COGNITIVE DEVELOPMENT

Cognitive Development 16 (2002) 907–928

Children’s solutions of logical versus empirical problems: What’s missing and what develops?[☆]

Bradley J. Morris[a,*], Vladimir Sloutsky[b]

[a] *Department of Psychology, Carnegie Mellon University and The Ohio State University, Columbus, OH 43210, USA*

[b] *Center for Cognitive Science, The Ohio State University, Columbus, OH 43210, USA*

Received 1 June 1999; received in revised form 1 August 2001; accepted 1 October 2001

Abstract

This research examined the development of the ability to differentiate logical from empirical problems and the different ways in which children solve these problems. Thirty-two 4- and 5-year-olds, thirty-four 8- and 9-year-olds, and thirty-five 11- and 12-year-olds were given five questions regarding an imaginary character’s predictions as to where a ball would land after being dropped through a ‘tautology machine’. The questions examined encoding and recall of problems, children’s understanding of when evidence was necessary, and children’s evaluation of form and evidence. Data were analyzed in two ways: (1) by comparing differences across participants on component questions and (2) an individual analysis examining the consistency of responses to component questions across the problem set. Overall, the results indicated that: (1) sixth graders tended to differentiate logical from empirical problems while preschool and third grade children rarely did; (2) young children tend to ignore both the logical connective and the second half of problems-termed a ‘cut;’ (3) these cuts are less frequent when a problem is compatible with one empirical possibility; (4) cuts do not stem from encoding or recall errors, but seem to be the product of incomplete problem processing and (5) from third

---

[☆] This research was a part of Ph.D. thesis of the first author submitted to the Ohio State University and was supported by the George I. Miller and Florence Levy Miller Memorial Fellowship.

[*] Corresponding author. Present address: Learning Research & Development Center, University of Pittsburgh, 3939 O’Hara Street, LRDC Building, Pittsburgh, PA 15260, USA. Fax: +1-412-684-1470.

E-mail address: bjmorris@pitt.edu (B.J. Morris).

PII: S0885-2014(01)00070-3

to sixth grade, children's understanding of logical form increased as the rate of cuts decreased.

Keywords: Children; Logical problem; Empirical statement

1. Introduction

This paper focuses on young children's ability to differentiate logical from empirical statements and solve simple problems with both types of statements. Specifically, we examine the relationship between children's ability to map verbal descriptions onto states of affairs in the world, which is a cornerstone of semantic development, and their ability to evaluate these mappings. These mappings can be evaluated by presenting participants with reasoning problems, such as: (1) deriving conclusions from a set of premises, (2) assigning of truth-values (true or false) to statements, or (3) assigning modal operators (Possible, Impossible, Necessary, and Unnecessary) to statements (see Evans, Newstead, & Byrne, 1993, for a review of research with each type of reasoning problem). In this research, we have specifically focused on the second type of reasoning problem, assigning truth-values.

The simplest reasoning problems involve mapping a verbal description onto a single state of affairs. This type of problem is empirically determinate. Reasoning is more complex if a verbal description can be mapped onto multiple states of affairs, making the problem empirically indeterminate. A problem is empirically determinate if it corresponds to exactly one empirical possibility; otherwise, it is empirically indeterminate (Piéraut-LeBonniec, 1980). For example, if an object is drawn from a box that has only blocks, and the goal is to determine whether or not a block was drawn from the box without seeing the drawn object, the problem is determinate — the drawn object has to be a block. On the other hand, if the box has blocks and circles, and the goal is the same, the problem is indeterminate — the drawn object may or may not be a block. Although young children are quite good at reasoning with simple empirically-determinate problems (Fay & Klahr, 1996), they often fail to appreciate empirical indeterminacy by mistaking indeterminate problems for determinate ones, and they have more difficulty solving indeterminate problems than determinate problems (Bindra, Clarke, & Schultz, 1980; Byrnes & Overton, 1986; Fay & Klahr, 1996; Piéraut-LeBonniec, 1980). For example, Fay and Klahr (1996) used a task requiring children to assign modal operators. These researchers found that young children mistook merely possible solutions (i.e., a block *can be* drawn from a box of blocks and circles) for necessary ones (i.e., a block *was* drawn from a box of blocks and circles rather than a box of blocks and triangles). Specifically, children often erred by drawing a conclusion based on the first match to the evidence instead of considering other possibilities. Thus, young children were found to confuse determinate problems (whose solutions are necessary) with indeterminate ones (whose solutions are possible).

There is also another class of problems that can yield a single solution. These are logical problems whose truth-values can be determined on the basis of their logical form, without mapping them onto states of affairs. These are logically determinate problems, which are true with logical necessity. Any valid logical argument is logically determinate if its truth-value can be determined based on its logical form. For example, in the argument of the form “If A then B. A”, the conclusion that “B is the case” is true with logical necessity. At the same time, an invalid argument “If A then B. B”, is both logically and empirically indeterminate because no firm conclusion can be drawn from this argument. A special case of logically determinate problems are simple tautologies and contradictions, which consist of a single statement, and whose truth-value can be determined on the basis of their logical form. At the same time, determining the truth-value of tautologies and contradictions by mapping them onto states of affairs is an impossible task: these statements correspond either to an infinite number of empirical possibilities (in the case of tautologies) or to no empirical possibilities (contradictions). For example, the tautology “Fido is a dog or is not a dog” is true independent of the empirical state of affairs but corresponds to any empirical possibility (i.e., it is indeterminate), whereas a contradiction “Fido is a dog and is not a dog” is false independent of empirical states of affairs and corresponds to no empirical possibility (i.e., it is impossible). Thus, these problems are determinate if they are solved logically; empirical evidence plays no role in establishing their truth-value. It seems that tautologies and contradictions are basic logical statements, and the ability to assign truth-values to these statements should precede the ability to assign truth-values to more lengthy multi-statement arguments. This research focuses on the ability to assign truth-values to tautologies and contradictions, and on the development of this ability from early childhood to preadolescence.

In order to correctly assign truth-values to logically determinate problems, children (as well as adults) must understand logical determinacy. However, researchers have demonstrated that children (and many adolescents) do not fully understand logical determinacy (Byrnes & Overton, 1986; Moshman & Franks, 1986; Morris & Sloutsky, 1998; Ruffman, 1999), often attempting to provide empirical solutions to logically determinate problems (Osherson & Markman, 1975; Markman, 1978; Morris & Sloutsky, 1998). It seems that understanding of logical determinacy includes the ability to differentiate logical and empirical problems, understand logical form, possibility and necessity, and the ability to ignore conversational interpretations of statements and arguments. However, children have a limited understanding of each of these requisite components.

First, children poorly differentiate logical from empirical problems. On the one hand, they often attempt to solve logical problems empirically (Morris & Sloutsky, 1998; Osherson & Markman, 1975). On the other hand, they often construct incomplete problem representations that limit the number of states of affairs compatible with the problem (Johnson-Laird & Byrne, 1991; Markovits, Fleury, Quinn & Venet, 1998). These incomplete representations are often “minimalist”, such that participants represent a single possibility compatible with a problem, thus transforming indeterminate problems into determinate ones (Sloutsky & Goldvarg, 1999). These incomplete representations lead to reasoning errors as participants fail to consider alternative possibilities compatible with a problem (Markovits, 1988; Wason & Johnson-Laird, 1972). For example, in conditional reasoning, children and adults have no difficulty concluding that “B is the case”, given that “If A then B and A”. On the other hand, they fail to recognize that the argument “If A then B and B” is indeterminate, and they erroneously conclude that “A is the case” (Rader & Sloutsky, in press; see also Evans et al., 1993, for reviews). These errors, however, can be reduced by explicitly introducing alternatives or by cueing counterexamples using a familiar content (Byrnes & Overton, 1986; Markovits et al., 1998). For example, Horobin & Acredolo (1989) found that young children could consider more than one possible solution (when given explicit prompts). However, even with prompts children often did not consider more than one possibility.

Participants in studies of scientific reasoning also construct incomplete problem representations, limit their search in both problem space and memory space, and seldom spontaneously search for counterexamples (Kuhn, Garcia-Mila, Zohar, & Andersen, 1995; Markovits, 1988; Mynatt, Doherty, & Tweney, 1977). Children and adults exhibit a ‘positive capture’ strategy by accepting one possibility before considering equally plausible alternatives (Bindra et al., 1980; Fay & Klahr, 1996; Klahr & Dunbar, 1988). Similarly, in practical reasoning (in a task to compare the size of foreign cities) people often made fast and frugal inferences, relying on a small number of the most salient predictors (such as the familiarity of the city), while ignoring other information (Gigerenzer & Goldstein, 1996).

Second, children have limited understanding of logical connectives and/or modal operators of possibility and necessity (Inhelder & Piaget, 1958; Osherson & Markman, 1975; however, see Braine & O’Brien, 1998). In particular, young children demonstrate a limited understanding of basic logical connectives in natural conversation by ages 4–5 with an understanding of *If* and *And* preceding an understanding of *Or* (Johansson & Sjolin, 1975; Scholnick & Wing, 1992; Braine & O’Brien, 1998). Children of this age have difficulty assigning correct truth-values for logical connectives (Inhelder & Piaget, 1958; Osherson & Markman, 1975; however, see Braine & O’Brien, 1998). They also have difficulty understanding that some statements happen to be true (they are merely possible), whereas others are true with necessity.

Third, children exhibit difficulty in distinguishing conversational from logical interpretations of connectives: they can often use logical connectives correctly within conversational pragmatics, yet fail to extend this understanding the assignment of truth-values (Scholnick & Wing, 1991, 1992; Braine & O’Brien, 1998). For example, preschool children interpreted contradictory statements pragmatically by treating them as if each half of the statement referred to a separate element of a larger whole (Sharpe, Eakin, Saragovi, & Macnamara, 1996). Given the statement “The dinner was good and wasn’t good”, preschool children offered interpretations such as “The salad was good, yet the dessert was bad”, resolving the apparent contradiction. While helping the child to cope with a statement, pragmatic considerations do not help when the task is to assign truth-values. In particular, to answer whether or not a proposition is true, one may have to ignore conversational or pragmatic considerations in order to focus on the formal elements of a problem (Scholnick & Wing, 1991; Braine & O'Brien, 1998). Children are more likely to rely on conversational factors than they are to rely on the formal structure.

In short, children have limited understanding of logical determinacy and empirical indeterminacy. Lacking understanding of logical determinacy, children might treat logically determinate problems as empirically solvable and therefore draw a conclusion from the first match between a part of the statement and the state of affairs — in a manner similar to their solutions to empirically indeterminate problems. This limited understanding of logical determinacy drastically limits young children's reasoning on logically determinate problems. For example, a child who treats a logically determinate problem (e.g., a tautology) as an empirical problem will encounter a challenge: there is an infinite number of compatible possibilities. Thus, the child might attempt to simplify the task by empirically evaluating each part of the sentence independently. In doing so, she might be expected to follow the same pattern as children in the Fay and Klahr (1996) study — she may evaluate the truth-status of a problem based only on the first piece of evidence that matches a part of the statement. For example, when presented with a statement "It will rain today OR it will not rain today", and asked if the statement is true, the child would look outside and will answer "yes" if it is raining and "no" if it is sunny (Sloutsky, Rader, & Morris, 1998).

The current research focuses on two principal limitations of children's reasoning: (1) their failure to distinguish logically determinate from empirical statements and (2) their construction of incomplete problem representations based on a very small number of possibilities compatible with a problem description. We argue that these limitations do not lead to random responding — children often exhibit consistent response patterns. One such pattern is a "cut" (Morris, 1998; Sloutsky, Rader, & Morris, 1998), which allows young children to map a tautology or a contradiction onto a state of affairs, ignoring the fact that these statements could NOT be mapped onto a state of affairs: when a statement does not afford the mapping (e.g., "It is Tuesday and it is not Tuesday"), children selectively focus on a part of the statement, while ignoring another part. At the same time, children do not do not engage in this selective focusing when a statement affords a simple mapping (e.g., "The ball will land on Red and I will open the book"). Thus, a cut will be defined as a response in which a conclusion is based on a mapping between the evidence and the first part of the problem only. Furthermore, if some forms of compound statements elicit small proportions of cuts, whereas other forms elicit large proportions of cuts, this would suggest that cuts do not stem from children's inability to attend to both parts of compound statements.

However, while several factors limit children's ability to reason, it remains unknown how these limitations change in the course of development and how young children and preadolescents differ in their solutions of logically determinate problems. The present study was designed to examine (a) the development of the ability to distinguish logical from empirical problems and (b) solution strategies and patterns of errors for different types of problems in young children and preadolescents.

2. Method

2.1. Participants

The participants were 32 preschool children enrolled in three child care centers (average age = 4.3 years; S.D. = 9 months; 16 girls and 22 boys), 34 third grade children in three elementary school classrooms (average age = 8.4 years; S.D. = 1.2 months; 19 girls and 15 boys), and 35 sixth grade children enrolled in two middle school classrooms (average age = 11.7 years; S.D. = 1.4 months; 16 girls and 19 boys) all from Columbus, OH.

2.2. Materials

The experimental tasks consisted of a series of predictions by an imaginary character, ZZ, as to the outcome of one of two separate items a) a ball dropped in a “Tautology Machine” and b) the experimenter opening (or not opening) a book. The Tautology Machine is a 21 in. × 24 in. board with a chute at the top in which a ball dropped will fall to one of two terminating points (in this experiment labeled “Red” and “Green”). The Tautology Machine has a switch (occluded from participants) in the back of the machine that moves a lever to one of two sides directing the ball to either the red or the green side.[1] The book and tautology machine were both placed between the experimenter and the participant.

ZZ made eight predictions regarding the outcomes of the game, two predictions for each of the following syntactic forms: tautologies, contradictions, disjunctions, and conjunctions. The following predictions were presented in a random order for each participant:

- *Tautology* 1: The ball will land on Red or will NOT land on Red.
- *Tautology* 2: The ball will NOT land on Red or will land on Red.
- *Contradiction* 1: The ball will land on Red and will NOT land on Red.
- *Contradiction* 2: The ball will NOT land on Red and will land on Red.
- *Conjunction* 1: The ball will land on Red and I will open the book.
- *Conjunction* 2: The ball will land on Red and I will NOT open the book.
- *Disjunction* 1: The ball will land on Red or I will open the book.
- *Disjunction* 2: The ball will land on Red or I will NOT open the book.

---

[1] The tautology machine was demonstrated so that children could see both red and green landing sites. However, children were not told that NOT red was green (or vice-versa) since other outcomes were possible (e.g., the ball could fall out of the machine).

The forms chosen do not represent all possible combinations of affirmations and negations for two reasons. First, several previous experiments on these forms demonstrated no differences in performance due to whether negation of affirmation came first (Morris, 1998; Sloutsky, Rader, & Morris, 1998). Second, attention constraints limit the number of problems children would consider.

2.3. Procedure

In this experiment, one within-subject factor was systematically manipulated, logical form of the prediction. The experiment was conducted in a single 10–15 min session that included two phases: warm-up/instruction phase and the experimental phase. Each participant was tested individually in a quiet room by a male researcher. In the instruction phase, each child was read a set of instructions that explained the purpose of the game as evaluating the predictions of an imaginary character named “ZZ”. The instruction phase was followed by the warm-up phase, and the experimental phase.

The experimenter presented ZZ, the Tautology Machine, and a book. ZZ made predictions pertaining (1) only to the ball’s landings (tautologies and contradictions) and (2) to the ball’s landing and to whether the book will be opened or closed (conjunctions and disjunctions). The participants were asked five questions (preschoolers were asked four questions) for each of the eight predictions: (1) To repeat the prediction (encoding measure). If participants made errors in encoding they were corrected before proceeding to the next question. (2) To make an initial (*a priori*) evaluation of ZZ’s predictions (True, Not True, or Can’t Tell). Participants were then asked (3) if the ball had to be dropped (and/or the book opened) to test the initial evaluation (Yes or No), and, if and only if such empirical verification was requested, the ball was dropped through the machine and a controlled outcome was obtained (e.g., the ball landed on red and the researcher did not open the book). After that, the participants were asked to (4) make an *a posteriori* evaluation (hereafter evidence evaluation) of ZZ’s prediction (Right, Wrong, or Can’t Tell). Finally, the researcher asked the participant to (5) repeat the initial prediction (recall measure). The recall measure was introduced only with elementary school and sixth graders since this measure was added after the preschool data had been collected. Children were told that the ball was dropped and the book opened by the experimenter according to a schedule provided by ZZ before the experiment.

2.4. Warm-up tasks

Four warm-up questions were given to participants: two logically determinate problems (a Contradiction and a Tautology) and two logically indeterminate problems (a Conjunction and a Disjunction). The form was slightly different in that the warm-up questions were related to immediately perceptible objects like the color of the Tautology Machine (established after asking children what the color was) or the interviewed child’s name. For example, a child was asked to name the color of the ball ("White"). Using this information, they were asked to evaluate the following statement by ZZ, "The ball is NOT white". Participants were eliminated if they gave "Can't tell" responses to both *a priori* and evidence evaluation questions on all four trials. Six preschool participants were eliminated on the basis of the warm-up tasks leaving thirty-two total preschoolers (no elementary or middle school participants were eliminated).

3. Results

In this section we present data pertaining to solution strategies, patterns of errors, and measures of accuracy. Two types of analyses will be used to examine participants' responses: (a) component analyses, or examining aggregated responses to each of the five questions and (b) analysis of individual patterns of responses across problems.

3.1. Component analysis: distinguishing logical determinacy/indeterminacy on the basis of logical form

First, we examined the ability of young children to distinguish logically determinate from logically indeterminate problems on the basis of logical form. An *a priori* evaluation was used to assess understanding of logical form before evidence and a request for empirical verification was used to assess children's understanding of when evidence is necessary and when it is unnecessary. The two are combined to form an aggregated measure of age-related differences in understanding of logical form.

Recall that for each prediction, the participants were asked (1) if the prediction was "true", "not true" or "can't tell" before requesting evidence (*a priori* evaluation), (2) if it was necessary to drop the ball or open the book (request for empirical verification), and (3) if the prediction was "true", "not true" or "can't tell" after requesting evidence.

Participants' mean number of correct *a priori* evaluations (from 0 to 2) by age and logical form are presented in Table 1. The following responses were coded as correct: Tautology — "True"; Contradiction — "False"; Disjunction and Conjunction — "Can't Tell".

A two-way 3 (age group) × 4 (logical form) repeated measures ANOVA was performed on the data presented in Table 1 with age as a between-participant factor and logical form as a repeated measure. The total number of correct responses was summed for each participant (range 0–2) and compared. The analysis yielded a significant main effect of age, $F(2, 98) = 28.8$, $P < 0.001$, form, $F(2, 98) = 5.3$, $P < 0.002$, and the interaction between age and logical form, $F(2, 98) = 3.1$, $P < 0.005$. Tukey's HSD post-hoc tests indicated that for conjunctions and tautologies, sixth graders made more correct *a priori* evaluations than both elementary and preschool children, Tukey's HSD, all $P < 0.01$. For contradictions, sixth graders

Table 1
Mean number of correct *a priori* evaluations

Age (year)	Logical form of proposition			
	Logically determinate		Logically indeterminate	
	Tautology, *M* (S.D.)	Contradiction, *M* (S.D.)	Disjunction, *M* (S.D.)	Conjunction, *M* (S.D.)
4–5	0.32 (0.06)	0.96 (0.13)	0.76 (0.50)	0.76 (0.17)
8–9	0.52 (0.10)	0.66 (0.21)	0.95 (0.37)	0.80 (0.08)
11–12	1.58 (0.30)	1.32 (0.21)	1.30 (0.42)	1.31 (0.21)

Note: Maximum score for each type of logical problem = 2.

made more correct *a priori* evaluations than third graders, Tukey's HSD, $P < 0.01$. Sixth graders made more correct *a priori* evaluations of disjunctions than preschool children, Tukey's HSD, $P < 0.02$. Overall, the analysis points to clear differences in *a priori* evaluations for logical forms as a function of age.

We also analyzed participants' requests for empirical evidence. Initial analyses indicated that there were no significant differences in the number of requests for empirical verification between contradictions and tautologies ($P > 0.10$) and between conjunctions and disjunctions ($P > 0.10$) for any age group. Therefore, participants' responses were collapsed into two groups, responses to logically determinate problems (i.e., tautologies and contradictions) and to logically indeterminate problems (conjunctions and disjunctions). Next, children's requests for empirical verification were coded as correct (1) or incorrect (0). For tautologies and contradictions a "no" response was correct and for conjunctions and disjunctions a "yes" response was correct. The correct responses were summed by formal type (determinate or indeterminate) with a maximum score of 8. These collapsed responses were subjected to a 3 (age group) × 2 (logical determinacy) mixed-measures ANOVA with logical determinacy as a repeated measure. The analysis yielded a main effect of age, $F(2, 98) = 75.9$, $P < 0.0001$, and form, $F(1, 98) = 51$, $P < 0.0001$, and an interaction between age and form, $F(2, 98) = 44.5$, $P < 0.0001$. For logically determinate forms, sixth graders made significantly more correct evidence requests than third grade and preschool children, Tukey's HSD, all $P < 0.001$. For logically indeterminate forms, sixth graders requested empirical verification less frequently than elementary and preschool children, Tukey's HSD, all $P < 0.01$, while third graders requested empirical verification less frequently than preschool children, Tukey's HSD, $P < 0.01$. Table 2 displays the mean number of correct requests for evidence. These data demonstrate that while preschool and third grade children frequently requested empirical verification for both logically determinate and logically indeterminate forms, sixth grade children requested empirical verification less frequently for logically determinate forms than for logically indeterminate forms.

These findings indicate that the majority of preschoolers and third graders did not distinguish between logical and empirical problems, and they attempted to use

Table 2
Mean number of correct requests for empirical verification

Age (year)	Logically determinate		Logically indeterminate	
	Tautology, *M* (S.D.)	Contradiction, *M* (S.D.)	Disjunction, *M* (S.D.)	Conjunction, *M* (S.D.)
4–5	0 (0)	0 (0)	4.0 (0)	4.0 (0)
8–9	0.64 (0.21)	0.36 (0.13)	3.64 (0.38)	3.68 (0.25)
11–12	3.44 (0.42)	2.76 (0.38)	2.72 (0.42)	3.16 (0.11)

Note: Correct responses: tautology, contradiction = No; disjunction, conjunction = Yes. Each correct response was scored as 1 while each incorrect response was scored as 0. Maximum score for each type of logical problem = 4.

empirical evidence to solve logical as well as empirical problems. At the same time, the majority of sixth graders distinguished between logical and empirical problems and less than a third of these children attempted to "empirically" solve logical problems.

As stated earlier, we hypothesized that the number of possibilities compatible with a problem would predict processing accuracy. Recall that we hypothesized that participants would err when the problem is compatible with more than one or less than one empirical possibility, i.e., that tautologies, contradictions, and disjunctions would elicit more errors than conjunctions. We also expected these errors to exhibit systematicity.

One particular pattern of errors that emerged in participants' evidence evaluation responses was the tendency to focus on the first part of the problem, while ignoring the second part of the problem. We identified such a pattern as a 'cut'. To analyze this and other patterns of errors, a rigorous decision procedure was introduced into the analysis. In order to exclude "logically appropriate" responses, we predetermined the landing outcomes so that we could distinguish 'cuts' from logically appropriate responses and other types of errors (can't tell). The predictions derived from these two decision strategies are presented in Table 3. As an example, consider the Tautology "The ball will land on red or it will not land on red". If the ball landed on red, then the cut is indistinguishable from the logically appropriate response. In both cases, the child would answer that ZZ was right. However, if the ball lands on green, the 'cut' could easily be distinguished from the logically appropriate response. In the first case, the response would be that ZZ was wrong, whereas in the second case the response would be that ZZ was right. Similar reasoning was applied to other logical forms presented in Table 3.

As Table 3 demonstrates, there are three possible evidence evaluation responses: "Right", "Wrong", and "Can't Tell". Coding for cuts varied by problem type; however, a response of "Can't Tell" was always coded as an error, but not a cut. Table 4 classifies each evidence evaluation response by its correctness, and whether, if incorrect, it is a cut. It is clear that errors classified as 'cuts' are the most common error types.

Table 3
ZZ's predictions, controlled outcomes, and evidence evaluations representing cuts

Prediction	Controlled outcome	Logically appropriate response	Response representing a "cut"
Tautology 1: The ball will land on Red or will NOT land on Red	Not red	Right	Wrong
Tautology 2: The ball will NOT land on Red or will land on Red	Red	Right	Wrong
Contradiction 1: The ball will land on Red and will NOT land on Red	Red	Wrong	Right
Contradiction 2: The ball will NOT land on Red and will land on Red	Not red	Wrong	Right
Conjunction 1: The ball will land on Red and I will open the book	Red and not book	Wrong	Right
Conjunction 2: The ball will land on Red and I will NOT open the book	Red and book	Wrong	Right
Disjunction 1: The ball will land on Red or I will open the book	Not red and book	Right	Wrong
Disjunction 2: The ball will land on Red or I will NOT open the book	Not red and not book	Right	Wrong

Table 4
Number of evidence evaluation errors versus cuts

Age (year)	Tautology			Contradiction			Conjunction			Disjunction			Total
	Correct	Errot	Cut	Correct	Error	Cut	Correct	Error	Cut	Correct	Error	Cut	
4–5	21	5	38	14	3	47	38	7	19	25	8	31	256
8–9	39	6	23	31	1	36	45	5	18	29	9	30	272
11–12	61	4	5	56	6	8	65	3	2	55	7	8	280
Total	121	15	66	101	10	91	148	15	39	109	24	69	808

Note: Each cell represents the total number of responses coded as either correct, a cut or a non-cut error. Row totals reflect the total number of responses (participants/age group × two trials/problem type).

A series of McNemar and Cochran's χ^2 were conducted to examine the differences in cut levels between age groups and types of logical form. Each evaluation response was coded as either correct, a cut, or an error. The total number of responses for each type were then compared by type of form and age in a Cochran's *Q*, thus each child contributed two data points for analysis. A series of pairwise McNemar tests were conducted to determine significant differences between specific items by either logical form or age group. Six pairwise comparisons were required within each age group to compare each of the four forms. In order to control for multiple comparisons, a Bonferroni adjustment was made to bring the overall alpha level back to 0.05. Thus, only those results that are significant at the $P < 0.008$ level will be considered significant. Fig. 1 displays the percent of total evidence evaluation responses that were coded as a cut.

For preschool age children, the types of forms differed significantly in the number of responses coded as a cut, Cochran's $Q\ (3, N = 64) = 22.3$, $P < 0.0001$. McNemar's pairwise comparisons indicate that conjunctions were cut significantly less frequently than tautologies, $\chi^2\ (1, N = 64) = 14.7$, $P < 0.0001$ and contradictions, $\chi^2\ (1, N = 64) = 28$, $P < 0.0001$. Disjunctions were significantly less likely to be cut than contradictions, $\chi^2\ (1, N = 64) = 10.3$, $P < 0.005$.

For third graders, the types of forms also differed significantly in the number of responses coded as a cut, $Q\ (3, N = 68) = 21.7$, $P < 0.0001$. McNemar's pairwise comparisons indicate that conjunctions were cut significantly less frequently than contradictions, $\chi^2\ (1, N = 68) = 9.6$, $P < 0.001$. Contradictions were cut significantly more than tautologies, $\chi^2\ (1, N = 68) = 8.9$, $P < 0.005$.

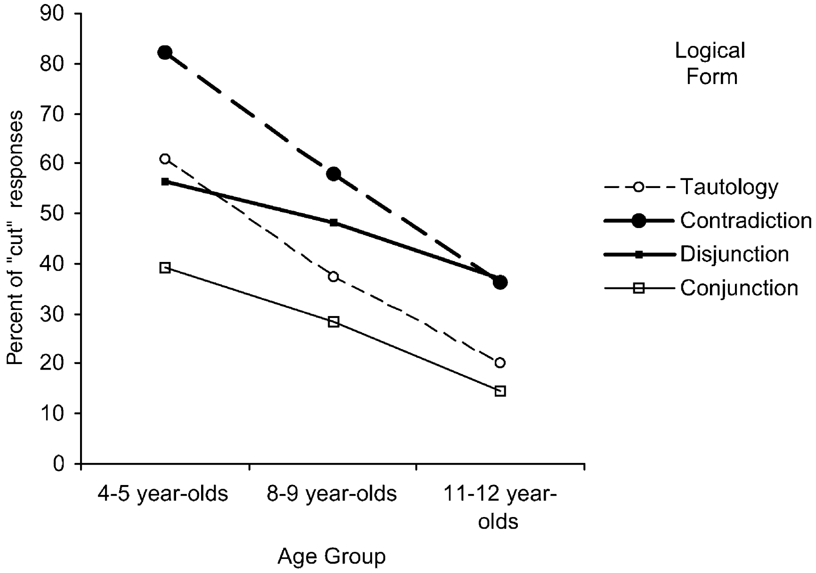

Fig. 1. Percent of "cuts" of logical propositions by age and logical form.

Sixth graders also significantly differed in cut rates across the logical forms, $Q\,(3, N = 70) = 24.6$, $P < 0.0001$. McNemar's pairwise comparisons indicate that conjunctions were significantly less frequently cut than disjunctions, $\chi^2\,(1, N = 70) = 15.9$, $P < 0.001$. Disjunctions were more frequently cut than conjunctions, $\chi^2\,(1, N = 70) = 9.4$, $P < 0.005$.

These data indicate that problems corresponding to exactly one empirical possibility (i.e., conjunctions) tended to elicit fewer cuts than problems corresponding to more than one or less than one empirical possibility. These data in Fig. 1 also indicate a marked developmental progression with respect to error rates. For sixth graders, disjunctions were cut at the highest rate while low levels of cuts of tautologies and contradictions seem to be due to recognition of logical form and therefore elimination of the need for empirical evidence and its effect on outcome.

We have contemplated several possibilities as to where in the course of information processing cuts take place. Three possibilities seem most plausible: (1) cuts occur in the course of encoding the propositions into working memory or (2) in the course of retrieving the propositions from long-term memory, or (3) they occur in the course of the creation of a problem space.

Table 5 presents the levels of encoding errors by logical form for all age groups. We are measuring correctly repeating the logical proposition as an indicator of encoding. The data demonstrate that the levels of encoding errors cannot account for cuts for two reasons: (1) encoding errors are much less frequent than cut errors and (2) the rates of encoding errors are lowest for those forms which children cut at the highest rates. Therefore, the evidence suggests that encoding is not responsible for cuts.

Overall, preschool children encoded approximately 20% of all predictions incorrectly while third and sixth graders encoded less than 10% incorrectly. The differences in encoding and cut rates and overall high rates of encoding for all forms suggest that cuts are not the product of errors in encoding. Finally, encoding errors were analyzed for systematic distortions (e.g., attending only to the first half, confusing AND and OR) that would provide possible explanations for the cut phenomenon. All errors were classified by the part (or parts) of the problem that were distorted. Each error was then grouped with errors that displayed identical distortions. The total number of errors that fit each type of classification were then divided by the total number of errors to find what percent of total errors were accounted for by a specific classification. Results indicated that no single pattern accounted for more than 5% of the errors.

Table 5
Percentage of encoding errors

Age (year)	Tautology	Contradiction	Conjunction	Disjunction
4–5	22	14	22	30
8–9	20	12	4	6
11–12	14	6	10	4

Table 6
Percentage of correctly recalled responses

Age (year)	Tautology	Contradiction	Conjunction	Disjunction
8–9	74	87	91	95
11–12	88	70	90	89

Table 6 shows recall rates for third and sixth graders. Recall rates indicate that children tend to recall the initial predictions correctly, even those with high cut rates. These results are rather surprising. For example, when given the prediction "the ball will land on red and not red" and a landing of red, a child demonstrating a cut would evaluate the statement as "true". However, when asked to recall the proposition, a large percentage recall "red and not red" as the initial prediction. These findings seem to suggest that participants represent a proposition veridically without veridically representing the corresponding state of affairs. This dissociation indicates that the reported cuts do not stem from memory limitations. It seems that a likely source of these cuts is an incomplete representation of empirical possibilities or states of affairs compatible with the proposition. Finally, like encoding errors, recall errors were analyzed for systematic distortions (e.g., repeating only to the first half, confusing AND and OR) that would provide possible explanations for the cut phenomenon. Results indicated that no single pattern accounted for more than 5% of the errors.

The preceding analyses provide data on trends derived from aggregated data that do not allow for the examination of patterns within individuals. Therefore, the next section will provide such an analysis. Individual analysis was performed to examine the consistency of response patterns within participants. Recall that each child was asked four or five questions on each of eight logical problems (32 or 40 total questions). The previous analyses were unable to provide evidence bearing on questions of processing because the data were aggregated across participants. The individual analysis will examine the consistency of individual responses across the problem set. This analysis will allow a detailed description of a child's current understanding of formal properties.

4. Individual response patterns across problems

This analysis examined overall patterns of responses across all problems. The analysis of individual patterns allows for an examination of underlying patterns of responding across the entire set of questions. The analysis of individual patterns focused on the consistency of each child's responses for (a) distinguishing logically determinate problems (LD) from logically indeterminate problems (LI) on the need for evidence and (b) coordinating the evidence given with form in order to draw correct conclusions. Coding individual patterns required (1) evalu-

Table 7
Across-problem coding decisions for evidence coordination and evidence requests

	Tautology, LD	Contradiction, LD	Conjunction, LI	Disjunction, LI
Require evidence?	No	No	Yes	Yes
Evidence evaluation	Right	Wrong	Wrong[a]	Right[a]

[a] Evidence was such that these were always correct answers.

ating raw responses as outlined in Table 7, and (2) matching coded responses to the across-problem (AP) codes outlined in Table 8. Table 7 displays the coding decisions for evaluating raw data.

The analysis of individual patterns points to four levels of underlying competence. As previously discussed, an understanding of logical form requires an understanding of two processes: (1) the mapping between evidence and form in order to draw correct conclusions and (2) an understanding of when evidence is necessary and when it is unnecessary. AP coding examines each process separately. The mapping between evidence and form was captured by the consistency with which children correctly evaluated the evidence requested. Understanding the necessity of evidence was operationalized as correctly requesting or not requesting evidence. Individual responses were coded as adhering to the pattern if 75% of the total responses conformed to the specifications detailed in Table 8. Each participant received one level code for the set of problems.

Children coded as Pattern I demonstrated a full understanding of logical form by giving correct answers to all three questions for LI and two questions for LD problems. These children distinguished LD and LI forms on the need for evidence and correctly coordinated evidence with form on LI problems, resulting in correct conclusions from the evidence given.

Children coded as Pattern II demonstrated a partial understanding of logical syntax. These children gave correct responses for LI problems. However, they requested evidence for LD problems and thus do not understand the need for evidence. Though evidence was requested, correct conclusions were drawn from the evidence. Thus, participants coded as Pattern II do not distinguish LD from LI problems on the need for evidence, yet draw correct conclusions on all problems.

Table 8
Coding decisions for individual across-problem patterns of response

Pattern	Correct evidence evaluation, LD	Correct evidence evaluation, LI	Distinguish LD from LI on need for evidence?
I	Yes	Yes	Yes
II	Yes	Yes	No
III	No	Yes	No
IV	No	No	No

Table 9
Across problem pattern use by age

Age (year)	Pattern I	Pattern II	Pattern III	Pattern IV	Other	Percentage accounted for by Patterns I–IV
4–5	0	0	3	22	7	80
8–9	1	7	16	6	4	87
11–12	22	2	6	2	3	92
Total	23	9	25	30	14	86

Note: Percentage accounted for by AP patterns does not include those coded as "Other".

Children coded as Pattern III demonstrated a limited understanding of logical form in two ways. First, a lack of syntactic understanding was demonstrated empirically by requests for evidence on LD problems as well as LI problems. In addition, Pattern III is characterized by correctly coordinating evidence with form for LI problems but not for LD problems. Children coded as Pattern IV demonstrated no understanding of logical form. These children demonstrated both an inability to distinguish logical sufficiency of LD problems and an inability to coordinate evidence on both LD and LI problems (resulting in incorrect responses on both). Children classified as Pattern IV requested evidence for all problems and drew incorrect conclusions from the evidence. Thus, for all problems there was a lack of coordination between evidence and form and a failure to distinguish LD from LI problems on the need for evidence. Given the coding guidelines for the individual pattern analysis, Table 9 displays the number of children coded as using each pattern by age.

All children ($N = 101$) were given one of five possible pattern codes (I–IV and "other"). Only those given a code ($N = 87$) were included in further analyses. A chi-square analysis indicated significant differences in the number of AP pattern codes between age groups, $\chi^2\,(6, N = 87) = 43.9$, $P < 0.001$. To determine which cells deviated significantly from the grand mean, we conducted the analysis of standardized residuals. Because standardized residuals approximate z-scores, we report them as z-score approximations. An examination of standard residuals indicated three significant differences in the likelihood of response patterns across age groups. Pattern I was significantly more likely to be found in sixth graders than in preschool and in third grade children, $z = 2.6$. Pattern III was significantly more likely to be found in third graders than in preschoolers and sixth graders, $z = 2.5$. Finally, Pattern IV was significantly more likely to be found in preschoolers than in third or sixth graders, $z = 1.9$.

5. Discussion

The findings suggest that there are age-related differences in how children solve logical and empirical problems. Preschoolers and third graders do not distinguish between logically determinate and logically indeterminate problems, often attempting to solve logical problems empirically, while many sixth graders distinguish the two. The individual analyses indicated four solution strategies that represent differences in children's ability to coordinate evidence with form and recognize the logical sufficiency of logically determinate problems. There is a dominant error pattern in all age groups: a "cut" pattern in which problems are treated as if the first part is an atomic logical statement. The number of cuts seems related to the number of empirical possibilities in that deviations from a single empirical possibility seem to increase the number of errors. Encoding and recall errors are not responsible for cuts.

The tendency to recognize logical form corresponded to reduction in cuts. Four sources of evidence support this claim. *A priori* evaluations demonstrated that preschoolers and third graders did not recognize the truth-status of problems before evidence, suggesting a limited understanding of form (Table 1). Preschoolers and third graders failed to recognize the sufficiency of logically determinate problems by requesting evidence for all (or nearly all) problems, failing to distinguish when evidence was unnecessary (Table 2). However, the same evidence suggests that sixth grade children differentiated logically determinate from indeterminate problems on *a priori* evaluations and evidence requests. The most compelling data is provided by the individual analyses. This analysis indicated that sixth graders were more likely to give consistently correct responses across the problems set than third graders or preschoolers. Thus, as in previous experiments (Fay & Klahr, 1996; Sloutsky & Morris, in review) preschool and third graders did not distinguish logically determinate from empirical statements.

The analysis of individual patterns of responses indicated that preschool and third grade children were more likely to err on component questions across all forms failing to differentiate logically determinate from logically indeterminate problems, and poorly coordinating evidence for both types of problems. There was also a tendency to make more errors on logically determinate problems (i.e., tautologies and contradictions) than on logically indeterminate problems, such as conjunctions.

Most of the errors observed exhibited a particular pattern — a tendency to cut the second half of the statement. The evidence also suggests that these cuts are not due to encoding or recall limitations. We suggest three possible explanations for this pattern of errors. First, children who "cut" complex problems may be using a simplification heuristic that helps them reduce problem complexity (i.e., indeterminacy) by limiting the number of compatible states to one. An alternative possibility is that children who "cut" complex problems simply fail to attend to the components of complex of verbal statements. The high levels of correct encoding and recall and lower levels of cuts in conjunctions eliminate the second possibility by demonstrating that children are capable of attending to both atomic components of complex of verbal statements. A third possibility is that the difficulty with complex problems may stem from the relative atypicality of tautologies, contradictions, and conjunctions. However, while tautologies and contradictions might be considered atypical, disjunctions are as comprehensible to young children as conjunctions (Johansson & Sjolin, 1975). Furthermore, atypical problems should be equally atypical for children of different age groups, whereas children exhibited marked age differences in their handling of propositions of different logical forms. Therefore, it seems likely that cuts function to reduce problem complexity.

Developmental trends suggest that the proportion of cuts decreases with the acquisition of an understanding of logical form. While preschoolers and third graders did not recognize *a priori* logical form, did not differentiate determinate from indeterminate problems in terms of empirical verification, and often failed to draw correct conclusions from evidence, sixth graders performed significantly better on all three measures. Additionally, levels of cuts were very low for sixth graders overall. Encoding and recall rates were similar for all groups and did not occur at levels high enough to account for the number of "cuts". Finally, the analysis of individual patterns indicates that children who perform cuts also err on other component questions. Taken together, these sources of evidence strongly support the notion that cuts are a part of a consistent individual response pattern. Therefore, developmental changes that seem to be related to the decrease in cuts are an increase in recognition of logical forms and a decrease in requests for empirical verification in logically determinate forms. This suggests that as children acquire an understanding of logically determinate problems, the need for limiting the number of possibilities compatible with the problem decreases. These findings support the notion that "the rule for generating correct solutions, once known, is trivially easy to execute, but inducing the rule in the first place is quite difficult" (Siegler, 1976, p. 483).

The tendency to decrease the number of possibilities (in this case a cut) indicates a tendency to decrease the problem's complexity. Perhaps this tendency is an adaptive heuristic that seeks to reduce redundancies, avoid losing information, and favor more informative constructions (Bar-Hillel, 1964; Grice, 1975). This strategy should bias the organism not only to seek more informative forms but also to seek empirical verification in reasoning. This strategy, while inappropriate in scientific thinking, is quite adaptive for the solution of practical problems. For example, Gigerenzer and Goldstein (1996) demonstrated that in many practical problems, where the problem space is very large, the consideration of only few possibilities may yield outcomes as good as systematic consideration of many or all possibilities. Other research has demonstrated a bias in young children to seek information from the environment (Bjorklund & Green, 1992). Connectionist simulations have also shown the value of limiting initial learning parameters to maximize the incoming information for the system to induce regularities from smaller samples (Elman, 1993). Perhaps an assumption that a problem is compatible with exactly one empirical possibility serves a similar function in limiting information to a level commensurate with the system's (or the child's) processing capabilities.

These findings are consonant with a position that young children's processing limitations may be adaptive for development in that they tailor experiences to existing memory and processing capacities. Incomplete problem space search might be a constraint on processing that helps one to get started, but it becomes a limitation in later ages. In particular, it may limit acquisition of mathematical and logical procedures (such as proofs, which are tautologous and totally uninformative, and yet are critically important). However, there is some evidence that the constraint could be easily overridden with appropriate training/instruction (Markman, 1990; Haygood & Bourne, 1965). For example, Haygood and Bourne (1965) demonstrated that the difficulty of learning disjunctive concepts is present only in initial training trials; differences between learning of new disjunctive and new conjunctive concepts diminished markedly after participants learned a few disjunctive concepts.

In short, three things seem to be missing in children's reasoning with logical and empirical statements: (1) veridical representations (children seem to be considering only one possibility), (2) an understanding of logical determinacy, and (3) coordination of evidence with the form of the statement. What seems to develop then are the following: (1) an emerging distinction between problems that require evidence and problems that do not, (2) better problem representation that allow for considering more possibilities, and (3) the ability to coordinate evidence and form to draw correct conclusions.

Acknowledgments

The authors would like to thank Amy Masnick, Jen Schnakenberg, Dan Osherson, Timothy Patrick Martin, David Klahr, Martha Alibali, Peter Bryant and two anonymous reviewers for their insightful suggestions. Manuscript preparation was supported by NIH Training Grant (T32 MH19102-10) to Carnegie Mellon University.

References

Bar-Hillel, Y. (1964). Semantic information and its measures. In Y. Bar-Hillel (Ed.), *Language and information* (pp. 298–312). Reading, MA: Addison-Wesley.

Bindra, D., Clarke, K., & Shultz, T. (1980). Understanding predictive relations of necessity and sufficiency in formally equivalent "causal" and "logical" problems. *Journal of Experimental Psychology: General*, *109*(4), 422–443.

Bjorklund, D. F., & Green, B. L. (1992). The adaptive nature of cognitive immaturity. *American Psychologist*, *47*(1), 46–54.

Braine, M. D. S., & O'Brien, D. P. (1998). *Mental logic*. Mahwah, NJ: Erlbaum.

Byrnes, J., & Overton, W. (1986). Reasoning about certainty and uncertainty in concrete, causal, and propositional contexts. *Developmental Psychology*, *22*(6), 793–799.

Elman, J. L. (1993). Learning and development in neural networks: The importance of starting small. *Cognition*, *48*(1), 71–99.

Evans, J. St. B. T., Newstead, S. E., & Byrne, R. M. J. (1993). *Human reasoning: The psychology of deduction*. Hillsdale, NJ: Erlbaum.

Fay, A. L., & Klahr, D. (1996). Knowing about guessing and guessing about knowing: Preschoolers' understanding of indeterminacy. *Child Development*, *67*(2), 689–716.

Gigerenzer, G., & Goldstein, D. (1996). Reasoning the fast and frugal way: Models of bounded rationality. *Psychological Review*, *103*(4), 650–669.

Grice, H. P. (1975). Logic and conversation. In P. Cole, & J. L. Morgan (Eds.), *Syntax and semantics, Vol. 3, Speech acts*. New York: Academic Press.

Haygood, R. C., & Bourne, L. E. (1965). Attribute and rule learning aspects of conceptual behavior. *Psychological Review*, *72*, 175–195.

Horobin, K., & Acredolo, C. (1989). The impact of probability judgments on reasoning about multiple possibilities. *Child Development*, *60*, 183–200.

Inhelder, B., & Piaget, J. (1958). *The growth of logical thinking from childhood to adolescence*. New York: Basic Books.

Johansson, B., & Sjolin, B. (1975). Preschool children's understanding of coordinators "And" and "Or". *Journal of Experimental Child Psychology*, *19*, 233–240.

Johnson-Laird, P. N., & Byrne, R. (1991). *Deduction*. Hillsdale, NJ: Erlbaum.

Klahr, D., & Dunbar, K. (1988). Dual space search during scientific reasoning. *Cognitive Science*, *12*, 1–55.

Kuhn, D., Garcia-Mila, M., Zohar, A., & Andersn, C. (1995). Strategies of knowledge acquisition. *Monographs of the Society for Research in Child Development*, *60*(4, Serial No. 245).

Markman, E. (1990). Constraints children place on word meanings. *Cognitive Science*, *14*(Special Issue 1), 57–77.

Markman, E. M. (1978). Empirical versus logical solutions to partŵhole comparison problems concerning classes and collections. *Child Development*, *49*(1), 168–177.

Markovits, H. (1988). Conditional reasoning, representation, and empirical evidence on a concrete task. *Quarterly Journal of Experimental Psychology*, *40*(3-A), 483–495.

Markovits, H., Fleury, M. L., Quinn, S., & Venet, M. (1998). The development of conditional reasoning and the structure of semantic memory. *Child Development*, *69*, 742–755.

Morris, A., & Sloutsky, V. (1998). Understanding of logical necessity: Developmental antecedents and cognitive consequences. *Child Development*, *69*(3), 721–741.

Morris, B. J. (1998). Maximizing informativeness: Conversions of logical propositions in preschool children. *Unpublished doctoral dissertation*, The Ohio State University, Columbus, OH.

Moshman, D., & Franks, B. (1986). Development of the concept of inferential validity. *Child Development*, *57*, 153–165.

Mynatt, C. R., Doherty, C. E., & Tweney, R. D. (1977). Confirmation bias in a simulated research environment: An experimental study of scientific inference. *Quarterly Journal of Experimental Psychology*, *29*, 85–95.

Osherson, D., & Markman, E. (1975). Language and the ability to evaluate contradictions and tautologies. *Cognition*, *3*(3), 213–226.

Piéraut-LeBonniec, G. (1980). *The development of modal reasoning: The genesis of necessity and possibility notions*. New York: Academic Press.

Rader, A., & Sloutsky, V. (in press). Processing of logically valid and logically invalid conditional inferences in discourse comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*.

Ruffman, T. (1999). Children's understanding of logical inconsistency. *Child Development*, *70*(4), 872–886.

Scholnick, E. K., & Wing, C. S. (1991). Speaking deductively: Preschoolers' use of If in conversation and in conditional inference. *Developmental Psychology*, *27*(2), 249–258.

Scholnick, E. K., & Wing, C. S. (1992). Speaking deductively: Using conversation to trace the origins of conditional thought in children. *Merrill-Palmer Quarterly*, *38*(1), 1–20.

Sharpe, D., Eakin, L., Saragovi, C., & Macnamara, J. (1996). Resolving apparent contradictions: Adults' and preschoolers' ability to cope with non-classical negation. *Journal of Child Language*, *23*, 675–691.

Siegler, R. S. (1976). Three aspects of Cognitive Development. *Cognitive Psychology*, *8*, 481–520.

Sloutsky, V. M., & Morris, B. J. (in review). Increasing informativeness and reducing indeterminacy: An adaptive constraint in human cognition.

Sloutsky, V. M., Rader, A. W., & Morris, B. J. (1998). Increasing informativeness and reducing ambiguities: Adaptive strategies in human information processing. In M. A. Gernsbacher, & S. J. Derry (Eds.), *Proceedings of the Twentieth Annual Conference of the Cognitive Science Society* (pp. 997–1002). Mahwah, NJ: Erlbaum.

Sloutsky, V. M., & Goldvarg, Y. (1999). Effects of externalization on representation of indeterminate problems. In M. Hahn, & S. Stones (Eds.), *Proceedings of the XXI Annual Conference of the Cognitive Science Society* (pp. 695–700). Mahwah, NJ: Erlbaum.

Wason, P. C., & Johnson-Laird, P. N. (1972). *Psychology of reasoning: Structure and content*. Cambridge, MA: Harvard.