variables, we focus on elementary school children in Experiments 1 and 2. In Experiment 3, we present data on adults in order to extend *Beyth-Marom’s (1982)* results to a data interpretation task.

In this research, we asked participants to explain how they reached their judgments for each covariation problem, and we inferred their strategies from these justifications. We chose this approach in light of limitations of other methods for assessing strategy use. First, *Shaklee’s* approach to identifying strategies on the basis of solution patterns across problems (e.g., *Shaklee & Mims, 1981; Shaklee & Paszek, 1985*) has the drawback that a high proportion of (especially younger) participants’ solution patterns do not fit one of the proposed strategy categories (e.g., 25% of the fourth graders in *Shaklee & Mims, 1981*). Second, there are alternative strategies that could produce the same solution patterns as the proposed ones. For example, the pattern from which the Sum of Diagonals strategy was inferred could also be produced by the A versus C comparison strategy (*Shaklee & Mims, 1981*). Thus, it is not clear whether participants really used the sophisticated Sum of Diagonals strategy or only used a two-cell comparison. Third, the pattern analysis approach is based on the assumption that people use a single strategy consistently across problems, and this assumption may be unfounded. Some adult studies have asked participants to provide a justification after each judgment, and these studies point to shifts in strategy use depending on the item structure (*Batanero et al., 1996; Mata, Ferreira, & Sherman, 2013*). For these reasons, rule modeling on the basis of solution

Table 2
Sample Interpretations of a Correlation as Used by Beyth-Marom (1982, p. 514)

Strategy	Sample sentence
Cell A	“Among all animals examined, many had the symptom and the disease.”
A vs. B	“Among all animals examined with the disease, there was a higher percentage with the symptom than without it.”
A + D	“Among all animals examined, there were many animals either with the disease and the symptom or without both.”
(A + D) vs. (B + C)	“Among all animals examined, there was a higher percentage of either animals with the symptom and the disease or animals without both <i>than</i> either animals with the disease, but without the symptom or animals without the disease but with the symptom.”
Conditional Probabilities	“The percentage of animals with the symptom among animals with the disease was higher than the percentage of animals with the symptom among animals without the disease.”

patterns does not seem to be the optimal approach with which to identify the strategies people use in interpreting contingency tables. We chose to ask participants to provide verbal explanations, because past research has shown that verbal reports yield valid information about children's strategy use in other mathematics tasks (e.g., Reed, Stevenson, Broens-Paffen, Kirschner, & Jolles, 2015; Robinson, 2001). Verbal reports have also been used in past research on adults' interpretations of contingency tables (Batanero et al., 1996; Mata et al., 2013, 2015).

Experiment 1

The purpose of Experiment 1 was to investigate whether children's interpretations of covariation data presented in contingency tables are influenced by the symmetry of variables and whether the magnitude of the symmetry of variables effect depends on age.

Beyth-Marom (1982) found that adults misinterpret relationships between asymmetrical variables more narrowly than relationships between symmetrical variables. Based on this finding, we hypothesized that children would perform better when interpreting covariation data with symmetrical variables than when interpreting covariation data with asymmetrical variables. This difference should be reflected in the accuracy of their judgments and in their justifications.

Method

Participants. Participants were 96 children, including 52 second graders (46.2% male; mean age $M = 8.23$ years, $SD = .33$, range = 7.47–8.98) and 44 fourth graders (45.5% male; mean age $M = 10.27$ years, $SD = .46$, range: 9.32–11.80). Children were recruited from four elementary schools in Munich, Germany.

Design. The experiment used a 2×2 design, with symmetry of variables and grade level as between-subjects factors. A paper-and-pencil pretest was conducted in class 4 to 10 weeks before the individual testing sessions. In this pretest, children were asked to judge eight covariation data sets. The problems were embedded in a context about crèmes against pimples and children were asked to indicate for each covariation problem which of two crèmes was more effective or if there is no difference (i.e., to choose one out of three options). For the study proper, within each grade level, the experimental groups were matched based on their mean pretest result and on gender. In this way we ensured that the experimental groups were comparable with respect to their data interpretation abilities. Table 3 shows the number of participants in each group.

Materials

Two series of nine pictures with 2×2 contingency tables were used. The context story was about the effectiveness of fertilizers.

Table 3
Number of Participants in Each Condition

Condition	Grade level		All
	2nd	4th	
Asymmetrical	26	21	47
Symmetrical	26	23	49
All	52	44	96

The rows and columns of the tables were labeled with small illustrations, indicating the levels of the independent and dependent variables. The rows in the asymmetrical variables condition were labeled *fertilizer present* and *fertilizer absent*, whereas the rows in the symmetrical variables condition were labeled *fertilizer A* and *fertilizer B*. The columns were labeled *plant is growing* and *plant is dying* in both conditions. The cell frequencies were depicted with small illustrations and numbers (see Figure 2).

The data in the tables were the same in both conditions. The data sets with their corresponding χ^2 values are presented in Figure 3. The data sets were developed to be more complex than the ones used in recent studies (e.g., Koerber et al., 2005). Thus, for most data sets it was not sufficient to attend only to the first row (equals A vs. B strategy) or to the first column (equals A vs. C strategy). Numbers ranged from 1 to 30. Three data sets (Items 5, 6, 9) showed no covariation, while all other items showed either positive (Items 1, 4, 7) or negative (Items 2, 3, 8) relationships. The numerical structure differed with respect to several characteristics. In some data sets, the relation between values was very salient. Five data sets included the same cell value twice, either in the first row (Item 2), in the first column (Item 3), in both rows (Item 5), in both columns (Item 6), or in one diagonal (Item 9). In addition, values in Item 9 were simple multiples. For the other items, relations were not as salient but could easily be estimated (e.g., 13 is about 2×6 in Item 1). There was the same difference between the cells in the two rows and the cells in the two columns for Items 5, 6, and 8, while this difference varied for the other data sets.

There were two orders in which the covariation problems were presented (order A: 1, 5, 3, 9, 4, 8, 6, 7, 2; order B: 1, 2, 7, 6, 8, 4, 9, 3, 5). Half of the children in each group received order A and the other half order B.

Procedure. Children were interviewed individually in a quiet place at their school. The items were presented on a laptop screen and the interviews were video-recorded and transcribed. Children in both conditions were told that they would hear a story about scientists and that there would be questions about the scientists and their data tables.

In the asymmetrical variables condition, the interview started with the context story about nine scientists who wanted to test their fertilizers. Children were told that each scientist gave his fertilizer to some plants and he did not give it to some other plants. After a few weeks, he checked if each plant grew well or died. After this introduction, the interviewer explained the meaning of the rows and columns based on a sample contingency table, which did not contain data:

Each scientist writes down his observations in such a table. In the first row (interviewer pointed to the first row), he notes plants that got the fertilizer and in the second row (interviewer pointed to the second row), he notes plants that did not get the fertilizer. In the first column (interviewer pointed to the first column), he notes plants that grew well and in the second column (interviewer pointed to the second column), he notes plants that died.

Two control questions were administered to make sure that the participants understood the meaning of the cells. A child who failed these questions received a second explanation of the table and a second set of control questions. After passing the control questions, the experiments and results of the nine scientists were presented one at a time. For each table, the children were asked for

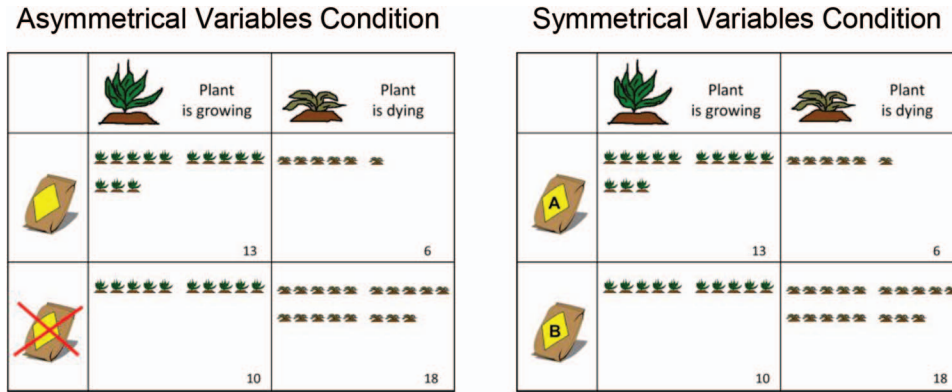


Figure 2. Sample items for the asymmetrical and symmetrical variables conditions in Experiments 1 and 2. See the online article for the color version of this figure.

a yes/no judgment: “Is his fertilizer good for plants?” and a justification: “Where do you see in the table that his fertilizer is (not) good for plants?”

In the symmetrical variables condition, the procedure differed in the following aspects. In this case, each of the scientists had invented two fertilizers and wanted to test which one was better for plants or whether there was no difference. This difference is reflected in the judgment question: “Is fertilizer A or fertilizer B better for plants or is there no difference?”

Coding. Children’s judgments were coded as correct or incorrect. Children’s justifications were analyzed in terms of the strategy they expressed. We decided to not restrict our coding scheme to the strategies postulated by Shaklee and colleagues (see Table 1) but to cluster children’s justifications in a bottom-up manner. This approach yields the advantage that strategies that were neglected in previous work could be identified. As described in Table 4, the justifications were categorized into three main categories based on the number of cells they referred to (four-cell justifications, two-cell justifications, other). These main categories were further divided into several subcategories.

The comparison of ratios category encompasses the traditional Conditional Probability strategy and also comparisons of simple proportions (e.g., A/C vs. C/D). Note that correct comparison of ratios leads to the correct judgment for all covariation data patterns. The comparison of differences category includes the Sum of Diagonals strategy and also row- and column-wise comparisons of differences. Justifications in this subcategory lead to the correct judgment for only a subset of the data patterns. Justifications in the comparison of marginal sums category are not adequate for any data patterns but might lead to the correct judgment by chance. Subcategories of two-cell justifications include attending to the first row, which corresponds to the A versus B strategy, as well as

attending to the other row and both columns. Attending to only two of the four cells is not adequate for any data pattern. However, comparing values of two cells, especially of the second row or column, can still be appropriate to explain a correct judgment in some data sets used in the present study (e.g., comparing cell B and cell D in Item 8).

All justifications were coded by two raters independently. Interrater agreement varied from 90.4% to 100% for the main categories, from 88.9% to 100% for the four-cell subcategories, and from 96.8% to 100% for the two-cell subcategories. Disagreements were resolved by discussion.

Results

All but one of the children passed both control questions after the first explanation of the table; the remaining child passed the control questions only after a second explanation of the table. There was no effect of gender on judgments (range: 0–1; male: $M = .66$, $SD = .26$; female: $M = .61$, $SD = .26$; $F(1, 94) = .83$, $p = .365$, $\eta_p^2 = .01$). However, gender was associated with justifications. Compared to girls, boys produced four-cell justifications more often (range: 0–9; male: $M = 6.30$, $SD = 3.00$; female: $M = 4.88$, $SD = 3.28$; $F(1, 93) = 4.74$, $p = .032$, $\eta_p^2 = .05$). There were no gender differences for any other justification category. With respect to the order of items, there was no effect on judgments (range: 0–1; order A: $M = .63$, $SD = .27$; order B: $M = .64$, $SD = .25$; $F(1, 94) < .01$, $p = .964$, $\eta_p^2 < .01$) but order did influence children’s justifications. Children with order A reported more four-cell justifications (range: 0–9; order A: $M = 6.35$, $SD = 2.82$; order B: $M = 4.70$, $SD = 3.41$; $F(1, 93) = 6.63$, $p = .012$, $\eta_p^2 = .07$) and fewer two-cell justifications (range: 0–9; order A: $M =$

Item	1	2	3	4	5	6	7	8	9
Cell frequencies	13 6 30 30	18 15 6 7	8 8 24 12	15 10 30 11	20 10 20 1	10 5			
χ^2 values	4.84	4.99	4.34	1.93	0	0	4.18	4.32	0

Figure 3. Cell frequencies and χ^2 values for the nine covariation data sets from Experiments 1 and 2.

Table 4
Main and Subcategories With Sample Justifications

Category	Justification type	Sample justification
Four-cell justifications	Four cells are mentioned	
Comparison of ratios	Ratios are compared	“Because there (<i>cell A and B</i>) are 20 good ones and half of them are not good, and in the second row there are also half that are not good”
Comparison of differences	Differences are compared (no proportional reasoning)	“Because here (<i>cell D</i>) are 5 more than here (<i>cell B</i>) and here (<i>cell A</i>) are 10 more than here (<i>cell C</i>)”
Comparison of marginal sums	Marginal sums are compared	“Because here (<i>cell A and C</i>) are together more growing plants than here (<i>cell B and D</i>)”
Other 4 cells	Does not fit in one of the four-cell subcategories	
Two-cell justifications	Two cells are mentioned	
First row	Reference only to the first row (cells A and B)	“Because there (<i>cell A</i>) are more than there (<i>cell B</i>)”
Second row	Reference only to the second row (cells C and D)	“Because 18 (<i>cell D</i>) is more than 10 (<i>cell C</i>)”
First column	Reference only to the first column (cells A and C)	“Because of those two (<i>cell A and C</i>) 13 is the larger number”
Second column	Reference only to the second column (cells B and D)	“Because more plants died without fertilizer than with fertilizer”
Other 2 cells	Does not fit in one of the two-cell subcategories	
Other	Does not fit in either the four-cell nor the two-cell category	

Note. Sample justifications for four-cell subcategories are from Item 9 (20/10/10/5), sample justifications for two-cell subcategories are from Item 1 (13/6/10/18).

2.35, $SD = 2.70$; order B: $M = 3.77$, $SD = 3.35$; $F(1, 93) = 5.12$, $p = .026$, $\eta_p^2 = .05$).

The proportion of correct judgments (range: 0–1) in each condition for participants in each grade level is presented in Figure 4. Please note that the probability of answering correctly by chance was not the same in both conditions, because the questions implied different answer alternatives (asymmetrical variables: yes/no, probability of chance success = .50; symmetrical variables: fertilizer A/fertilizer B/no difference; probability of chance success = .33). Despite the lower probability of chance success, children were more accurate in the symmetrical variables condition, $F(1, 92) = 29.35$, $p < .001$, $\eta_p^2 = .24$. In addition, fourth-grade students were more successful than second-grade students, $F(1, 92) = 15.81$, $p < .001$, $\eta_p^2 = .15$. There was also a significant interaction; the effect of symmetry of variables was greater among second graders than among fourth graders, $F(1, 92) = 4.57$, $p = .035$, $\eta_p^2 = .05$. As seen in Figure 4, second graders’ performance

suffered more from asymmetrical variables than did fourth graders’.

Table 5 presents the proportion of correct judgments for each item over both grade levels. For the comparison between conditions, we report p values for one-tailed, exact χ^2 tests, as Bühner and Ziegler (2009) suggested. There were significant effects for

Table 5
Proportion of Correct Judgments (in %) and χ^2 Test for Each Item

Item	Proportion of correct judgments		χ^2 test				
	Asymmetrical variables	Symmetrical variables	$\chi^2(df = 1)$	p (1-tailed)	w		
1	13 10	6 18	95.7	100	2.13	.237	.15
2	30 27	30 10	46.8	77.6	9.67**	.002	.32
3	18 18	15 4	53.2	87.8	13.87**	<.001	.38
4	6 10	7 29	38.3	71.4	10.65**	.001	.33
5	8 16	8 16	76.6	75.5	0.02	.546	.01
6	24 24	12 12	53.2	100	29.76**	<.001	.56
7	15 16	10 30	68.1	87.8	5.43*	.018	.24
8	30 20	11 1	6.4	42.9	17.02**	<.001	.42
9	20 10	10 5	21.3	36.7	2.78	.074	.17

* $p < .05$. ** $p < .01$.

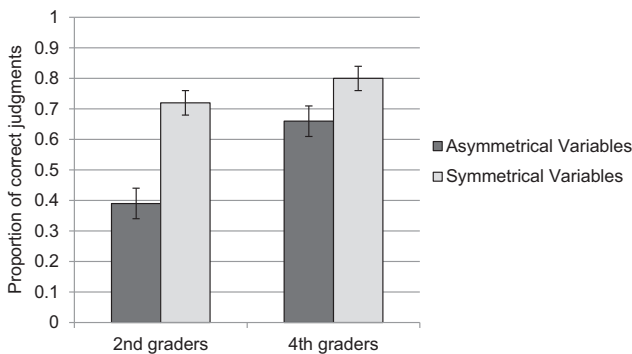


Figure 4. Average proportion of correct judgments for both conditions and grade levels. Error bars represent standard errors.

Items 2, 3, 4, 6, 7, and 8, and no significant effects for Items 1, 5, and 9. All significant differences were in favor of the symmetrical variables condition (see Table 5).

Children's justifications of their judgments were analyzed to better understand the differences between conditions. One second grader in the symmetrical variables group was excluded because some justifications were not audible and thus could not be transcribed. We focus first on the main categories of four- and two-cell justifications, and we then present more detailed analyses concerning four-cell justifications as well as two-cell justifications.

Justifications were initially coded as four-cell justifications, two-cell justifications, or other. Table 6 presents the mean number of justifications in each category for both conditions. A multivariate analysis of variance (ANOVA) with condition as factor and the three main categories as dependent variables showed two significant effects: Four-cell justifications occurred more frequently in the symmetrical variables condition, and the residual (other) category occurred more frequently in the asymmetrical variables condition (see Table 6). Grade level was excluded from this analysis because there were no significant effects of grade level on any of the main categories (four-cell justifications: $F(1, 93) = 1.71, p = .194, \eta_p^2 = .02$; two-cell justifications: $F(1, 93) = 1.03, p = .313, \eta_p^2 = .01$; other: $F(1, 93) = .92, p = .340, \eta_p^2 = .01$).

We examined whether four-cell justifications tended to lead to correct judgments. The data from further 11 children had to be excluded from this analysis because they did not report any four-cell justifications. The probability of correct judgments given four-cell justifications was calculated for each condition. This value was significantly higher for the symmetrical variables condition (asymmetrical: $M = .56, SD = .34$; symmetrical: $M = .85, SD = .14$; $F(1, 82) = 28.18, p < .001, \eta_p^2 = .26$). Thus, four-cell justifications are associated with correct judgments, but more so in the symmetrical variables condition than in the asymmetrical variables condition.

To follow up this effect, we classified the four-cell justifications into four subcategories: comparison of ratios, comparison of differences, comparison of marginal sums, and other four-cell justifications. To consider differences between conditions with respect to those subcategories, a multivariate analysis of covariance with condition as factor was calculated. The number of four-cell justifications was included as a covariate because children produced such justifications more often in the symmetrical variables condition, and the different baselines of four-cell justifications for the experimental groups would influence the prevalence of the four-cell subcategories in the two groups. Comparison of differences occurred more often in the symmetrical variables condition, while comparison of marginal sums occurred more often in the asymmetrical variables condition (see Table 6). It seems that symmetrical variables led not only to more four-cell justifications, but also to more sophisticated four-cell justifications.

We classified two-cell justifications into five subcategories: first row (A vs. B), second row (C vs. D), first column (A vs. C), second column (B vs. D), and other two-cell justifications. A multivariate analysis of covariance with condition as factor was conducted. The number of two-cell justifications was included as a covariate because the different baselines for the experimental groups might influence the prevalence of the subcategories in the two groups. This analysis indicated several significant effects: Children in the asymmetrical variables condition produced more first row justifications than did children in the symmetrical variables condition. In contrast, children in the symmetrical variables condition produced more first column and second column justifications than did children in the asymmetrical variables condition (see Table 6). This data pattern suggests that children tend to reason more row-wise with asymmetrical variables and more column-wise with symmetrical variables.

Table 6
Mean Values (Range: 0–9; SD in Parentheses) for Each Justification Category for Both Conditions and Corresponding Analysis of Variance

Category	Condition		Analysis of variance ^a		
	Asymmetrical	Symmetrical	$F(df_c, df_e)$	p	η_p^2
Four-cell justifications	4.87 (3.25)	6.19 (3.09)	4.10* (1, 93)	.046	.04
Comparison of ratios	.26 (.71)	.35 (.60)	<.01 (1, 92)	.998	<.01
Comparison of differences	3.70 (2.86)	5.52 (2.80)	7.42** (1, 92)	.008	.08
Comparison of marginal sums	.70 (1.84)	.21 (.50)	5.67* (1, 92)	.019	.06
Other 4 cells	.21 (.66)	.10 (.31)	1.88 (1, 92)	.173	.02
Two-cell justifications	3.45 (3.06)	2.67 (3.13)	1.51 (1, 93)	.223	.02
First row	1.81 (2.46)	.06 (.32)	22.70** (1, 92)	<.001	.20
Second row	.02 (.15)	.15 (.51)	3.06 (1, 92)	.084	.03
First column	.94 (1.95)	1.29 (2.48)	4.13* (1, 92)	.045	.04
Second column	.51 (1.32)	1.15 (1.71)	8.11** (1, 92)	.005	.08
Other 2 cells	.17 (1.03)	.02 (.14)	.59 (1, 92)	.444	<.01
Other	.68 (1.45)	.15 (.41)	6.07* (1, 93)	.016	.06

Note. df_c = degrees of freedom for condition; df_e = degrees of freedom for error.

^a Three multivariate analyses of variance were calculated. The first included the main categories as dependent variable. The second included the four-cell subcategories as dependent variables and four-cell justifications as covariate, and the third included the two-cell justification subcategories as dependent variables and two-cell justifications as covariate.

* $p < .05$. ** $p < .01$.

Discussion

The symmetry of variables influences elementary school children's interpretations of covariation data presented in summary form in contingency tables. This influence is reflected in the accuracy of children's judgments as well as in their justifications. Children draw more correct inferences when interpreting covariation data that involve symmetrical variables. This was true for the average proportion of correct judgments as well as for six out of nine individual items (Items 2, 3, 4, 6, 7, and 8). Furthermore, children used four-cell strategies more often, they used more sophisticated four-cell strategies, and they reasoned based on columns rather than on rows more often when analyzing relations between symmetrical variables than when analyzing relations between asymmetrical variables.

The data suggest that symmetrical variables foster column-wise reasoning and four-cell comparisons, while asymmetrical variables lead to row-wise reasoning. This pattern is especially striking when contrasting Items 5 and 6, both of which show no effect (Item 5: 8/8/16/16; Item 6: 24/12/24/12). Item 5 could be solved by row-wise two-cell comparisons or by four-cell comparisons; on this item, children in both conditions performed equally well. In contrast, Item 6, which required column-wise reasoning or four-cell comparisons, led to a large difference in favor of the symmetrical variables condition. Our data suggest that the two present levels of the symmetrical variables emphasize the importance of all four cells and lead to column-wise comparisons, while the one present level of the asymmetrical variables attracts children's attention, and they consequently neglect the absent level.

However, an alternative explanation is that the difference between conditions was caused by the differing question formats. The question in the asymmetrical variables condition ("Is the fertilizer good for plants?") aims only implicitly at the comparison between the rows, while the question in the symmetrical variables condition ("Is fertilizer A better for plants or is fertilizer B better for plants or is there no difference?") explicitly highlights the comparison between the two types of fertilizer. To address this possibility, in Experiment 2, we examine children's performance in an asymmetrical variables condition with an explicit comparison question.

Experiment 2

Experiment 2 was designed to rule out the possibility that the symmetry effect found in Experiment 1 was caused by differences in question format between the asymmetrical and the symmetrical variables condition and not by the symmetry of variables itself. If the effect is caused by the symmetry of variables, it should be found even if the question format is held constant. We focused on children in Grade 2 because the symmetry effect was most pronounced at this age.

Method

Participants. Twenty-four second graders were interviewed. They were recruited from four elementary schools in Munich, Germany. We compared their data to that of the second graders in the symmetrical variables condition from Experiment 1 ($n = 26$). Thus, the total sample consisted of 50 second graders (46.0% male; mean age $M = 8.25$ years, $SD = .29$, range: 7.76–8.77).

Design. The additional group was matched with the existing group based on gender and performance on the pretest of contingency table interpretation (see Experiment 1 for a description of the pretest).

Materials and procedure. The materials in the new asymmetrical variables condition were the same as in the asymmetrical variables condition from Experiment 1 (see Figure 2). A series of nine pictures showed covariation data in 2×2 contingency tables. The rows were labeled with *fertilizer present* versus *fertilizer absent* and the columns with *plant is growing* versus *plant is dying*. The presented covariation data tables can be seen in Figure 3. As in Experiment 1, half of the children in each group received order A (1, 5, 3, 9, 4, 8, 6, 7, 2) and the other half order B (1, 2, 7, 6, 8, 4, 9, 3, 5).

The procedure was also similar to the asymmetrical variables condition in Experiment 1; the only change was in the wording of the target question. In contrast to the asymmetrical variables condition from Experiment 1, it included an explicit comparison: "Is his fertilizer good for plants, or is it not good for plants, or is there no difference?"

Results

Neither gender (range: 0–1; male: $M = .62$, $SD = .28$; female: $M = .60$, $SD = .29$; $F(1, 48) = .03$, $p = .868$, $\eta_p^2 < .01$) nor order of items (range: 0–1; order A: $M = .62$, $SD = .27$; order B: $M = .60$, $SD = .29$; $F(1, 48) = .05$, $p = .825$, $\eta_p^2 < .01$) were associated with children's judgments. The proportion of correct judgments (range: 0–1) was $M = .50$ ($SD = .30$) in the asymmetrical variables, explicit question condition and $M = .72$ ($SD = .21$) in the symmetrical variables, explicit question condition. An ANOVA with condition as a factor indicated a significant effect, $F(1, 48) = 9.23$, $p = .004$, $\eta_p^2 = .16$. Thus, children performed better in the symmetrical variables condition, even when question format was held constant.

Discussion

Experiment 2 demonstrates that symmetry of variables affects performance, even with question format controlled. It can be inferred that the symmetry effect found in Experiment 1 was not caused by the differences in question format, but by the symmetry of variables.

The present research is the first evidence that symmetry of variables influences children's interpretations of covariation data presented in contingency tables. However, it is also not known whether adults' interpretations of covariation data presented in contingency tables are also influenced by the symmetry of variables. To date, the only existing study that has investigated symmetry of variables in adults did not use contingency tables, but instead used a sentence-choice paradigm (Beyth-Marom, 1982).

Experiment 3

Experiment 3 was designed to test the influence of symmetry of variables on adults' interpretations of covariation data presented in contingency tables. The experiment was conducted using an online questionnaire. Based on prior evidence that correlations between asymmetrical variables are interpreted more narrowly than corre-

lations between symmetrical variables in a sentence-choice task (Beyth-Marom, 1982), we hypothesized that adults would perform better with symmetrical variables than with asymmetrical variables.

Method

Participants. The original sample consisted of 122 undergraduate education students. Thirty-seven were excluded because they did not finish the questionnaire and 23 were excluded because they indicated that they used a calculator or did not answer the calculator question. The final sample consisted of 62 students (30.6% male; mean age $M = 22.39$ years, $SD = 2.35$, range: 19.00–29.00).

Design. The experiment used a one-factor design, with symmetry of variables as a between-subjects factor. Participants were randomly assigned to one of the two groups (asymmetrical: $n = 32$; symmetrical: $n = 30$).

Materials. The materials were similar to the materials from Experiment 1. Two series of 10 pictures showing covariation data in 2×2 contingency tables were embedded into a context story about the effectiveness of fertilizers. The rows and columns of the tables were labeled in the same way as in Experiment 1. However, the four cells were additionally labeled with the capital letters A, B, C, and D and the cell frequencies were depicted with numbers (see Figure 5).

The data sets with their corresponding χ^2 values are presented in Figure 6. To avoid ceiling effects, the number range was extended (1 to 990) and ratios were less salient and more difficult to calculate (e.g., Items 3 and 7) as compared to the data sets in the child experiments. Furthermore, for Items 6, 7 and 9, the ratio between the first column and the second column was below 1 (corresponding with smaller values in the first column and larger ones in the second column). Two data sets (Items 5 and 6) showed no covariation, whereas all other items showed either positive (Items 1, 7, 8, and 9) or negative (Items 2, 3, 4, and 10) relationships.

Items were presented to all participants in the same order (2, 10, 9, 6, 3, 8, 1, 4, 5, 7).

Procedure. Testing was conducted using an online questionnaire system, which assigned the participants randomly to one of the two conditions. After receiving information about the aim of the study and being asked to work without the help of a calculator, the participants were presented with demographic questions (gen-

der, age). Next, the context story was introduced and the table was explained based on an example. In the asymmetrical variables condition, the story was about 10 scientists, each of whom had invented one fertilizer and wanted to test its effectiveness. In the symmetrical variables condition, each scientist had invented two fertilizers and wanted to test which one was better. Then, the 10 items were presented in fixed order. In the asymmetrical variables condition, participants were asked to judge whether the fertilizer is good for plants (answer alternatives: yes/no). In the symmetrical variables condition, they were asked to judge whether it is better for plants to get fertilizer O or fertilizer P or if it makes no difference (answer alternatives: fertilizer O/fertilizer P/no difference). The justification question was the same in both conditions: "Where do you see this in the table?" Participants answered the justification question by typing freely into a text window. Participants were encouraged to use the alphabetical labels of the cells for their explanations. At the end of the questionnaire, participants were asked if they had used a calculator.

Coding. Justifications were categorized into the same main categories as in Experiment 1: four-cell justifications, two-cell justifications, and other (see Table 4); most adults reported four-cell justifications. Thus, only the four-cell justifications were further divided into the subcategories comparison of ratios, comparison of differences and other four-cell justifications. Simple comparisons of marginal sums did not occur.

All justifications were coded by two raters independently. Interrater agreement varied from 88.7% to 98.4% for the main categories and from 88.0% to 100% for the four-cell subcategories. Disagreements were resolved by discussion.

Results

Gender was associated with the average proportion of correct judgments (range: 0–1; male: $M = .83$, $SD = .15$; female: $M = .65$, $SD = .27$; $F(1, 60) = 7.07$, $p = .010$, $\eta_p^2 = .11$) but not with justifications (range: 0–9; four-cell justifications: male: $M = 7.74$, $SD = 2.79$; female: $M = 7.05$, $SD = 3.27$; $F(1, 60) = .64$, $p = .426$, $\eta_p^2 = .01$; two-cell justifications: male: $M = .74$, $SD = 1.82$; female: $M = 1.47$, $SD = 2.14$; $F(1, 60) = 1.66$, $p = .202$, $\eta_p^2 = .03$; other: male: $M = 1.53$, $SD = 2.44$; female: $M = 1.49$, $SD = 2.37$; $F(1, 60) < .01$, $p = .954$, $\eta_p^2 = .01$).

The average proportion of correct judgments did not differ between the asymmetrical variables condition (range: 0–1; $M = .71$, $SD = .19$) and the symmetrical variables condition ($M = .70$,







		Asymmetrical Variables Condition		Symmetrical Variables Condition	
		 Plant is growing	 Plant is dying	 Plant is growing	 Plant is dying
	A	190	130	190	130
	C	990	930	990	930

Figure 5. Sample item for the asymmetrical and symmetrical variables condition in Experiment 3. See the online article for the color version of this figure.

Item	1		2		3		4		5	
Cell frequencies	600 730	300 480	30 20	11 1	454 299	299 144	450 150	400 100	20 10	10 5
χ^2 values	8.90		4.32		6.21		3.88		0	

Item	6		7		8		9		10	
Cell frequencies	38 56	114 168	11 1	96 87	190 990	130 930	32 91	11 70	8 46	35 73
χ^2 values	0		5.06		6.72		4.54		5.71	

Figure 6. Cell frequencies and χ^2 values for the 10 covariation data sets in Experiment 3.

$SD = .31$; $F(1, 60) = .04$, $p = .846$, $\eta_p^2 < .01$). However, the probability of answering correctly by chance differed between conditions. In the asymmetrical variables condition, the participants chose between two answer alternatives (yes/no), so the probability of answering correctly by chance was .50. In the symmetrical variables condition, there were three answer alternatives (fertilizer A/fertilizer B/no difference) and therefore the probability of answering correctly by chance was .33. To correct for chance responding, the proportion of correct judgments was divided by 1/2 for participants in the asymmetrical variables condition and divided by 1/3 for participants in the symmetrical variables condition (so that chance would be equal to 1 in each condition). Then, these calculated values were z-standardized. The mean z-values were $M = -.43$, $SD = .48$ for the asymmetrical variables condition and $M = .45$, $SD = 1.20$ for the symmetrical variables condition. This difference was significant, $F(1, 60) = 14.64$, $p < .001$, $\eta_p^2 = .20$. Thus, when correcting for chance responding, the supportive effect of symmetrical variables was revealed.

Table 7 presents the mean frequencies for the main categories and the four-cell subcategories of justifications. With respect to the main categories, it can be seen that adults produced four-cell justifications far more often than two-cell justifications. Although there were more two-cell justifications in the asymmetrical variables condition, $F(1, 60) = 6.11$, $p = .016$, $\eta_p^2 = .09$, there was no significant difference for the four-cell justifications, $F(1, 60) = .45$, $p = .506$, $\eta_p^2 < .01$, nor for the residual (other) category, $F(1, 60) = 1.40$, $p = .242$, $\eta_p^2 < .02$.

Table 7
Mean Values (Range: 0–10; SD in Parentheses) for Each Justification Category for Both Conditions

Category	Condition	
	Asymmetrical	Symmetrical
Four-cell justifications	7.00 (3.16)	7.53 (3.10)
Comparison of ratios	5.25 (3.65)	6.63 (3.66)
Comparison of differences	1.56 (2.44)	.80 (1.79)
Other 4 cells	.19 (.54)	.10 (.31)
Two-cell justifications	1.84 (2.54)	.60 (1.10)
Other	1.16 (1.87)	1.87 (2.80)

It should be noted that the most sophisticated strategy - comparison of ratios - was the most common justification type in both conditions. We classified participants as ratio-users if they used this strategy for more than half of the 10 items (i.e., at least six times). In the asymmetrical variables condition, 46.9% of participants were ratio-users, and in the symmetrical variables condition, 73.3% of participants were ratio-users. A one-tailed, exact χ^2 test indicated a significant difference between conditions, $\chi^2(1, N = 62) = 4.50$, $p = .031$, $w = .27$. Thus, a higher proportion of participants reported ratio-based justifications on a majority of items in the symmetrical variables condition.

Discussion

Symmetry of variables affected adults' interpretations of covariation data presented in summary form. Adults produced more correct judgments for symmetrical variables than for asymmetrical variables, although this difference was only revealed after correcting for chance responding in each condition. In the symmetrical variables condition, adults produced fewer two-cell justifications, and a greater proportion of the adults reported ratio-based justifications on a majority of problems. These results are in line with our findings with children and with related findings about adults' concepts of correlation (Beyth-Marom, 1982).

General Discussion

Empirical Summary

Given diverging results on young children's sensitivity to covariation on the one hand (e.g., Kushnir & Gopnik, 2007; Sobel & Kirkham, 2006) and adults' difficulties in reasoning about covariation on the other hand (e.g., Shaklee, 1983), we investigated the influence of task characteristics on children's and adults' interpretation of covariation data presented in contingency tables. We examined the effects of symmetry of variables in three experiments. Taken together, these experiments demonstrate that symmetry of variables influences contingency table interpretation, both in children and in adults. Experiment 1 showed that symmetrical variables support elementary school children's interpretations of data patterns, and that the symmetry effect is larger in Grade 2 than

in Grade 4. For tables with symmetrical variables, second graders judged on average about 70% of our covariation problems correctly. Thus, elementary school children display a basic competency in interpreting covariation data presented in contingency tables. The discrepancy between implicit judgments of experienced covariation and explicit judgments of descriptive data may be less pronounced than has generally been assumed.

The difference between asymmetrical and symmetrical variables was also reflected in children's justifications. When interpreting tables with symmetrical variables, children reported more four-cell justifications and also more sophisticated four-cell justifications, such as the comparison of ratios. In contrast, they offered more inadequate four-cell justifications (e.g., comparison of marginal sums) with asymmetrical variables. With respect to two-cell justifications, there was more column-wise reasoning and less focus on the first row in the symmetrical variables condition.

The evidence presented in Experiment 2 indicates that the difference between conditions in Experiment 1 was caused by the symmetry of variables and not only by the differing question formats. Experiment 3 showed that symmetry of variables also affects adults' interpretations of covariation data presented in contingency tables. Furthermore, our adult sample showed good performance in dealing with highly complex data patterns (i.e., they displayed a high proportion of correct judgments and comparisons of ratios).

Our Findings in Light of Prior Research

Overall, the children performed better than would be expected in light of prior research on the interpretation of covariation data. Prior studies that presented children with each data point serially showed that 4- and 5-year-olds as well as 8- and 12-year-olds can interpret perfect covariation, but have substantial difficulties with noncovariation (Amsel & Brock, 1996; Koerber et al., 2005; Piekny et al., 2014; Ruffman et al., 1993). In our sample, children solved two out of three noncovariation items with high success rates in the symmetrical variables condition. Shaklee and colleagues presented the data in summary tables and classified most elementary school children as A versus B strategy users (Shaklee & Mims, 1981; Shaklee & Paszek, 1985). In our child samples, most justifications referred to all four cells of the table.

Performance in our adult sample was also surprisingly good, relative to prior studies. Whereas past studies reported quite low judgment accuracy and use of inadequate strategies (e.g., Batanero et al., 1996; Smedslund, 1963), our results indicate a high level of proficiency in interpreting data. Participants interpreted about 70% of the complex data patterns correctly, and they reported four-cell justifications in most of the cases. Moreover, most of the adults' justifications included the comparison of ratios.

The reasons for participants' better-than-expected performance in our studies cannot be determined conclusively because prior studies differed from the present studies in multiple ways (e.g., data patterns, task format). However, our findings suggest that the use of asymmetrical variables contexts in most studies (e.g., Amsel & Brock, 1996; Arkes & Harkness, 1983; Kao & Wasserman, 1993; Shaklee & Mims, 1981; Shaklee & Wasserman, 1986;

Smedslund, 1963; Ward & Jenkins, 1965) likely contributed to participants' difficulties.

What Drives the Symmetry Effect?

The present study is the first to investigate the effects of symmetry of variables on the interpretation of covariation data. We showed that the symmetry effect is not only relevant for adults' data interpretation, but it is also relevant for elementary school children's. The effect of symmetry of variables was stronger in younger children than in older ones. Our results also suggest that the effect is smaller in adults than in elementary school children. However, methodological differences (e.g., different items, data collection with one-on-one interviews vs. online questionnaire) might have contributed to this pattern of findings, so that a decrease in the effect in adulthood cannot be inferred conclusively.

One possible explanation for participants' difficulties with asymmetrical variables is that they may have held the implicit hypothesis that the present value has a positive effect (e.g., that the fertilizer is effective). Previous studies have shown that a prior belief about the relationship between variables impedes interpretation if the data are not in line with this belief (Amsel & Brock, 1996; Kuhn et al., 1988). The implicit hypothesis that fertilizer is effective might affect performance in the same way as a prior belief. However, the item-based analysis in Experiment 1 does not provide strong support for this explanation. For example, children's performance on Item 5 (8/8/16/16) did not differ between conditions, although this item shows no effect and thus is not in line with the hypothesis that the fertilizer is effective. It seems that the interaction between the condition-specific strategies (asymmetrical: row-wise; symmetrical: column-wise, four-cell) and the numerical structure of the items influences performance more than the data's relation to the hypothesis that the fertilizer is effective.

A second possible explanation is that the higher "status" of the present value in the asymmetrical case leads to neglect of the data concerning the absent value. In contrast, the similar "status" of both values in the symmetrical case provides a context for a meaningful comparison of both rows. This idea is in line with the observed differences in justifications: Symmetrical variables were not only associated with more four-cell reasoning, but also with more sophisticated comparisons. Thus, with symmetrical variables, children did not focus only on the most salient cells of the contingency table, but tended to include all available information in their reasoning process. The higher frequency of more adequate four-cell strategies suggests that in the symmetrical variables context, meaningful comparisons between relevant cells are more salient for children. Moreover, the idea that symmetrical variables provide a context for meaningful comparisons is supported by the higher prevalence of column-wise (i.e., across row) justifications and the lower prevalence of first-row justifications in this condition. Thus, our data provide evidence that the symmetry of variables effect is driven by the differing salience of relevant comparisons.

The interpretation of the symmetry effect in terms of salience can be extended to experimentation tasks. It may be that difficulties in interpreting data in asymmetrical contexts are associated with more general problems understanding contrastive tests. For a

contrastive test, one needs to compare a case in which the variable of interest is present with a case in which it is absent (i.e., one needs to vary the target variable). Studies of young children's experimentation have revealed that they have difficulty making contrastive tests. For example, in a study in which children were asked to choose which of three experiments was best, Koerber, Sodian, Kropf, Mayer, and Schwippert (2011) reported that most second graders did not choose a contrastive test, but instead chose an alternative in which no variables were varied. Most fourth graders did choose the contrastive test in this context, but other evidence suggests that many elementary school children continue to have difficulties understanding the importance of varying only one thing at a time (e.g., Croker & Buchanan, 2011; Piekny & Maehler, 2013). We argue that once children fully grasp the idea of contrastive testing, even asymmetrical variables should provide a context for meaningful comparisons. The second graders' problems with contrastive tests and the developmental trend found by Koerber and colleagues (2011) align with our finding that second graders' data interpretation is more strongly influenced by asymmetrical variables than fourth graders'.

Limitations and Future Directions

Our work demonstrates that children's interpretations of contingency tables are influenced by the symmetry of variables, even when the target question is identical across conditions. Our findings suggest that adults also display the symmetry effect; however, the target question was not identical across conditions in our adult experiment. Future work is needed to ascertain that adults do indeed display the symmetry effect, even when the target question is held constant. Preliminary work (Magee, 2015; Osterhaus, Magee, Saffran, & Alibali, 2016) suggests that this is indeed the case; however, more data are needed to establish this point more definitively.

Some other limitations of the adult experiment should also be noted. The online questionnaire method allows less experimental control than individual interviews, like those we conducted with the children. Indeed, we had to exclude many participants from the online experiment because they acknowledged using a calculator to obtain their answers. Future work with adults in an individual interview setting would avoid this issue.

This body of work could also be criticized on the grounds that we inferred people's strategies from their verbal justifications (in the case of the children) or their written justifications (in the case of the adults). It is possible that the participants—especially children—were not able to accurately report how they reached their judgments, and that the reported justifications do not reflect their actual strategies. However, the differences in judgments between conditions corresponded with the justifications, supporting the validity of this approach. Furthermore, the analysis of children's justifications led to the identification of strategies such as the comparison of marginal sums, which have not been described in the literature on the interpretation of covariation data so far.

An additional limitation of the present study is its failure to definitively identify the critical conceptual differences that underpin the performance differences between the symmetrical and asymmetrical conditions. Why does the salience of the comparison

differ in the symmetrical and asymmetrical conditions? One possibility has to do with the salience of action. In our research, the symmetrical condition contrasts two different actions (applying fertilizer A and applying fertilizer B), whereas the asymmetrical condition contrasts an action (applying fertilizer) with no action. It could be argued that a different action is generally more salient than no action and that the observed effect may be due to this asymmetry. However, a "no action" condition could easily be framed as the more salient one, for instance if the research question addressed the effects of not doing something that one normally does (e.g., not eating breakfast) on some outcome (e.g., performance on a math test). Future research should investigate this issue more thoroughly.

The present studies offer several other starting points for future research. First, the interaction between the symmetry of variables effect and age needs further investigation. Based on the second graders' good performance in the symmetrical variables condition, it seems reasonable to assume that even preschoolers might be able to make explicit judgments based on covariation data embedded in symmetrical contexts. The interaction with age in elementary school children suggests that the symmetry effect might be even larger in younger children. Second, the hypothesis that difficulties with data interpretation in asymmetrical contexts are related to experimentation skills should be examined. This would help to determine if the differing salience of the comparison really drives the effect of variable symmetry. Third, it would be interesting to test whether the supportive effect of symmetrical reasoning would carry over to asymmetrical contexts, if symmetrical contexts were encountered first. It is possible that interpreting covariation data in the supportive symmetrical context would facilitate subsequent interpretations of covariation with asymmetrical variables.

Applications

The present study has implications, both for the practice of science and for education. Researchers who are interested in data interpretation or in the statistical concepts of covariation or correlation should be aware that the symmetry of variables is an influential factor. Whether tasks are embedded in an asymmetrical or symmetrical context may affect how children and adults understand statistical concepts and where their attention is drawn. Evaluations of children's reasoning that rely solely on asymmetrical variables items may underestimate children's abilities, and this may hold true for adults, as well. Thus, symmetry of variables is a factor that should be considered or controlled for in research on data interpretation and covariation understanding.

These findings also have implications for educational practices. Curriculum developers and textbook authors should attend to the symmetry of variables in examples and in lessons about data interpretation. Our data suggest that symmetry of variables influences the ways in which children think about data and about the concept of covariation. Symmetrical contexts support more sophisticated reasoning, so it might be useful for instruction to begin with symmetrical contexts and then move on to asymmetrical contexts. Understanding covariation in a symmetrical context could serve as a scaffold for understanding covariation more generally, and also

for helping students to build a richer understanding of the value of contrastive tests.

Conclusion

Symmetry of variables is an influential factor in both children's and adults' interpretations of covariation data presented in contingency tables. With the support of symmetrical variables, children as young as second grade are able to interpret data presented in contingency tables, indicating that children's ability to interpret covariation data patterns is not limited to implicit, experiential tasks. Symmetry of variables continues to play a role in data interpretation, even into adulthood, affecting both the strategies that people use and the conclusions they draw.

References

- Amsel, E., & Brock, S. (1996). The development of evidence evaluation skills. *Cognitive Development, 11*, 523–550. [http://dx.doi.org/10.1016/S0885-2014\(96\)90016-7](http://dx.doi.org/10.1016/S0885-2014(96)90016-7)
- Arkes, H. R., & Harkness, A. R. (1983). Estimates of contingency between two dichotomous variables. *Journal of Experimental Psychology: General, 112*, 117–135. <http://dx.doi.org/10.1037/0096-3445.112.1.117>
- Arkes, H. R., & Rothbart, M. (1985). Memory, retrieval, and contingency judgments. *Journal of Personality and Social Psychology, 49*, 598–606. <http://dx.doi.org/10.1037/0022-3514.49.3.598>
- Batanero, C., Estepa, A., Godino, J. D., & Green, D. R. (1996). Intuitive strategies and preconceptions about association in contingency tables. *Journal for Research in Mathematics Education, 27*, 151–169. <http://dx.doi.org/10.2307/749598>
- Beyth-Marom, R. (1982). Perception of correlation reexamined. *Memory & Cognition, 10*, 511–519. <http://dx.doi.org/10.3758/BF03202433>
- Bühner, M., & Ziegler, M. (2009). *Statistik für Psychologen und Sozialwissenschaftler* [Statistics for Psychologists and Social Scientists]. München, Germany: Pearson Deutschland GmbH.
- Collins, D. J., & Shanks, D. R. (2002). Momentary and integrative response strategies in causal judgment. *Memory & Cognition, 30*, 1138–1147. <http://dx.doi.org/10.3758/BF03194331>
- Croker, S., & Buchanan, H. (2011). Scientific reasoning in a real-world context: The effect of prior belief and outcome on children's hypothesis-testing strategies. *British Journal of Developmental Psychology, 29*, 409–424. <http://dx.doi.org/10.1348/026151010X496906>
- Gopnik, A., Sobel, D. M., Schulz, L. E., & Glymour, C. (2001). Causal learning mechanisms in very young children: Two-, three-, and four-year-olds infer causal relations from patterns of variation and covariation. *Developmental Psychology, 37*, 620–629. <http://dx.doi.org/10.1037/0012-1649.37.5.620>
- Inhelder, B., & Piaget, J. (1958). *The growth of logical thinking from childhood to adolescence*. New York, NY: Basic Books.
- Jenkins, H. M., & Ward, W. C. (1965). Judgment of contingency between responses and outcomes. *Psychological Monographs: General and Applied, 79*(1), 1–17.
- Kao, S.-F., & Wasserman, E. A. (1993). Assessment of an information integration account of contingency judgment with examination of subjective cell importance and method of information presentation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19*, 1363–1386. <http://dx.doi.org/10.1037/0278-7393.19.6.1363>
- Klayman, J., & Ha, Y.-W. (1987). Confirmation, disconfirmation, and information in hypothesis testing. *Psychological Review, 94*, 211–228. <http://dx.doi.org/10.1037/0033-295X.94.2.211>
- Koerber, S., Sodian, B., Kropf, N., Mayer, D., & Schwippert, K. (2011). Die Entwicklung des wissenschaftlichen Denkens im Grundschulalter [The Development of Scientific Thinking in Elementary]. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie, 43*, 16–21. <http://dx.doi.org/10.1026/0049-8637/a000027>
- Koerber, S., Sodian, B., Thoermer, C., & Nett, U. (2005). Scientific reasoning in young children: Preschoolers' ability to evaluate covariation evidence. *Swiss Journal of Psychology/Schweizerische Zeitschrift für Psychologie/Revue Suisse de Psychologie, 64*, 141–152. <http://dx.doi.org/10.1024/1421-0185.64.3.141>
- Kuhn, D., Amsel, E., O'Loughlin, M., Schauble, L., Leadbeater, B., & Yotiv, W. (1988). *The development of scientific thinking skills*. San Diego, CA: Academic Press.
- Kushnir, T., & Gopnik, A. (2007). Conditional probability versus spatial contiguity in causal learning: Preschoolers use new contingency evidence to overcome prior spatial assumptions. *Developmental Psychology, 43*, 186–196. <http://dx.doi.org/10.1037/0012-1649.43.1.186>
- Levin, I. P., Wasserman, E. A., & Kao, S.-F. (1993). Multiple methods for examining biased information use in contingency judgments. *Organizational Behavior and Human Decision Processes, 55*, 228–250. <http://dx.doi.org/10.1006/obhd.1993.1032>
- Magee, J. (2015). *Effects of symmetry and context on covariation assessment* (Unpublished Bachelor's thesis). University of Wisconsin-Madison, Madison, WI.
- Mata, A., Ferreira, M. B., & Sherman, S. J. (2013). Flexibility in motivated reasoning: Strategic shifts of reasoning modes in covariation judgment. *Social Cognition, 31*, 465–481. http://dx.doi.org/10.1521/soco_2012_1004
- Mata, A., Garcia-Marques, L., Ferreira, M. B., & Mendonça, C. (2015). Goal-driven reasoning overcomes cell D neglect in contingency judgments. *Journal of Cognitive Psychology, 27*, 238–249. <http://dx.doi.org/10.1080/20445911.2014.982129>
- Osterhaus, C., Magee, J., Saffran, A., & Alibali, M. W. (2016, July). *Judging the contingency between variables: Where scientific thinking and mathematics come together*. Paper presented at the 4th Annual Midwest Meeting on Mathematical Thinking, Madison, WI.
- Piekny, J., Grube, D., & Maehler, C. (2014). The development of experimentation and evidence evaluation skills at preschool age. *International Journal of Science Education, 36*, 334–354. <http://dx.doi.org/10.1080/09500693.2013.776192>
- Piekny, J., & Maehler, C. (2013). Scientific reasoning in early and middle childhood: The development of domain-general evidence evaluation, experimentation, and hypothesis generation skills. *British Journal of Developmental Psychology, 31*, 153–179. <http://dx.doi.org/10.1111/j.2044-835X.2012.02082.x>
- Reed, H. C., Stevenson, C., Broens-Paffen, M., Kirschner, P. A., & Jolles, J. (2015). Third graders' verbal reports of multiplication strategy use: How valid are they? *Learning and Individual Differences, 37*, 107–117. <http://dx.doi.org/10.1016/j.lindif.2014.11.010>
- Robinson, K. (2001). The validity of verbal reports in children's subtraction. *Journal of Educational Psychology, 93*, 211–222. <http://dx.doi.org/10.1037/0022-0663.93.1.211>
- Ruffman, T., Perner, J., Olson, D. R., & Doherty, M. (1993). Reflecting on scientific thinking: Children's understanding of the hypothesis-evidence relation. *Child Development, 64*, 1617–1636. <http://dx.doi.org/10.2307/1131459>
- Schulze, C., & Hertwig, R. (2016). *Statistical intuitions: Smart babies, stupid adults?* [Manuscript submitted for publication].
- Shaklee, H. (1983). Human covariation judgment: Accuracy and strategy. *Learning and Motivation, 14*, 433–448. [http://dx.doi.org/10.1016/0023-9690\(83\)90026-7](http://dx.doi.org/10.1016/0023-9690(83)90026-7)
- Shaklee, H., & Elek, S. (1988). Cause and covariate: Development of two related concepts. *Cognitive Development, 3*, 1–13. [http://dx.doi.org/10.1016/0885-2014\(88\)90027-5](http://dx.doi.org/10.1016/0885-2014(88)90027-5)
- Shaklee, H., & Mims, M. (1981). Development of rule use in judgments of covariation between events. *Child Development, 52*, 317–325. <http://dx.doi.org/10.2307/1129245>

- Shaklee, H., & Mims, M. (1982). Sources of error in judging event covariations: Effects of memory demands. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 8, 208–224. <http://dx.doi.org/10.1037/0278-7393.8.3.208>
- Shaklee, H., & Paszek, D. (1985). Covariation judgment: Systematic rule use in middle childhood. *Child Development*, 56, 1229–1240. <http://dx.doi.org/10.2307/1130238>
- Shaklee, H., & Tucker, D. (1980). A rule analysis of judgments of covariation between events. *Memory & Cognition*, 8, 459–467. <http://dx.doi.org/10.3758/BF03211142>
- Shaklee, H., & Wasserman, E. A. (1986). Judging interevent contingencies: Being right for the wrong reasons. *Bulletin of the Psychonomic Society*, 24, 91–94. <http://dx.doi.org/10.3758/BF03330513>
- Smedslund, J. (1963). The concept of correlation in adults. *Scandinavian Journal of Psychology*, 4, 165–173. <http://dx.doi.org/10.1111/j.1467-9450.1963.tb01324.x>
- Sobel, D. M., & Kirkham, N. Z. (2006). Blickets and babies: The development of causal reasoning in toddlers and infants. *Developmental Psychology*, 42, 1103–1115. <http://dx.doi.org/10.1037/0012-1649.42.6.1103>
- Sobel, D. M., Tenenbaum, J. B., & Gopnik, A. (2004). Children's causal inferences from indirect evidence: Backwards blocking and Bayesian reasoning in preschoolers. *Cognitive Science*, 28, 303–333.
- Ward, W. C., & Jenkins, H. M. (1965). The display of information and the judgment of contingency. *Canadian Journal of Psychology/Revue canadienne de psychologie*, 19, 231–241. <http://dx.doi.org/10.1037/h0082908>
- Wasserman, E. A., Dorner, W. W., & Kao, S. F. (1990). Contributions of specific cell information to judgments of interevent contingency. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 16, 509–521. <http://dx.doi.org/10.1037/0278-7393.16.3.509>
- Wasserman, E. A., & Shaklee, H. (1984). Judging response-outcome relations: The role of response-outcome contingency, outcome probability, and method of information presentation. *Memory & Cognition*, 12, 270–286. <http://dx.doi.org/10.3758/BF03197676>

Received November 11, 2015

Revision received June 24, 2016

Accepted July 12, 2016 ■

Correction to Cowan et al. (2006)

In the article “Life-Span Development of Visual Working Memory: When Is Feature Binding Difficult?” (*Developmental Psychology*, 2006, Vol. 42, No. 6, pp. 1089–1102. <http://dx.doi.org/10.1037/0012-1649.42.6.1089>) by Nelson Cowan, Moshe Naveh-Benjamin, Angela Kilb, and J. Scott Saults, there were two errors in experiment 1a. The mean for color item information in older adults was incorrectly calculated. As a result, Figure 3 shows a mean of over .70. The true mean was .63 ($SEM = .04$). This change diminishes the magnitude of the aging deficit for associative information, although this deficit still appears to remain, to a smaller extent. (For a conceptual replication see Peterson & Naveh-Benjamin, 2016). There also was an error in the experimental procedure of Experiment 1a. The older adults in that experiment received only half the number of trials specified in the methods section, and half as much as the other groups. For all groups, when there were 4 or 6 items and the probe was a binding change, the probed location was matched by the same color at 1 other location but, when there were 8 or 10 squares, the probed location was matched by the same color at 1, 2, or 3 other locations. For 8 squares the number of trials was identical for these three trial subtypes whereas, for 10 squares, most of the trials had the same color at just 1 other location. These errors suggest that the experiment should be taken as only preliminary evidence that there is an aging deficit in color-location binding in visual working memory when color and binding trials are mixed in the same trial blocks.

Reference

- Peterson, D. J., & Naveh-Benjamin, M. (2016). The role of aging in intra-item and item-context binding processes in visual working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. Advance online publication. <http://dx.doi.org/10.1037/xlm0000275>.

<http://dx.doi.org/10.1037/dev0000222>