
Volume 116, Issue 1, July 2010
ISSN 0010-0277
pp. 1–154
ELSEVIER
COGNITION

Cognition 116 (2010) 1–14

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/COGNIT

How children use examples to make conditional predictions

Charles W. Kalish *

University of Wisconsin-Madison, Department of Educational Psychology, 1025 West Johnson St., Madison, WI 53705, United States

ARTICLE INFO

Article history:
Received 15 July 2009
Revised 2 March 2010
Accepted 3 March 2010

Keywords:
Probability learning
Inductive inference
Cognitive development

ABSTRACT

Two experiments explored children's and adults' use of examples to make conditional predictions. In Experiment 1 adults ($N = 20$) but not 4-year-olds ($N = 21$) or 8-year-olds ($N = 18$) distinguished predictable from un-predictable features when features were partially correlated (e.g., necessary but not sufficient). Children did make reliable predictions given perfect correlation. In the context of categorization and property projection in Experiment 2, children of both ages (both $N = 31$) and adults ($N = 30$) did use partial correlation in examples to make conditional predictions. However, predictions of category membership given property possession were more reliable than were predictions of property possession given category membership. Children generally showed good memory for frequency information, but did not always use this information as the basis of predictions. Results suggest that young children may have difficulty selectively using the relations they observe in experience.

1. Introduction

Infants and young children are able to learn patterns of contingency and covariation in events they experience (Aslin, Saffran, & Newport, 1998; Kirkham, Slemmer, Richardson, & Johnson, 2007; Xu & Denison, 2009). The current study explores the features of experience that lead to learning a correlation or contingency. What makes a correlation easy or difficult to detect? What sorts of contingencies do children learn? The current study focuses on conditional predictions. Conditional prediction is the expectation of one feature given, or conditional upon, some other(s). Categorization (e.g., if it barks is it a dog?) and property projection (e.g., if it is a dog, will it bark?) are kinds of conditional predictions. Within the developmental literature there has been substantial debate about the role of theoretical knowledge vs. statistical information in children's categorization and inference (Gelman & Kalish, 2006; Rogers & McClelland, 2004; Sloutsky, 2003). However, this debate is best understood as questioning whether children's judgments involve theoretical knowledge in addition to statistical, not instead (Gelman & Medin, 1993). Most researchers agree that expectations about barking given dog, and dog given barking, are somehow reflective of encounters with barking and non-barking dogs and non-dogs. But how do children use experience of past examples to make predictions about novel cases?

One of the simplest, and most classic, specifications of a prediction problem involves relations between two binary variables. Such a problem can be represented in a 2×2 contingency table (see Table 1). The row and column headings specify values of two feature dimensions (e.g., features of a set of birds: Diet, either berries or bugs; and egg-type, either brown or white). The cells of the table indicate frequencies with which examples possessing combinations of these feature-values are encountered (The cells are conventionally labeled left-to-right, top-to-bottom as A, B, C, D). We may ask what people learn from various frequency distributions of examples. This approach is familiar from the literature on association (Kao & Wasserman, 1993), causal learning (Cheng & Novick, 1992), and probability judgment (Schlottmann, 2001). Frequency distributions also provide bases for conditional prediction: Will a bird who eats berries lay brown eggs? Will a bird with white eggs eat bugs?

* Tel.: +1 608 262 9920; fax: +1 608 262 0843.
E-mail address: cwkalish@wisc.edu

0010-0277/$ - see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.cognition.2010.03.008

Table 1
Example of 2×2 contingency table illustrating frequencies of different examples (combinations of features).

	Berries	Bugs
Brown	8 A	2 B
White	2 C	4 D

In a recent paper, Girotto and Gonzalez (2008) explored children's use frequency information to make conditional predictions. Girotto and Gonzalez presented children with examples composed of two binary features (black and white, squares and circles, akin to the structure in Table 1). They asked two kinds of questions. Prior probabilities involved judgments of marginal frequencies, predictions of individual features. For example, from Table 1, is one more likely to find a bird that eats berries or one that eats bugs? Girotto and Gonzalez were particularly interested in judgments of "posterior probabilities", or conditional probabilities. What if one knows the value of one feature and is asked to guess the value of the other? For example, if we know that a bird lays white eggs, is it more likely to eat bugs or berries? In answering this question one has to select a sub-set of the exemplars to compare. Indeed with the cell frequencies in Table 1 (like those used by Girotto and Gonzalez) the answers to the prior and the posterior questions differ. Overall more birds eat berries, but among the white-egged birds, more eat bugs. Girotto and Gonzalez found that children as young as age four could correctly answer the posterior question suggesting they were appropriately conditionalizing their comparisons.

Although Girotto and Gonzalez (2008) found evidence that preschool-aged children did make the conditional comparisons, the task seemed difficult. Despite working with a small number of items (eight total) and having representations of the cell frequencies visible throughout the task, young children often responded at chance levels. Even into second grade, children were answering incorrectly about 40% of the time in some conditions. In some ways the difficulty of the conditional prediction task is incompatible with other work showing that young children are adept at making categorization judgments, and are quick to learn to predict properties from category membership (Gelman, 2003; Gelman & Kalish, 2006; Markman, 1989). The way to address this apparent inconsistency is to explore what makes tasks easier or more difficult. How do young children use frequency distributions to make conditional predictions? The literature suggests four strategies, four ways people may go about answering a conditional prediction question given frequency information like that shown in Table 1.

2. Strategies for conditional prediction

An early hypothesis, developed from Piaget's theory, held that young children focus on a single kind of exemplar, one cell in the table, because they are unable to coordinate multiple representations (see Reyna & Brainerd, 1994 for review). In answering questions about the exemplars represented in Table 1, preschool-aged children would focus on the berry-eating brown-egged birds because they are the most frequent, and essentially ignore the frequencies in the other cells. This focus might lead them to always expect berry-eating, a prediction in contrast to Girotto and Gonzalez's conditionalizing results. Indeed, the general consensus in the field is that young children are not limited to attending to one cell at a time; they do compare multiple cell frequencies when making predictions (see Reyna & Brainerd, 1994; Schlottmann, 2001). However, it remains possible that modal exemplars play an important role in children's predictions. Perhaps children will tend to make predictions consistent with the features of the modal exemplar. Tasks are easier when correct performance involves predicting the features shown by the modal exemplar.

A second strategy is to focus on marginal frequencies, or base-rates. From Table 1, most of the birds lay brown eggs, and most eat berries. Following a base-rate strategy, a child would tend to predict one of these two features. Girotto and Gonzalez (2008) demonstrated that even preschool-aged children are not limited to attending to marginal frequencies; under some conditions they will consistently predict the property with the lower base-rate. Girotto and Gonzalez (2008) did not directly compare conditional predictions involving high and low base-rate properties. Thus it remains an open question whether children are more likely to predict high base-rate features in conditional prediction tasks. Perhaps such tasks are easier when correct performance involves predicting the features with the highest base-rates.

The third strategy is the "correct" one of calculating the probability of one feature conditional on the other. This strategy is based on the definition of a conditional probability: $p(X|Y) = p(X \text{ and } Y)/P(Y)$. Assuming that the frequencies of observed exemplars and properties are the bases for assignments of probabilities, then making a conditional prediction involves computing the proportion of all exemplars having the given property who also had the to be predicted property. From Table 1, the probability that a bird with white eggs eats berries is: C/(C + D) or 2/(2 + 4). Since the current study concerns only binary properties, the probability of the alternative outcome is always the complement (e.g., $p(\text{bug}|\text{white}) = 1 - p(\text{berry}|\text{white})$). Exemplar models of adult cognition suggest that people make just this sort of comparison when asked for a conditional prediction, albeit not always perfectly (Dougherty, Gettys, & Ogden, 1999). As this strategy is the normative one used to define correct responding, participants in a prediction task will be successful to the extent they consistently select the property with the larger conditional probability. Classically, the tendency to notice and respond to a difference depends on the magnitude of that difference (Luce, 1963). People will be more likely to select the property with the greater conditional probability if that probability is much larger than the alternative. For example, if $p(\text{berry}|\text{white}) = .9$ (vs. $p(\text{bug}|\text{berry}) = .1$) people will be more consistent in predicting berry given white than they would be if $p(\text{berry}|\text{white}) = .55$ (vs. $p(\text{bug}|\text{berry}) = .45$). Thus, the prediction is that children will find tasks easier when they involve large (rather than small) differences in conditional probabilities.

Finally, a fourth way of generating a conditional prediction is to focus on the general association between two variables. An association is a bi-conditional relation. In a 2×2 table, association is a function of the difference between the two diagonals: The more exemplars fall on only one diagonal, the greater the association between the variables. In Table 1 there is an association between the variables of diet and egg color; brown and berries go together, and white and bugs go together. Associations, like modal exemplars, base-rates, and conditional probabilities, could be used to make conditional predictions. For example, participants in Girotto and Gonzalez's study could have made the correct posterior probability prediction by attending to the association between features. The association in Table 1 supports the prediction that a bird with white eggs will eat bugs. Note, that even with a strong association, the conditional probability of bugs given white eggs need not be high, and could be less than p(berry|white). The converse holds as well: A large difference in conditional probability does not mean there has to be a strong association. This is because only two cells contribute to a particular conditional probability, while all four cells contribute to the association. Thus the fourth hypothesis, that children use association to make conditional predictions, is distinct from the previous three. This hypothesis states that children will find tasks easier when correct performance involves predicting features consistent with the overall association between feature dimensions.

Of course the four strategies are not mutually exclusive. Children may attend to modal exemplars, base-rates, conditional probabilities, and associations when making predictions. Moreover, there are clearly "non-statistical" features that affect task performance, such as memory demands and task content (e.g., intuitions about causal relations between features). Nonetheless, by presenting different frequency distributions, it is possible to distinguish the different strategies or the degree to which the different strategies drive children's predictions. These strategies are also of interest because of competing developmental hypotheses.

3. Developmental hypotheses

One developmental perspective, already mentioned, is that children's cognition moves from simple strategies to more complex ones. In the context of the four prediction strategies described above, the modal exemplar strategy is often considered the simplest. Only one piece of information is attended to; only one cell matters. The base-rate and conditional probability strategies involve comparison of two quantities (two marginal or cell frequencies). An association involves coordinating all four cell frequencies. On this view, association is the most complex strategy.

Evidence for the simple-to-complex developmental hypothesis comes from Shaklee's work on contingency or causal learning (Shaklee & Goldston, 1989; Shaklee & Mims, 1981; Shaklee & Paszek, 1985). Shaklee presented children with stimuli like that described in Table 1. She asked them for judgments of association (e.g., are white-egg birds more/less/equally likely to eat bugs as brown-egg birds?) or of causal relation (e.g., does eating bugs cause brown eggs?). Both of these judgments are understood to involve something like a comparison of on- and off-diagonal frequencies (e.g., a phi-coefficient, or Δp, see Cheng & Novick, 1990). Shaklee found that school-aged children tended to use simple conditional probability to make these judgments. For example, if p(brown|bug) was high, children would assert there was a causal relation. Shaklee's data show that when asked a question that requires comparison of four cells, children answer by comparing two. Children are more likely to attend to conditional probabilities than to associations.

A different perspective on learning relations from examples comes from work on conditioning, especially, the animal learning literature (De Houwer, Vandorpe, & Beckers, 2005; Vadillo & Matute, 2007; Vadillo, Miller, & Matute, 2005). Here the hypothesis is that association learning is basic; organisms are sensitive to patterns of covariation. Yet it is also clear that humans, at least adult humans, are able to learn and use other relations, such as conditional probability. Vadillo and Matute (2007) suggest that the conditional probability judgment requires a controlled, higher-order, process of selective combination of associative information. As the conditional prediction requires controlled processing, the expectation is that it would be more difficult or later emerging developmentally.

Vadillo and Matute (2007) suggest that conditional probability judgments are more complex because they involve combining different associations. However, associative responding may be easier or apparent earlier than conditional prediction because the latter requires more selective processing. This perspective would be consistent with fuzzy-trace theory which suggests that people form a general, gist, impression of the relation between the features, such as an association (Reyna, 2005). Children especially might have difficulty selectively encoding or using specific representations to make different judgments. Note that in terms of an association there are really only two kinds of examples in the 2×2 table (see Table 1); examples in cells A and D support the (positive) association, those in cells B and C weaken the association. The cells/examples have the same implications for all judgments based on the association. What is associated with brown eggs? What is associated with eating bugs? A and D cells support one answer; B and C cells the other. In contrast, the cells have different significance for different conditional predictions. If a bird lays brown eggs, what does it eat? A-cell examples support berries, B-cell examples bugs, with C-cell and D-cell examples either irrelevant, or somewhat supportive of berries and bugs (respectively) if base-rates are considered. If a bird eats berries, what color are its eggs? This question involves A compared to C. A plausible "association-basic" hypothesis is that the selective and differential usage of examples is difficult. Young children would be expected to use association (A and D vs. C and B) to make all conditional predictions.

There is both theoretical and empirical support for the association-basic hypothesis. Models of probability judgment include an expectation that people will base predictive judgments on the wrong exemplars, consistent with an association (Dougherty et al., 1999). A classic result in the adult literature is the "inverse fallacy" in which people confuse evidence for $p(x|y)$ with evidence for $p(y|x)$ (Dawes, Mirels, Gold, & Donahue, 1993; Villejoubert & Mandel, 2002). The literature on conditional inference ("if ... then" reasoning) often suggests a bias toward bi-conditional interpretations, especially in children (Barrouillet & Lecas, 2002). Finally, young children are reasonably good at comparing gambles (Acredolo, O'Connor, Banks, & Horobin, 1989; Spinillo, 2002; see Reyna & Brainerd, 1994 for review). At least by age seven, children produce ratings or rankings of gambles, such as containers with different numbers of red and blue marbles, which reflect the relative probabilities (e.g., of drawing a red marble). This kind of comparison involves a judgment of difference in conditional probabilities, roughly an association or correlation.[1] Interestingly when children are asked to produce equivalent gambles (by adding marbles of one color to an urn to match the proportion in another urn) they do not reliably succeed until about age 13 (Falk & Wilkening, 1998). One interpretation is that the adjustment task requires a precise (e.g., numerical) representation of probability while the rating tasks call on more intuitive or qualitative representations. However, it is also suggestive that the adjustment task requires attending to a single "cell" of the 2×2 problem (i.e., "How does one adjust A to make A/B = C/D?").

The literature presents two broad views of the development of conditional prediction. Children may start simple, by attending to or using only a sub-set of the relevant information, and build to more complex judgments that integrate more and more of the evidence. Alternatively, children may start with a general, holistic, representation of relations in the evidence (e.g., association) and develop abilities to tune or alter their use of evidence in response to specific task demands. The critical predictions of the two hypotheses concern the relative ease or salience of the associative and conditional probability strategies for making conditional predictions. The holistic-to-selective hypothesis predicts that children will tend to use the degree of association between two variables to make conditional predictions. They will have difficulty selectively focusing on the sub-set of examples relevant to a conditional probability judgment. In contrast, the simple-to-complex hypothesis predicts that young children will either focus on a single, modal, exemplar type or use a strategy of comparing two types of exemplars.

With these hypotheses described, the remainder of this paper reports the results of two experiments exploring young childrens' and adults' conditional predictions. The logic of the studies follows Shaklee's work on diagnosing rules for association judgments. Participants see different frequency distributions of exemplars, and then make conditional predictions. By assessing which distributions support or weaken which predictions we may diagnose the strategies people are using.

[1] If the goal is getting a red marble, then high frequencies of [red in gamble 1] and [blue in gamble 2] support higher ratings for gamble 1, while high frequencies of [red in gamble 2] and [blue in gamble 1] support higher ratings for gamble 2.

4. Experiment 1

This experiment compares conditional predictions under two different frequency distributions of examples. In the "Correlation" distribution all examples fell on one diagonal of a contingency table (e.g., A and D). Thus there is a perfect association between the two feature dimensions (i.e., egg color perfectly predicts diet and vice versa). In the "Partial" distribution only one cell is empty; some examples appear in one of the off-diagonal cells (B-cell, see Fig. 1). The association between the two features is weaker in the Partial than Correlation distributions. However, the conditional probabilities of some features are actually higher (more distinct) in the Partial distribution. Other features have chance-level conditional probabilities in the Partial. The Partial distribution also contains a single modal example-type and some large differences in the base-rates of individual features. In contrast there is no single mode and base-rates are more balanced, hence less informative, in the Correlation distribution.

The Introduction described four strategies for conditional prediction: Modal, base-rate, Conditional Probability, and association. These strategies yield distinctive patterns of responses when applied to the Partial and Correlation distributions (see Fig. 1 for graphical representation of strategies). The modal strategy is to always predict the features of the most frequent exemplar in the Partial condition; since there is no mode in the Correlation distribution, predictions cannot be made (chance-level performance). The base-rate strategy yields the same pattern. Two of the features have higher base-rates than the others in the Partial distribution so children should always predict those. The base-rates are all about equal in the Correlation distribution, so no consistent predictions can be made. The key implication of the Association strategy is that all predictions have the same magnitude (difference from chance). Because there is a strong association between the features in the Correlation distribution, each feature on one dimension can be used to predict the feature on the other dimension. The association is weaker in the Partial distribution, so predictions should be weaker in all cases. The Conditional Probability strategy yields the same predictions as the Association strategy for the Correlation distribution. However, the conditional probabilities vary from certain (1) to chance (.5) in the Partial distribution. Thus the Conditional Probability strategy implies that some predictions in the Partial condition will be strong, while some will not differ from chance.

Participants for this experiment were preschool-aged children, young school-aged children, and adults. Past research has shown some inconsistent results regarding preschooler's probability judgments. These children have not been studied in contingency learning paradigms, so little is known about their processing of frequency distributions. Girotto and Gonzalez (2008) study suggests there is wide variability in how preschool-aged children make conditional predictions. School-aged children have been found to make conditional predictions (e.g., A/[A + B]), rather than association judgments ([A + D]/[B + C]) on causal and contingency learning tasks, suggesting a bias toward

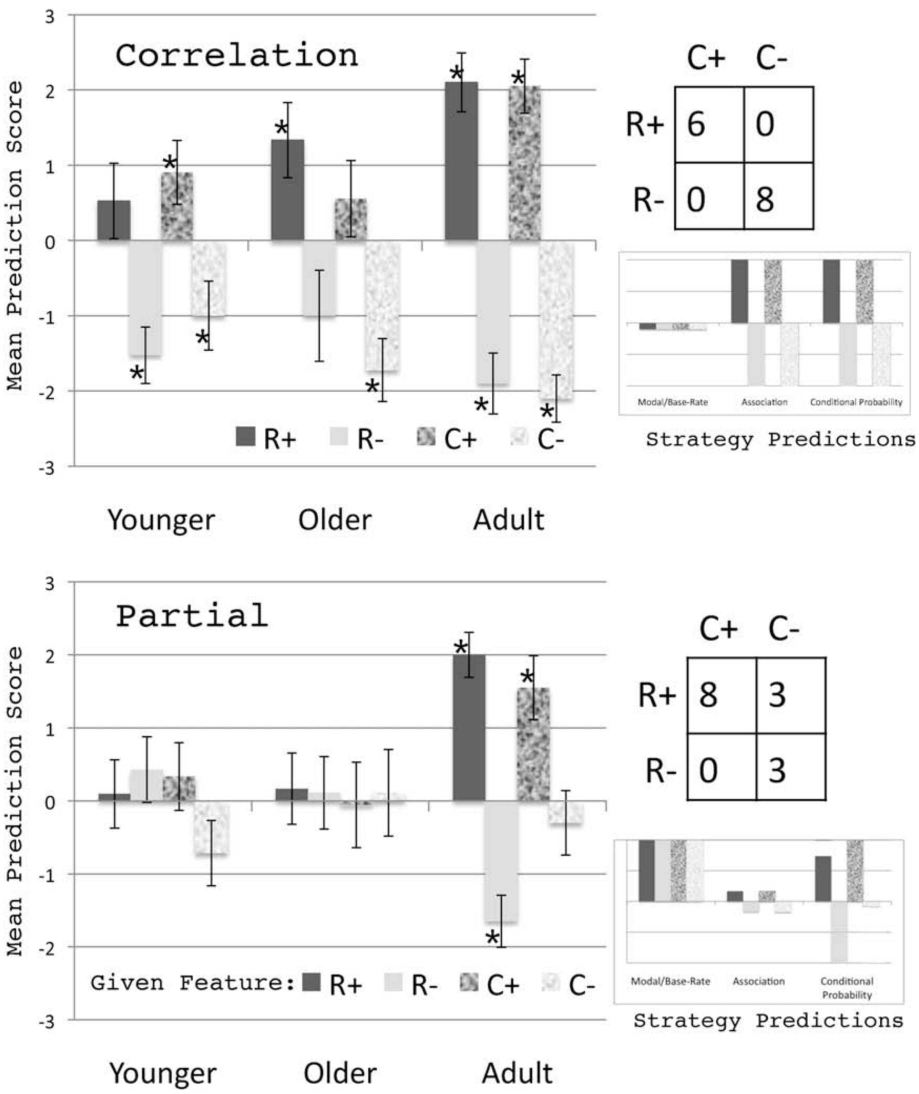

Fig. 1. Mean prediction scores, Experiment 1. Distributions of examples in the learning phases are indicated to right of each pane. Higher scores indicate more frequent/confident predictions of positive feature (R+ or C+, Max 3); Negative scores are net predictions of negative features (Min −3). Error bars indicate one standard error. * Differs from chance (0), $p < .05$, 1-tailed t-test.

attending to conditional probability rather than association in this group.

4.1. Methods

4.1.1. Participants

Twenty-one younger children (Mean Age = 4:6, Range = 4:1–5:2), 18 older children (Mean Age = 8:3, Range = 7:2–8:10), and 20 adults participated. Children were recruited from daycares and afterschool programs serving a largely middle-class population in a mid-sized Midwestern city. Adults were college students participating for extra-credit in courses at a large Midwestern university.

4.1.2. Design

Participants learned about two distributions of exemplars: A set of birds, and a set of fish. The birds had two dimensions: Diet (bugs or berries) and egg color (white or brown). The fish also had two dimensions: Habitat (river or pond) and behavior (active at night, active in the day). In each set one dimension was arbitrarily designated "Row" (diet and habitat), the other: "Column". One level of each dimension was arbitrarily designated the "positive feature", the other "negative feature." Individual features are referred to by their dimension and sign (e.g., R+, R−, C+, C−) Thus the examples can be described as R+C+, R+C−, R−C+, R−C− (corresponding to A, B, C, D cells). The specific assignments of features were: Bugs (R+), Berries (R−), White Eggs (C+), Brown Eggs (C−), Lake (R+), River (R−), Day (C+), Night (C−). Each participant encountered two different frequency distributions (see Fig. 1). In the Correlation distribution, each level of one dimension was paired with a single level of the other. Thus all and only birds who ate bugs laid brown eggs (while berries co-occurred with white eggs), and all and only river fish were active at night (pond co-occurred with day). In the Partial distribution, one level of each dimension predicted perfectly (e.g., bug eaters always had brown eggs, but some berry eaters also laid brown eggs).

Participants encountered both sets of animals, one having a Correlation distribution, and one having a Partial distribution. Content (fish/bird) was randomly paired with distribution across participants, and order of presentation was also randomized. Presentation of each set involved two phases. In the Learning phase, a participant guessed the features of 14 examples (one at a time) and received feedback. For example, in the Bird set, a participant guessed whether the next bird would eat berries or bugs, and would lay brown or white eggs. After making the guess, the actual example appeared. A tally of encountered examples remained visible throughout the Learning phase. After each Learning trial a bird/fish icon was added to one of four virtual boxes, effectively filling in a contingency table. The tally of encountered examples was not visible during the next, Prediction, phase of the experiment.

In the Prediction phase participants were given one feature and asked to predict the other feature. Four questions asked for single-item predictions. Each question presented an example known to have one feature (i.e., brown, white, bugs, berries). The participant then predicted the feature of the other dimension. For example, "We found a bird that eats bugs (R+). Do you think it will lay brown eggs (C+) or white eggs (C−)?" Each participant made all four predictions in random order. Participants also indicated their confidence ("know for sure", "think maybe", or "just guessing). Following the four single-item predictions, participants made two frequency predictions. Participants saw 10 individuals with one feature and predicted how many would have a particular value of the other feature dimension. For example, "We have 10 birds that eat bugs. How many do you think will lay brown eggs?" Participants used a sliding "scroll-bar" to indicate a number; moving the slider caused the appropriate number of examples to be displayed (e.g., moving the slider to five showed five brown eggs). Frequency predictions were made for the two features composing the D-cell exemplars (R− and C−).

4.1.3. Procedure

Adults participated in groups in a computer classroom equipped with 12 individual workstations. Children were tested individually in a quiet location within their child-care sites. Instructions presented the task as game in which explorers were learning about animals on an island. During the Learning phase, participants watched as one of the explorers discovered examples of the animals (birds or fish).

Prior to each discovery the participant guessed what the next example would be like (the two properties it would have). Upon discovery participants learned the true features of the example and the explorer added it to the appropriate box (e.g., "This bird eats berries and lays brown eggs. It goes in this box here."). For children, Learning phase responses earned "Explorer Points" (which could be redeemed at the end of the task for time to play a simple computer game). Critically, at the conclusion of the Learning phase, experimenters had children count the different types of examples. Experimenters read all text for children, explained the displays, and described response options and results of the guesses. In many cases children pointed to the screen and experimenters used a computer mouse to indicate a response. Children comfortable with the interface were allowed to make their own responses.

4.2. Results

Participants' responses were converted into "prediction scores" by combining the feature prediction and confidence rating. The three levels of confidence were assigned numeric values (1, 2, 3; guess, think, know) and assigned a sign based on feature prediction (positive for positive features, negative for negative features). Thus prediction scores ranged from −3 to 3 (no 0). A high score indicates confident prediction of the positive feature. Mean prediction scores are presented in Fig. 1. Primary analyses used these confidence ratings in parametric tests (ANOVA, *t*-tests). Secondary analyses involved non-parametric tests (sign, chi-square) of the binary feature prediction data. Results from secondary analyses always matched the primary analyses, and so are not reported. Condition order (Correlation first or Partial first) did not effect predictions.

Participants' prediction scores may be compared against patterns predicted by the four strategies (see Fig. 1). As an example of how strategy predictions were generated, consider predictions given feature C+ in the Correlation condition of Experiment 1 (first panel of Fig. 1). In this case, the participant is told that the object has feature C+ (e.g., lays brown eggs). The task is to decide whether the object has feature R+ or R− (e.g., eats bugs or berries). On the modal strategy R+ and R− are both possible, because there is no real mode (there were six instances of R+C+, and eight of R−C−). Similarly, the base-rate strategy would also be split (overall six exemplars had R+, while eight had R−). The conditional probability strategy would make a clear prediction of R+ because of the six exemplars with C+, six had R+ and 0 had R−. The association strategy also predicts R+ because the correlation between the two dimensions is perfect: C+ is always associated with R+. Note that these strategy predictions are qualitative only; they suggest when predictions should be different from chance, and the direction (+ or −) of that difference.

The basic quantitative measure of learning from the exemplars is whether predictions differed given one feature value (e.g., berries) than given the other (e.g., bugs). "Given-feature" refers to the known feature in the single-item prediction (e.g., "this bird eats bugs, what color eggs does it lay?", "bug" is the Given-feature). To simplify analyses, data from the Correlation and Partial conditions were analyzed separately. The first analysis considered single-item predictions in the Correlation condition. An ANOVA with Age and Content (Bird, Fish) as between-subjects variables and Given-Feature (four levels R+, R−, C+, C−) as a within-subjects variable revealed a main effect of Given-Feature, $F(3, 159) = 30.2$, $\eta_p^2 = .36$, $p < .001$. Consistent with the distributions shown in the Learning phase, participants were more likely to predict the positive value of one feature when given the positive value of the other than when given the negative value (e.g., R+ given C+ rather than C−, see Fig. 1). Similarly participants made more negative feature predictions given negative features than given positive (e.g., R− given C− than C+, all comparisons $p < .05$, Tukey's HSD). The ANOVA also revealed a small

Age X Given-Feature interaction, $F(6, 159) = 2.7$, $\eta_p^2 = .09$, $p < .05$. Simple-effects showed that Adults made a greater distinction in their predictions given one of the feature-values (C+) than did children. However, the general pattern of more + predictions given + features, and more – predictions given – features held for all three age groups (all comparisons $p < .05$, Tukey's HSD). The comparisons against chance-level responding (see Fig. 1) also indicate that adults were more consistent in their predictions, but that participants in all age groups made predictions in accord with the distribution shown in the Learning phase. The only other significant effect from the ANOVA was a small Content by Given-Feature interaction, $F(3, 159) = 3.5$, $\eta_p^2 = .06$, $p < .05$ indicating somewhat greater distinction among Given-Features for the Bird than for the Fish items. Overall, participants learned the relations in the Correlation distribution and used this distribution to make consistent conditional predictions. This pattern is consistent with both the Association and Conditional Probability strategies, and inconsistent with a focus on modal exemplars or base-rates.

Participants showed sensitivity to the distribution presented in the Learning phase when making Frequency predictions as well. Fifty-percent represents a plausible chance rate for each level of a binary feature. On average, adults predicted that 81% of individuals known to have R− would have C−, and that the frequency of R− in a group known to all have C− would be 77%, both $p < .05$, 1-tailed t-tests vs. chance. Children predicted C− at greater than chance levels given R− (younger: 72%, older: 67%), but did not reliably predict greater than chance levels of R− given C− (younger: 40%, older: 53%). Finally, guesses in the Learning phase also indicate sensitivity to the distribution. Recall that participants predicted the features of the next example to be encountered in this phase. Did people predict likely combinations of features (R+C+ and R−C−) more than unlikely ones (R+C− or R−C+)? This test compares the "on" and "off" diagonals and is akin to a Phi-coefficient or a Fisher's exact test. Adults predicted the frequent feature combinations on 85% of trials (vs. 15% infrequent combinations, $t(19) = 9.5$, $p < .001$). Older children predicted frequent combinations on 63% of trials, and younger children did so on 69% of trials, $t(17) = 2.3$, $p < .05$, and $t(20) = 4.3$, $p < .001$, respectively.

In contrast to the good learning displayed in the Correlation condition, children made no systematic predictions in the Partial condition (see bottom panel of Fig. 1). An ANOVA with Age and Content as between-subjects variables, and Given-Feature as a within-subjects variable showed a weak main effect of Given-Feature, $F(3, 159) = 3.1$, $\eta_p^2 = .05$, $p < .05$. This main effect was conditioned by a stronger interaction between Age and Given-Feature, $F(6, 159) = 4.5$, $\eta_p^2 = .15$, $p < .005$. The simple-effect of Given-Feature was significant only for adults, $F(3, 159) = 12.0$, $\eta_p^2 = .18$, $p < .001$. Children's predictions did not differ across the four given features, younger: $F(3, 159) = .84$, older: $F(3, 159) = .02$. Children's responding most closely matches the Association strategy; adults' best matches Conditional Probability.

In the Partial condition, the predictive power of Given-features varies; C+ strongly predicts R+ (8:0 in the Learning Phase), R− somewhat less strongly predicts C− (3:0), R+ inconsistently predicts C+ (8:3), and C− is not predictive (3:3 for R+ and R−). As indicated by the chance comparisons for Single questions (in Fig. 1), adults recognized all three predictive relations. Also consistent with examples in the Learning phase, adults responded at chance given the non-predictive feature (C−). Children's predictions did not differ from chance for any of the Given-features. The same pattern is apparent on the Frequency questions. Adults predicted that 65% of the examples with R− would have C−, but did not expect difference from chance (47%) given C−, $t(19) = 1.9$, $p < .05$ (1-tailed), and $t(19) = -.6$, *ns*, respectively. Moreover, predictions given R− differed significantly from predictions given C−, $t(19) = 2.0$ $p < .05$ (1-tailed). Neither older nor younger children made significantly different predictions in the two cases.

Although children did not make systematic predictions in the Partial condition, there is evidence that they did encode the distribution in the Learning phase. R+C+ examples were the most common; R−C+ examples never occurred. During the Learning phase, adults and younger children were significantly more likely to predict R+C+ than R+C−, 36% vs. 20% $t(19) = 2.3$, and 34% vs. 22% $t(20) = 2.6$, both $p < .05$, respectively. Older children also selected R+C+ more often (31% vs. 22%), but the difference did not reach statistical significance.

4.3. Discussion

Participants in all three age groups learned predictive relations when there was a perfect correlation between two binary feature dimensions. When each feature always and only occurred with another, even preschool-aged children used this relation to make predictions. Adults also learned predictive relations when only some features perfectly co-varied. When the distribution supported partial prediction, neither preschool-aged nor young school-aged children reliably used one feature to predict another. This failure occurred despite the fact that some feature combinations occurred much more frequently than did others in the Learning phase.

Participant's predictions were inconsistent with the Modal and Base-rate strategies. These strategies would have led to more reliable predictions in the Partial than the Correlation distributions, the opposite of the observed pattern. Adults clearly responded using the Conditional Probability strategy: They made all four conditional predictions in the Correlation distribution but all and only those involving a conditional probability difference in the Partial distribution. Children's responses seemed most consistent with the Association strategy. The overall association between features decreased from the Correlation to Partial distribution. This decrease had a general effect on children's predictions; they made reliable predictions when the association was strong, but no reliable predictions when the association was weaker. Modal, Base-rate, and Conditional Probability strategies all would have supported at least one reliable prediction in the Partial distribution (i.e., R+ given C+). However, children did not make this, or any other, prediction given a partial correlation in the Learning phase. Note that a combination of these three "non-Association" strategies would also have lead to reliable prediction in the Partial distribution. Children did not reliably predict features when there was a strong conditional probability of that feature, when that feature had a higher base-rate, and when that feature characterized the modal exemplar. The results from Experiment 1 support the Association hypothesis: Children succeed in conditional prediction tasks when there is a strong association in the data (and fail otherwise).

Children's failure to make consistent predictions in the Partial distribution is somewhat surprising. There was a reasonably strong association between the features ($\chi^2(1) > 4$) so even the Association strategy might have been expected to yield reliable predictions. Perhaps the Partial distribution was just difficult. One possibility is that children did not encode the frequency distribution in the Learning phase of the condition. There is some evidence against this possibility in that young children did preferentially select the highest frequency examples (older children also selected these examples most frequently, but the difference was not statistically significant). Still, it remains possible that children did not remember the frequencies by the time they got to the Prediction phase. It would be useful to have better information about what children remembered from the Learning phase. A classic question in the literature is whether developmental differences in predictions are due to changes in the accuracy of children's memories for frequencies, or are due to the ways children use those memories (see Reyna & Brainerd, 1994 for review). A second, related, possibility is that the information in the Partial condition was more complex than in the Correlation condition. In the Correlation condition the Learning phase presented only two kinds of exemplars (R+C+ and R−C−). In the Partial condition there were three kinds of exemplars. In the Correlation condition, the two feature dimensions could be recoded into a single dimension (as there is a perfect correlation). The Partial condition required keeping all four features distinct. More carefully probing memory for examples will also provide information about the role of complexity of the stimuli. It is also possible to support encoding by making some features more salient than others. Both these strategies are pursued in Experiment 2.

5. Experiment 2

Did poor performance in the Partial condition of Experiment 1 indicate reliance on the Association strategy, or was it just an indication that the task was too difficult? Would children use other strategies (i.e., Conditional Probability) if task demands were reduced or would they continue to predict based on association? Conditional probability (as opposed to association) seems particularly relevant in thinking about properties and categories. For example, there are many instances in which a feature that is diagnostic of category membership (sufficient) is not necessary (and vice versa). We might expect that children would be sensitive to such asymmetries and distinguish *p*(category|property) from *p*(property|category). Perhaps predicting properties from category membership, or membership from property possession, would facilitate children's use of examples. Experiment 1 required children to attend to four distinct feature-values. Experiment 2 simplified the demands of the task by using only two features, each of which could be present or absent. Thus, Experiment 2 asks whether young children can learn that category membership predicts property possession even in contexts in which property possession does not predict membership (and vice versa).

Experiment 2 uses a more balanced frequency distribution. In Experiment 1, one kind of example appeared much more frequently than did any other. On the one hand, past research suggested that children might have focused on this modal example (Shaklee & Mims, 1981). However, children may also have failed to attend to the exact nature of the non-modal examples. That is, they may have simply remembered that most of the birds ate berries and laid brown eggs, but not all did. Encoding the stimuli as "berry/brown" or "not" provides no basis for distinguishing the different conditional prediction

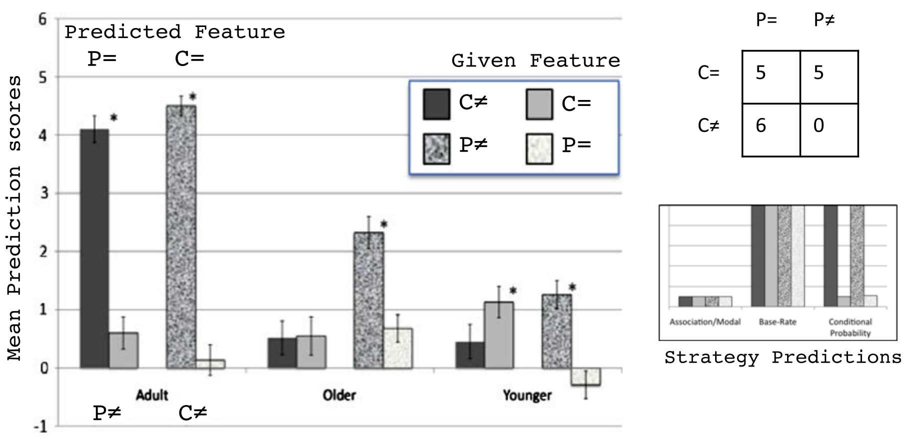

Fig. 2. Mean prediction scores, Experiment 2. Higher scores indicate more frequent/confident predictions of the high frequency features (*P*= and *C*=). Given indicates which feature was known (e.g., "We know this shell has spots" is given *P*=). "*P*" and "*C*" refer to property and category, respectively. "≠" and "=" refer to the magnitude of the frequency differences in the distributions. Error bars indicate one standard error. * Differs from chance (0), $p < .05$, 2-tailed *t*-test. Scores range from −6 to 6 because each participant made two predictions of each type (with fish and with shell content).

questions (what happens if the eggs are white? What happens if the bird eats bugs?). A more balanced distribution might allow children to identify specific, partial relations. That is, children may pay more attention to just which kinds of non-modal exemplars they actually encountered. This concern about encoding illustrates that knowing just what children were attending to during the Learning phase is critical to understanding their patterns of predictions. Experiment 2 included a memory probe at the end of the experiment, asking for recall of the frequencies of different types of examples encountered during learning.

The basic question motivating Experiment 2 was whether children would use conditional probabilities to make predictions when the task demands were reduced (relative to Experiment 1). Use of conditional probabilities implies differential prediction: The distribution of exemplars allows reliable prediction of some features but not others. For example, property possession might predict category membership, but category membership may not predict property possession. In contrast, use of association implies the same reliability of prediction for all features (as was shown in Experiment 1): The association between property possession and category membership is symmetric. The changes from Experiment 1 to Experiment 2 may make it easier for children to detect an association, or more difficult, but that relative change should be the same for all four conditional predictions. Although the specific cell frequencies have changed, the qualitative pattern of responses generated by the Conditional Probability and Association strategies remain the same as in Experiment 1. The more balanced frequency distribution means there is no modal exemplar, thus the Modal strategy would result in chance-level performance. Finally, base-rates of the individual features do vary (by as much as 3:1). In Experiment 2, the Base-rate strategy would be to predict the feature with the balanced distribution (indicated by the '=' symbol in Fig. 2), as there are always more instances with this feature than the other.

5.1. Methods

5.1.1. Participants

Thirty-one younger children (Mean Age = 4:8, range 4:1–5:6) and 31 older children (Mean Age 7:10, range 6:8–8:9) participated in the experiment. Children were recruited from the same local preschool and afterschool program population as were those children participating in Experiment 1. Thirty adults also participated. Adults were sampled from the same classes in a university population as used in Experiment 1.

5.1.2. Design

Experiment 2 was generally similar to Experiment 1. The major differences were the frequency distributions of examples, and the labels provided for the examples. One feature dimension involved category labels (e.g., *Krisser* shell). The other dimension involved property labels (e.g., is spotted). Each dimension had a single value that could be present or absent (e.g., Krisser shell or not, spotted or not). These feature-values are designated: P=, $P\neq$, C=, and $C\neq$. The = and $\neq$ designation refers to the frequency distribution.[2] The overall distribution of exemplars was similar to the Partial structure (from Experiment 1), with the difference that absolute frequencies were roughly balanced (see Fig. 2). Thus a distribution contained three feature combinations that appeared with roughly equal frequency, and a single feature combination that never appeared. The critical feature of this design is that there are two cases of equal conditional probabilities and two cases of unequal conditional probabilities. These relations are illustrated in Fig. 2. Of the exemplars with the feature P = (e.g., "spotted"), about half have the property C = (e.g., Krisser shell) and half have the feature $C\neq$ (e.g., not a Krissershell). Of the exemplars with the feature "$P\neq$" (e.g., "not spotted"), all have the property C=, none have the property $C\neq$. The experiment has a 2×2 design: Dimension (category or property) crossed with Frequency Distribution (= or $\neq$).

The distribution of exemplars means that a strategy of using conditional probabilities to make conditional predictions would result in consistent and confident predictions given "$\neq$" frequency features ($C\neq$ and $P\neq$) and chance-level performance given "=" frequency features (C= and P=). With this distribution there is a relatively weak association between property and category (weaker than the Partial condition from Experiment 1) thus the prediction of the Association strategy is chance-level performance for all predictions. There is no modal exemplar type; the Modal strategy would also yield chance-level performance on all tasks. A strategy of attending to Base-rates would yield non-random performance because the marginal frequencies are unequal. However this strategy predicts no difference between the "=" and "$\neq$" cases (e.g., C= predicted for both P= and $P\neq$). These strategy predictions are illustrated in Fig. 2.

As in Experiment 1, in the Prediction phase participants heard about one example with each feature and predicted presence/absence of the other feature (the single-item predictions). Each single-item prediction may be identified by the given dimension (property or category) and by the frequency distribution characteristic of this feature in the Learning phase (= or $\neq$). An example of a P= single-item prediction was: "We know that this shell has spots. Is it a Krisser shell or is it not a Krisser shell?" This is a "P" item

[2] The actual assignment of features to = and $\neq$ distributions was counterbalanced across participants. Half the participants encountered positive property distributions in which all category members possessed the property: There was a large frequency difference in favor of property possession given category membership (5:0). Half the participants encountered positive category distributions in which all examples with the property were category members. Critically each distribution allows prediction of property from one level of category (either presence or absence), and prediction of category from one level of property (either presence or absence). The "predictable" conditionals in the positive property distribution were p(category|property possession) and p(property|category non-membership). The predictable conditionals in the positive category distribution were p(category|property absence) and p(property|category membership). Each participant encountered two blocks of the same distribution (once with fish examples, and once with shell examples, see Materials, below). The main purpose of the condition manipulation was to control for any differences between making inferences from or to present versus absent features (e.g., "is a Targa" vs. "is not a Targa"). This condition manipulation did not affect any results so is not considered further.

because the property is given ("spots"). This is an = item because in the Learning phase half the spotted shells were Krissers and half were not.

One major design difference in Experiment 2 was inclusion of a memory probe in the Prediction phase. Rather than also asking for frequency predictions (as in Experiment 1), Experiment 2 asked for memory of the Learning phase distribution. Participants responded to a series of six pairwise comparisons. Each comparison presented two feature combinations and asked whether the participant had seen more of one or the other (or both equally) in the Learning phase. For example, one comparison asked whether they had seen more Krisser shells that were spotted, or more Krisser shells that were not spotted (or about the same).

5.1.3. Materials and procedure

The basic procedure was similar to that of Experiment 1. Participants encountered a number of examples in a Learning phase, and then made predictions about future examples in Prediction phase. The Learning phase differed slightly from Experiment 1 in that participants did not guess each feature value separately, but rather guessed which of the four example-types (feature combinations) would appear in the next trial. The Learning Phase began with 10 exposure trials in which one of the four possible examples appeared. Participants confirmed their understanding by moving the example to the correct "box" for its type. Six guessing trials followed the 10 exposures. Pictures of all four possible example appeared and the participant selected one (by clicking with the mouse) as the one that would be next encountered. The actual example then appeared, and was moved to the correct box. The Prediction phase followed the Learning phase. The full experiment involved two blocks, each consisting of a Learning phase and a Prediction phase. Materials for one block were fish examples that varied in size (big or not big) and color (red: Targa fish, white: not Targa fish). Materials for the other block were shells varying in pattern (spotted or not) and contour (spiky: Krisser shells, smooth: not Krisser shells). Note that in one block shape indicates category, while in the other it indicates property. Similarly, in one block color indicates category, and in the other it indicates property. Order of blocks was randomized across participants. In all other respects materials and testing conditions were identical to those of Experiment 1.

5.2. Results

Following Experiment 1, feature predictions and confidence ratings were combined to yield prediction scores. Fig. 2 presents mean scores for the four single-item predictions (given *C*=, *C*≠, *P*=, and *P*≠) predictions. Differences from zero indicate a tendency to predict one feature value rather than the other. Participants' mean prediction scores were analyzed in an ANOVA with Age as a between-subjects variable, and Frequency distribution (=, ≠) and given Dimension (category, property) as a within-subjects variable. This ANOVA revealed a significant effect of Frequency distribution, $F(1, 89) = 34.9$, though this effect was conditioned by an interaction with Age, $F(2, 89) = 14.5$, $\eta_p^2 = .24$, $p < .001$. Only adults made stronger or more consistent predictions for ≠ frequency items than for = frequency items, $F(1, 89) = 29.6$, $p < .001$. both F's for children ⩽1. Dimension did interact with Frequency distribution, $F(1, 89) = 5.8$, $\eta_p^2 = .06$, $p < .05$. Overall participants were more sensitive to Frequency distribution for property-given items than for category-given items. Although children did not make reliable predictions for ≠ frequency items overall, they did show the frequency effect for property-given items. Children reliably predicted the high frequency category for property-given items when there was a unequal frequency in the Learning phase, Both $F(1, 30) = 4.4$, $p < .05$. Neither older nor younger children reliably predicted the high frequency property for category-given items. This same pattern is evident in the comparisons against chance-level responding reported in Fig. 2. Adults predicted high frequency properties at rates greater than chance for both ≠ category-given and property-given items (but showed no reliable difference in prediction for either of the two = Frequency distribution items). Children showed reliable prediction of high frequency properties only for ≠ property-given items.

In summary, children recognized when it was possible to predict an individual's category identity from a property it possessed, and when such predictions could not be made. Moreover, unlike Experiment 1, children were able to recognize the partially predictive structure of their experience. They predicted category membership from properties even when they could not (and did not) predict properties from category membership.

What did children encode from the Learning phase of the task? In each block of the Learning phase, three feature combinations appeared with roughly equal frequencies, and one combination was absent. Overall, participants reliably remembered that the absent combination appeared less often than did the others. Table 2 presents the mean frequencies of correct memory responses for comparisons of absent vs. present combinations. Also shown in Table 2 are the mean prediction scores given both correct and incorrect memory. Even young children showed good memory for the frequency differences in the Learning phase. However, this good memory did not translate into accurate prediction. Even when children remembered that they had seen fewer of one kind of example than another, they did not reliably predict that the more frequent example would occur in the future (i.e., prediction scores did not differ from chance for memory-correct items). Moreover, young children were not significantly more likely to make accurate predictions when they remembered correctly than when they remembered incorrectly, $t(13) = 1.0$, *ns*. Older children, however, did have significantly higher prediction scores when they remembered the frequency differences correctly, $t(11) = 2.6$, $p < .05$. Too few adults ($N = 7$) ever misremembered the frequencies to make a meaningful comparison possible.

5.3. Discussion

Young children did learn partially predictive relations in Experiment 2. They distinguished predictable and un-predictable features based on frequencies of examples

Table 2
Memory performance, Experiment 2.

	Mean proportion of correct memory[a]	Mean prediction score (range −3 to 3)	
		When memory-correct	When memory incorrect
Adult	.92	2.2 (N = 30)	1.3 (N = 7)
Older	.83	.75 (N = 26)	−.74 (N = 12)
Younger	.77	.41 (N = 26)	.31 (N = 14)

[a] Only comparisons involving absent vs. present feature combinations included.

encountered. Specifically, when all of the examples that had a property (e.g., spots) belonged to a category (e.g., Krisser shell) children later predicted that an object with the property would belong to the category. They made this prediction even in the absence of an inverse relation; it was not the case that most category members had the property (e.g., the number of spotted and non-spotted Krisser shells was equal). Children were able to selectively use the encountered frequency distribution to support particular conditional predictions. The results of Experiment 2 contrast with those of Experiment 1 in which children did not learn predictive relationships in the absence of perfect correlation. At least in some circumstances young children will learn specific conditional relations from patterns of co-occurrences, and will distinguish predictive relations from non-predictive ones.

The results of Experiment 2 indicate that children will attend to conditional probabilities when making predictions. Children distinguished predictions involving large probability differences, from those involving small or no differences. The changes from Experiment 1 to Experiment 2 allowed children to detect conditional probabilities in the frequency distributions encountered in the Learning phase of the experiment. Critically, predictions were not always more confident or reliable in Experiment 2 than in Experiment 1. It was not simply that the association between the two dimensions was easier to detect. Rather than responding to the general association in the exemplars, children responded differentially to specific conditional probabilities. Children did not focus on the base-rates or marginal frequencies of features. That Krisser shells were more common overall (about 2:1) did not lead children to always predict a new shell would be a Krisser. Rather their predictions were sensitive to the frequency of Krisser shells conditional on coloration (spots or not). Finally, the results of Experiment 2 are also inconsistent with use of modal exemplars, as there was no single mode in the distribution yet children still showed some reliable predictions.

Despite the attention to conditional probabilities, Experiment 2 also revealed significant limitations in children's predictions. The most striking finding was that children did not learn to predict properties from category membership. This finding is especially curious given other demonstrations in the literature that children readily make category-based inductions, and may have an easier time predicting properties given category membership, than predicting category membership given properties (Gelman, Collman, & Maccoby, 1986). One direction for future research is to explore the task features that might make one conditional prediction more salient than another. Perhaps in tasks involving learning from multiple examples (like the current study) categorization is the more salient prediction. Tasks involving verbally presented information about classes of objects (e.g., including generics such as "birds lay eggs", Gelman, Star, & Flukes, 2002) might dispose children to focus on predicting properties of category members.

Overall, children were less consistent in distinguishing predictive and non-predictive relations than were adults. One hypothesis is that children might simply have a poor memory for example frequencies (Brainerd, 1981). The results of Experiment 2 suggest that memory difficulties cannot entirely account for children's predictions. Children were generally accurate in remembering the frequency differences (over 75% correct for even the youngest group). Moreover only for older children did the consistency and strength of predictions depend on memory performance. Even for older children, though, accurate recall did not always lead to prediction of the higher frequency outcome in the Prediction phase. The dissociation between memory and prediction suggests some other process at work. Children differ from adults in the way they use their memories of example frequencies to generate conditional predictions. The general discussion considers some explanations for children's patterns of performance, and the nature of the developmental differences between children and adults.

6. General discussion

Almost all accounts of learning and cognition emphasize detecting patterns in experience and then projecting those patterns into the future. This kind of inductive inference is likely an important mechanism for children's cognitive development. Thus it is important to understand how children identify patterns, and how they project those patterns into the future. The results from the current study suggest that preschool-aged children can detect and use conditional probabilities among features. However, relative to adults, children may be have greater difficulty detecting such probabilities, and may be less consistent in how they use such observed probabilities to make predictions.

The basic question motivating the current study was how children would use examples to make conditional predictions (predicting one feature given another): Would they focus on a single salient example, associations, base-rates, or conditional probabilities? In Experiment 1, adults, young school-aged, and preschool-aged children all learned and used a consistent association between two two-level dimensions (e.g., diet and egg color). When each level of one dimension always and only co-occurred with a single level of the other dimension, all participants used this relation to make predictions about the composition of novel exemplars. Participants detected the association and projected it into the future. However, when there was only a partial relation between the dimensions, children's future predictions did not differ from chance levels. In the Partial condition features were either necessary or sufficient, but not both: The association was not consistent, but at least one of the four simple conditional probability relations was very consistent in the encountered examples. Children did not reliably use the conditional probability differences to make future predictions. Adults did, recognizing when they could make a confident prediction and when they could not. There was some evidence from children's performance in the Learning phase that they had detected the conditional probabilities in examples. However, they did not use these probability differences to make predictions about new cases. The overall conclusion is that children rely on the association between features when making conditional predictions.

Experiment 2 went on to ask whether children could use conditional probabilities under less demanding conditions. Two major changes were framing the task within the context of properties and category membership, and reducing the number of features to keep track of by using positive and negative feature levels. Thus Experiment 2 presented category members and non-members that either had or lacked a property (e.g., spotted and not spotted Krisser shells and non-Krisser shells). In this case, both preschool-aged and school-aged children did use conditional probabilities; they recognized that some features could be predicted from others, but not all could. This selective prediction is the critical result. In Experiment 2 it was never possible to predict each feature from every other one. On some of the prediction measures the best one could do was a random guess, and, indeed, children's performance was close to chance in such cases. In other cases, one outcome was more likely than the other (at least from the examples encountered in training). Children did recognize at least some of these predicable cases. Thus they distinguished conditional inferences that were supported by experience, from those that were not. Children did not use the overall association as the basis of predictions. In neither experiment did children seem to rely on the modal exemplar or base-rates when making conditional predictions.

The current study complements Girotto and Gonzalez's (2008) recent demonstration that young children do attend to posterior probabilities when making conditional predictions. In that work children did use information about one feature to predict the value of another (i.e., they did not rely on base-rates or unconditional probabilities). However, Girotto and Gonzalez's study was not designed to uncover the basis of those conditional predictions; children could have answered correctly by attending to the association between the features or by attending to specific conditional probabilities. The current study explored the basis of predictions by comparing children's success given different frequency distributions of examples. The overall conclusion is that children have difficulty selecting the relevant pieces of information to use as the basis for a conditional prediction. When a there is a perfect correlation in experience, all examples "point" toward the same prediction. If all and only if spotted shells are Krisser shells (and thus all and only Krissers are spotted) using any sub-set of the examples as the basis of prediction yields the same response. In contrast, when there is only a partial correlation (e.g., one feature is necessary but not sufficient for another) different examples support different predictions. The claim is not that young children are unable to selectively attend to particular sub-sets of examples, but rather that such selection is difficult. With sufficient support (e.g., in Experiment 2) children can focus on a relevant sub-set of examples to make reliable predictions. One important direction for future research is to explore the conditions that support this kind of selection. For example, the current study used a relatively small number of examples. The partially predictive structure of experience may become clearer with more experience.

Young children did attend to conditional probabilities in experience, but they did not do so readily (Experiment 1) or completely (Experiment 2). In some ways these difficulties seem inconsistent with the infancy literature. The literature on infant learning is extremely diverse, but some notable findings are directly relevant to the current experiment. Infants are sensitive to patterns of correlated attributes (Bhatt, WIlk, & Rovee-Collier, 2004; Younger & Cohen, 1983). The key feature of a correlated attribute structure is that there are multiple cues providing redundant information. Kloos and Sloutsky (2008) note that correlated attribute structures are relatively easy to learn because they do not demand selective attention. What makes learning a relation difficult is attending to informative features and screening out uninformative ones. Babies do display sensitivity to very specific statistical relationships (e.g., transition probabilities) in otherwise noisy data (Aslin et al., 1998; Kirkham et al., 2007). These paradigms display information serially, so babies experience one feature regularly following another. Thus one of the possible statistical relations (transition probability of one element following another) is particularly salient and useful in these tasks. Interestingly, recent research suggests that infants are also sensitive to backwards transition probabilities (the probability of one element preceding another, Pelucchi, Hay, & Saffran, 2009). That infants attend to both of these conditional probabilities means that they face the problem of coordinating the two statistics, of deciding which is relevant and what happens when they conflict. The results of the current study suggest that tasks requiring selective attention to, and usage of, different statistical relations may be difficult for infants as they seems to be so for preschool-aged children.

A major difference between the methods in the current study and methods used in the infancy literature is the nature of the response required from participants. The current study asked for a prediction. Methods used with infants typically rely on a same/different or old/new judgment. There was evidence in the current study that young children did detect conditional probability relations in their experience. Children showed sensitivity to the frequency differences both in their (unconditional) guesses about exemplars (in the Learning phase of Experiment 1) and in their memories for exemplar frequencies (Experiment 2). We might expect good performance on tasks that require detection of differences in conditional probabilities. The current study, and prior research on developing probability judgments, suggests that developmental differences appear in the process of forming a response (making a prediction, stating a probability; Reyna & Brainerd, 1994). Thus children may be relatively good at detecting conditional relations, but poor at using such information to make predictions.

The introduction contrasted two general views of the development of inductive reasoning. The first, perhaps more traditional, suggests that young children have difficulty integrating multiple pieces of information; their judgments incorporate only a sub-set of the information used by adults. The alternative is that selection is the key developmental variable. Young children may attend to multiple and diverse pieces of information, but they have difficulty picking out specific relations that are relevant to specific tasks. Children do well when there are multiple redundant cues that support an overall "gist" representation (Kloos & Sloutsky, 2008; Reyna, 2005; Reyna & Brainerd, 1994). In the categorization and inductive inference literatures the "selection development" view is often associated with similarity-based accounts. Young children categorize and make predictions based on the overall similarity between objects, rather than focusing on a subset of criterial features, that might be characterized as a "rule" (Sloutsky, 2003). There is some debate regarding how well children distinguish those features that are merely associated with a category from those that are truly necessary (e.g., causal, essential; see Gelman, 2003). The current study provides a slightly different perspective on this debate by focusing on statistical relations rather than content (e.g., perceptual vs. causal features). Do young children distinguish associated features from those that are necessary (or sufficient)? That is, association is a more general relation composed of two more specific conditional relations: Degree of necessity and degree of sufficiency. If children only encode the more general "associated" relation then they may have difficulties on tasks in which the component relations (necessity and sufficiency) are disassociated, as in the current study. For example, focusing on the general level of association between a feature and a category may give a misleading impression of the ability to predict the feature from the category or the category from the feature.

The results from Experiments 1 and 2 present something of a mixed picture of young children's abilities to use more specific statistical relations of conditional probability (necessity and sufficiency). On the one hand they suggest that necessity supports sufficiency and vice versa. Learning that feature A predicts feature B is easier when feature B also predicts feature A. On the other hand, children were able to learn one conditional relation in the absence of another, at least when predicting category membership. Children learned that spotted shells were likely to be Krissers, even though Krisser shells were not especially likely to be spotted. However, this selective attention to specific conditional probabilities seemed somewhat fragile, dependent on task demands and content. Interestingly, though, even when children did not use specific conditional probabilities, they seemed to encode the relevant information. They remembered the frequency distributions, but did not always use them.

If children generally have accurate memory for the relative frequencies of different kinds of examples, and use such information in some cases, how should we characterize the development of conditional predictions? Does development involve increases in ability, overcoming of biases, or changes in beliefs? The "selection development" accounts suggest an increasing ability to select and focus, to apply different judgment strategies in different contexts (Kloos & Sloutsky, 2008; Reyna, 2005). Yet these accounts do not explain why children showed more selective use of examples in some tasks than in others (e.g., predicting category membership given property rather than vice versa). There might be some general biases to process or use experience for particular purposes; children may just encode information in terms of category prediction, for example. Alternatively, different priors may account for the different predictions made by children and adults. For example, a prior expectation of symmetric or bi-conditional relations (Barrouillet et al., 2002) may account for children's failure to learn relations in the Partial condition of Experiment 1. Placing the task in the context of categorization (in Experiment 2) may have activated different assumptions about different types of relations. Prior knowledge that all but not only members of a category have a particular property (e.g., all zebras are striped, but not all striped animals are zebras) or that only but not all category members have a given property (e.g., only mammals have fur, but not all do) may provide children a good basis for reasoning about conditional probabilities. In contrast, novel properties (especially simple binary features in an experimental setting) may be assumed to have a symmetric and perfectly correlated relation. The current study demonstrates that young children can pick out conditional probabilities from frequencies of co-occurrence in experience, but that detecting such probabilities is somewhat difficult. Whether such difficulty is best understood as a capacity limitation, a bias, or as the result of specific beliefs and knowledge remain questions for future research.

This study explored the kinds of relations children learn from experience. At least for the task used in the two experiments, young children tended to use bi-conditional relations of association. The two experiments did not reveal any tendency to focus on absolute frequencies either of features (base-rates) or of example-types (modal exemplars). Young children were able to learn and use conditional probabilities, at least in the context of predicting category membership from properties. The critical ability demonstrated by attending to conditional probabilities is selection. Only some kinds of example are informative about conditional probabilities; to make conditional predictions children must selectively attend to relevant evidence. In contrast, all examples contribute to an association and, every example is informative about the absolute frequencies of the properties it instantiates. Put another way, evidence may fail to support a consistent association but still allow (some) conditional predictions. Determining which pieces of evidence are relevant to some hypothesis, and how different types of evidence are to be weighted, are really the hard problems of inductive inference. The positive results of Experiment 2 demonstrate that young children do use evidence selectively, depending on the judgment involved. They may not always select and use evidence in the same ways as adults, but they are not indiscriminate. A promising direction for future research is to describe the principles that guide children's attention to, and use of, examples when making inductive judgments.

Acknowledgments

Thanks to Sunae Kim for her help with design and data collection and to Chris Lawson for comments on a previous draft. This research was supported by a Grant from the NSF (DRM/DLS) 0745423.

References

Acredolo, C., O'Connor, J., Banks, L., & Horobin, K. (1989). Children's ability to make probability estimates: Skills revealed through application of Anderson's functional measurement methodology. *Child Development, 60*, 933.

Aslin, R. N., Saffran, J. R., & Newport, E. L. (1998). Computation of Conditional Probability statistics by 8-month-old infants. *Psychological Science, 9*, 321–324.

Barrouillet, P., & Lecas, J.-F. o. (2002). Content and context effects in children's and adults' conditional reasoning. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology, 55*, 839.

Bhatt, R. S., WIlk, A., & Rovee-Collier, C. (2004). Correlated attributes and categorization in the first half-year of life. *Developmental Psychology, 44*, 103–115.

Brainerd, C. J. (1981). Working memory and the developmental analysis of probability judgment. *Psychological Review, 88*, 463–502.

Cheng, P. W., & Novick, L. R. (1990). A probabilistic contrast model of causal induction. *Journal of Personality and Social Psychology, 58*, 545–567.

Cheng, P. W., & Novick, L. R. (1992). Covariation in natural causal induction. *Psychological Review, 99*, 365–382.

Dawes, R. M., Mirels, H. L., Gold, E., & Donahue, E. (1993). Equating inverse probabilities in implicit personality judgment. *Psychological Science, 4*, 396–400.

De Houwer, J., Vandorpe, S., & Beckers, T. (2005). On the role of controlled cognitive processes in human associative learning. In A. J. Wills (Ed.), *New directions in human associative learning* (pp. 41–63). Mahwah, NJ: Lawrence Erlbaum Associates Publishers.

Dougherty, M. R. P., Gettys, C. F., & Ogden, E. E. (1999). MINERVA-DM: A memory processes model for judgments of likelihood. *Psychological Review, 106*, 180.

Falk, R., & Wilkening, F. (1998). Children's construction of fair chances: Adjusting probabilities. *Developmental Psychology, 34*, 1340–1357.

Gelman, S. A. (2003). *The essential child: Origins of essentialism in everyday thought*. Oxford, New York: Oxford University Press.

Gelman, S. A., Collman, P., & Maccoby, E. E. (1986). Inferring properties from categories versus inferring categories from properties: The case of gender. *Child Development, 57*, 396–404.

Gelman, S. A., & Kalish, C. W. (2006). Conceptual development. In D. Kuhn & R. S. Siegler (Eds.), *Handbook of child psychology: Cognition, perception, and language* (6th ed., Vol. 2, pp. 687–733). Hoboken, NJ: Wiley.

Gelman, S. A., & Medin, D. L. (1993). What's so essential about essentialism? A different perspective on the interaction of perception, language, and conceptual knowledge. *Cognitive Development, 8*, 157–167.

Gelman, S. A., Star, J. R., & Flukes, J. E. (2002). Children's use of generics in inductive inferences. *Journal of Cognition and Development, 3*, 179–199.

Girotto, V., & Gonzalez, M. (2008). Children's understanding of posterior probability. *Cognition, 106*, 325–344.

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.

Kirkham, N. Z., Slemmer, J. A., Richardson, D. C., & Johnson, S. P. (2007). Location, location, location: Development of spatiotemporal sequence learning in infancy. *Child Development, 78*, 1559–1571.

Kloos, H., & Sloutsky, V. M. (2008). What's behind different kinds of kinds: Effects of statistical density on learning and representation of categories. *Journal of Experimental Psychology: General, 137*, 52–72.

Luce, R. D. (1963). Detection and recognition. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.). *Handbook of mathematical psychology* (Vol. 1). New York: Wiley.

Markman, E. (1989). *Categorization and naming in children*. Cambridge, MA: MIT Press.

Pelucchi, B., Hay, J. F., & Saffran, J. R. (2009). Learning in reverse: Eight-month-old infants track backward transitional probabilities. *Cognition, 113*, 244–247.

Reyna, V. F. (2005). Fuzzy-trace theory, judgment, and decision-making: A dual-processes approach. In C. Izawa & N. Ohta (Eds.), *Human learning and memory: Advances in theory and application: The 4th Tsukuba international conference on memory* (pp. 239–256). Mahwah, NJ: Lawrence Erlbaum Associates Publishers.

Reyna, V. F., & Brainerd, C. J. (1994). The origins of probability judgment: A review of data and theories. In G. Wright & P. Ayton (Eds.), *Subjective probability* (pp. 239–272). Oxford, England: John Wiley & Sons.

Rogers, T. T., & McClelland, J. L. (2004). *Semantic cognition: A parallel distributed processing approach*. Cambridge, MA: MIT Press.

Schlottmann, A. (2001). Children's probability intuitions: Understanding the expected value of complex gambles. *Child Development, 72*, 103.

Shaklee, H., & Goldston, D. (1989). Development in causal reasoning: Information sampling and judgment rule. *Cognitive Development, 4*, 269–281.

Shaklee, H., & Mims, M. (1981). Development of rule use in judgments of covariation between events. *Child Development, 52*, 317.

Shaklee, H., & Paszek, D. (1985). Covariation judgment: Systematic rule use in middle childhood. *Child Development, 56*, 1229–1240.

Sloutsky, V. M. (2003). The role of similarity in the development of categorization. *Trends in Cognitive Sciences, 7*, 246–251.

Spinillo, A. G. (2002). Children's use of part–part comparisons to estimate probability. *The Journal of Mathematical Behavior, 21*, 357–369.

Vadillo, M. A., & Matute, H. (2007). Predictions and causal estimations are not supported by the same associative structure. *The Quarterly Journal of Experimental Psychology, 60*, 433–447 (Empirical Study Quantitative Study).

Vadillo, M. A., Miller, R. R., & Matute, H. (2005). Causal and predictive-value judgments, but not predictions, are based on cue-outcome contingency. *Learning & Behavior, 33*, 172–183 (Empirical Study Quantitative Study).

Villejoubert, G., & Mandel, D. R. (2002). The inverse fallacy: An account of deviations from Bayes's theorem and the additivity principle. *Memory & Cognition, 30*, 171–178.

Xu, F., & Denison, S. (2009). Statistical inference and sensitivity to sampling in 11-month-old infants. *Cognition, 112*, 97–104.

Younger, B. A., & Cohen, L. B. (1983). Infant perception of correlations among attributes. *Child Development, 54*, 858–869.