

Children's Beliefs About Everyday Reasoning

Jennifer Amsterlaw

University of Washington

Two studies investigated children's metacognition about everyday reasoning, assessing how they distinguish reasoning from nonreasoning and "good" reasoning from "bad." In Study 1, 80 1st graders (6–7 years), 3rd graders (8–9 years), 5th graders (10–11 years), and adults (18+ years) evaluated scenarios where people (a) used reasoning, (b) solved problems with nonreasoning approaches, or (c) reacted appropriately but automatically to events. All age groups distinguished reasoning from type (b) nonreasoning cases, but age-related improvement occurred for type (c) cases. Study 2 tested 160 1st, 3rd, 5th graders' and adults' evaluation of good and bad reasoning processes, finding 2 developmental changes: initial improvement in discriminating thinking processes by 3rd grade, and emergence of an adult-like, process-focused (vs. outcome-focused) concept of thinking quality by 5th grade.

Children's progress toward more sophisticated understandings of the thinking mind is described by a large and varied literature that includes research on children's early "theories of mind" (Flavell & Miller, 1998; Wellman, 2002), their beliefs about the nature and sources of knowledge, or epistemology (Montgomery, 1992; Pillow, 1999), and their metacognitive knowledge about the goals, strategies, and demands of various cognitive tasks (Kuhn, 2000b). The current research focuses on children's knowledge about everyday reasoning, an important domain of metacognitive development that has not been studied previously.

Reasoning can be broadly defined as "mental activity that consists of transforming given information in order to reach conclusions" (Galotti, 1989, p. 335). It is a staple of our everyday cognition. Planning, evaluating arguments, making decisions, drawing inferences, testing hypothesis, and solving logic puzzles all involve reasoning. Understandably, a great deal of research in cognitive science has been

devoted to characterizing and explaining human reasoning processes (for reviews, see Baron, 2000; Galotti, 1989), tracking their development over the life span (e.g., DeLoache, Miller, & Pierroutsakos, 1998; Jacobs & Klaczynski, 2002; Moshman, 1998), and attempting to improve people's reasoning skills (e.g., Baron & Sternberg, 1987).

Research on children's knowledge about reasoning has the potential to inform our understanding of both theory of mind development and reasoning development. From a theory of mind perspective, more detailed information about the global organization of children's concepts of thinking processes, and about the development of their "specific theories" (Wellman, 1990) of particular mental functions, such as reasoning, is necessary to adequately describe development. As Flavell, Green, and Flavell (1995a) note, a focus on children's knowledge about mental *states* (e.g., beliefs, desires, intentions) has left important questions concerning children's knowledge about mental *processes* and *activities* (e.g., learning, reasoning, paying attention) unanswered. This is especially true for the older child's theory of mind, as developments beyond the preschool years are not as well studied. According to Wellman and Gelman (1998), theory of mind development involves both *integration* and *differentiation* of key concepts. Thus, the organization of children's concepts of mental activity can reveal important qualitative changes in their theories of mind: for example, by reflecting older children's emerging awareness of the mind's constructive and interpretive functions (Schwanenflugel, Fabricius, & Alexander, 1994). If developments observed in children's concepts of

This research served as part of the author's doctoral dissertation in psychology at the University of Michigan, Ann Arbor.

Portions of the data were presented at the biennial meeting of the Cognitive Development Society, October 2003, in Park City, UT, and at the biennial meeting of the Society for Research in Child Development, April 2005, in Atlanta, GA.

Support during research and writing was provided by NIH grant HD-22149 and NSF grant SBE-0354453.

I would like to thank the schools, parents, and children who participated in the studies, and Sandra Iler and Melanie Stevenson for assistance with data collection and coding. I am grateful to Henry Wellman and Andy Meltzoff for helpful comments on the paper.

Correspondence concerning this article should be addressed to Jennifer Amsterlaw, Institute for Learning and Brain Sciences, University of Washington, Box 357988, Seattle, WA 98195. Electronic mail may be sent to amsterj@u.washington.edu.

reasoning parallel those in other areas of theory of mind, this suggests a coherent system of conceptual change.

From the perspective of reasoning development, some researchers hypothesize a link between metacognitive knowledge and the development of reasoning skills—an idea that reflects the thrust of earlier metamemory research (for a review, see Joyner & Kurtz-Costes, 1997). Moshman (1994, 1998), for example, has argued that children's ability to reason according to the norms of rational thought depends in large part upon their skills of "metareasoning," which includes knowledge about logical principles and beliefs about appropriate reasoning strategies. Kuhn (2000a, 2000b; Kuhn, Katz, & Dean, 2004) has similarly posited a strong relationship between children's emerging metacognitive knowledge and the development of strategic reasoning during middle childhood, calling this metalevel "the locus of developmental change" Kuhn (2000a, p. 179). Although there is preliminary evidence that metacognition and reasoning are related developmentally (e.g., Kuhn & Pearsall, 1998), research directly investigating developmental changes in children's knowledge about reasoning has been sparse. Such evidence will be necessary to refine and evaluate metacognitive accounts of reasoning development.

The goal of the current research was to investigate two aspects of developmental change in children's knowledge about everyday reasoning: (1) how children distinguish reasoning from other kinds of thinking and (2) how they differentiate "good" reasoning from "bad" reasoning.

Reasoning Versus Nonreasoning

The first research question concerns children's fundamental *concept* of reasoning—that is, how they understand reasoning as a mental activity. Being able to recognize instances of reasoning and distinguish them from other kinds of thinking constitute some of the most basic knowledge about reasoning children might be said to possess. Earlier work with U.S. college-age adults (Amsterlaw, in preparation) shows that when given descriptions of people thinking in everyday situations, adults draw clear distinctions between cases that involve reasoning (e.g., weighing the pros and cons of a choice, informal hypothesis testing) and various nonreasoning cases, such as using shortcut strategies (e.g., flipping a coin), responding to events automatically (e.g., pulling one's hand from a hot stove), or engaging in purposeful thinking of some other kind (e.g., learn-

ing from a book, rehearsing a grocery list). For adults, reasoning/nonreasoning distinctions are rooted in judgments about underlying cognitive processes; reasoning cases are regarded as involving more thinking, mental effort, and logic, more use of strategies, clearer goals, and less automatic responding than nonreasoning cases. Study 1 asks: Do school-age children use these same process features to distinguish between reasoning and nonreasoning?

It may be that even young children distinguish reasoning from nonreasoning as adults do. Prior research has found that children aged 5–7 years already possess important insights about complex cognitive activities such as memory (e.g., Kreutzer, Leonard, & Flavell, 1975). Studies of children's understanding of inference, a concept closely related to reasoning, suggest that children first recognize inference as a source of knowledge at around 5 or 6 years of age (Gopnik & Graf, 1988; O'Neill & Gopnik, 1991; Sodian & Wimmer, 1987). Thus, we might expect children to have basic knowledge about reasoning by this point. However, other research indicates that key metacognitive developments occur only later in childhood. For example, developmental research on children's and adults' concepts of mental activities (e.g., Schwanenflugel et al., 1994; Schwanenflugel, Henderson, & Fabricius, 1998) finds that children as old as 8–10 years do not use inference or constructive processing as organizing dimensions in their categories of mental activities as adults do. Although children respect similarities in mental activities based on memory and attention (e.g., linking "making a list at home of all the kids in your new class" with "saying Happy Birthday on the right day to your friend"), they do not respect underlying similarities in constructive mental processes like inferring (e.g., "seeing a puddle on the ground and realizing it must have rained last night"), guessing (e.g., "filling in a question when you don't know the answer"), and reasoning (e.g., "pointing out why joining the club is better than not joining it"). More generally, in research with 5–13-year-olds, Flavell and his colleagues report late-emerging comprehension of phenomena such as the controllability of mental states (Flavell & Green, 1999; Flavell, Green, & Flavell, 1998), attentional focus (Flavell et al., 1995a), and distinctions between conscious and unconscious mentation (Flavell, Green, Flavell, & Lin, 1999). Developmental differences in children's and adults' ontologies of mental functioning may well be reflected in their concepts of reasoning, with younger children showing less differentiated categories of reasoning and nonreasoning.

“Good” Reasoning Versus “Bad” Reasoning

The second question deals with children’s ability to distinguish between good and bad reasoning: Given a description of a person’s thinking process, can children appropriately assess its quality? This question targets children’s beliefs about the norms and standards of rational thought, the kind of metacognitive knowledge Moshman (1994, 1998) describes as closely tied to reasoning development. The issue of what constitutes “good thinking” has been a contentious one (for a discussion, see Stanovich & West, 2000, and replies). In this work, I draw on distinctions given by Baron (2000), Perkins, Jay, and Tishman (1993b), and Kuhn (1996), among others, who emphasize that good thinking is typically *strategic*, *thorough*, and *unbiased*. Previous research on adults’ beliefs about thinking (Amsterlaw, in preparation; Lizotte & Amsterlaw, 2001) has found similar themes in their characterizations of good versus bad reasoning: Good reasoning involves a search for information or evidence to draw conclusions, the consideration of multiple viewpoints, being alert to biases, and high amounts of thinking and mental effort, whereas bad reasoning involves automatic and emotional responses, incomplete or biased appraisals of information, and minimal thinking, strategy, or effort. Study 2 asks whether children also use these criteria to distinguish between good and bad reasoning.

Prior research in this area has focused on children’s metacognition in formal reasoning situations, such as solving logic problems (e.g., Moshman & Franks, 1986) or testing scientific hypotheses (e.g., Kuhn & Pearsall, 1998). A small literature on children’s acquisition of “metalogue,” or explicit knowledge of logical reasoning principles, such as inferential validity and logical coherence (see, e.g., Morris, 2000; Moshman & Franks, 1986; Tunmer, Nesdale, & Pratt, 1983), indicates that the ability to meaningfully discriminate logical and illogical forms of argument emerges relatively late in childhood, around the age of 10 years. However, this research has not fully considered what children might know about good and bad thinking in everyday contexts with which they are more familiar. There appears to have been only one study directly investigating children’s awareness of flaws in everyday thinking (Perkins, Tishman, Richart, Donis, & Andrade, 2000, Study 2). In that study, sixth graders read stories that described people showing two common shortcomings in thinking—neglecting alternative options and failing to seek reasons on both sides of a case—and had to describe any problems they noticed. In an-

other condition, problems were highlighted ahead of time and children generated solutions to them. Perkins et al. found that children’s ability to detect flawed thinking was quite low; however, they were better at generating solutions. These results imply that distinguishing between good and bad reasoning can be difficult even for older children. Yet because the study did not examine developmental effects, we do not know whether younger children would have performed still more poorly, or whether adults—who may also ignore flaws in thinking (see Jacobs & Klaczynski, 2002)—would have fared any better. The current work compares the performance of children at three ages, as well as adults, on multiple judgments about thinking.

A related question is whether children can distinguish a good thinking *process* from simply getting a good *outcome*. In general, good thinking processes are expected to yield good outcomes (and bad processes, bad outcomes). Yet this is not always so. If a person’s initial information is incorrect or incomplete—not an unusual event in many real-life situations—any well-intentioned efforts to determine the best solution may fall short. Similarly, correct solutions can be obtained fortuitously, not necessarily as a result of good thinking. Adult judgments of thinking quality in such cases tend to prioritize thinking processes over final outcomes (e.g., bad thinking that results in a good outcome is still bad thinking)—although outcomes do bias ratings toward the valence of the outcome (e.g., Baron & Hershey, 1988). Outcome biases have not been investigated in young children previously (but see Klaczynski, 2001, for evidence that the bias declines from early to middle adolescence). Conceivably, young children who are naïve to process-based distinctions in reasoning quality will show extreme outcome biases in their evaluations of thinking. A second condition of Study 2 evaluates this hypothesis.

Study 1: Development of the Reasoning/ Nonreasoning Distinction

Study 1 investigated children’s ability to make process-based distinctions between reasoning and nonreasoning cases. Specifically, two kinds of nonreasoning cases were assessed. The first involves instances of *shortcut problem solving*—cases where a person uses a shortcut nonreasoning approach to solve a problem that is typically solved with reasoning (e.g., flipping a coin to make a decision). The second involves instances of *automatic responding*—cases where a person responds appropriately to

some stimulus in the environment, but does so unconsciously, without any real thinking at all (e.g., removing one's hand upon touching a hot stove). Both cases are especially clear examples of nonreasoning in adults, but they offer interesting test cases for children because they contain situational features—the presence of a problem to solve, generating an appropriate response—that are often associated with reasoning. Young children with less developed knowledge about thinking may in fact assume a reasoning response is occurring just because a person needs to solve a problem or because they respond to an event appropriately. Such a focus on salient characteristic features, in advance of appropriate attention to the defining features that reliably determine category membership, is often reported in conceptual development research (Keil, 1989).

“Reasoning” itself is not a term that children can be expected to know. Schwanenflugel et al.'s (1998) research on cognitive verb extensions suggests that third and fifth graders do not apply the verb “reason” in the same way adults do. This poses obvious challenges for studying children's ideas about the cognitive process the term references. One way to judge whether children's concepts of reasoning are like those of adults, however, is to examine whether they make similar distinctions over key features of the underlying thinking process. In this study, questions about *how much* and *how hard* people are thinking provide measures of children's sensitivity to process features that distinguish reasoning from nonreasoning, although the term “reasoning” itself is not used. To check children's understanding of the other cognitive terms that are used, a separate measure of children's cognitive word knowledge is included.

Method

Participants

A total of 60 children (20 in each of first, third, and fifth grades) and 20 adults from a small Midwestern city participated. There were 11 females and 9 males from the first grade (M age = 6.5 years, SD = 0.4); 9 females and 11 males from the third grade (M age = 8.8 years, SD = 0.6); 11 females and 9 males from the fifth grade (M age = 10.7 years, SD = 0.3); and 10 female and 10 male adults (M age = 19.3 years, SD = 2.1). Children were recruited from public elementary schools through letters to parents (\sim 45% acceptance rate). Adult participants were college students who participated to fulfill an Introductory Psychology course requirement. The final sample was approximately 65% Caucasian, 13% African American, 9% Hispanic, 5% Asian, and 8% other or unknown backgrounds. No direct measures of

socioeconomic status were obtained, but the populations were generally middle to upper-middle class.

Measures and Procedures

Children were tested individually in a quiet area of their school, receiving the reasoning/not-reasoning task followed by the cognitive word task. Sessions lasted about 20 min and were audio-taped. Adults received the tasks as a written questionnaire.

Reasoning/not-reasoning task. Participants received nine scenarios portraying everyday thinking events that participants of all ages were likely to understand. Scenario order was randomized within age groups (see the Appendix for scenarios). There were three types of scenarios: (1) *reasoning* scenarios described cases where the character faced a problem that required reasoning and used reasoning to solve it (identifying a secret admirer, figuring out why plants are wilting, deciding which bike to buy); (2) *shortcut problem-solving* scenarios described cases where the character faced a problem that typically required reasoning to solve, but he or she solved it in some other way (setting an alarm clock by trial-and-error, deciding which summer camp to go to by flipping a coin, simply recalling the answer to a math problem); and (3) *automatic action* scenarios described cases where the character responded to a stimulus in the environment appropriately but automatically, without thinking (removing hand from a hot stove, jumping away from an oncoming truck, eating some appetizing cookies).

For children, a 5" \times 8" drawing of a child's face was shown with each scenario. The researcher read the scenarios aloud and asked children to respond to a series of questions for each one: (1) Was X thinking—yes or no?; (2) How much was he or she thinking—a whole lot, a medium amount, or not really thinking at all?; (3) How hard was he or she thinking—very hard, kind of hard, or not hard at all?; (4) How long did it take him or her—a long time, a medium amount of time, or no time at all?; (5) In that story, did X have some kind of problem to figure out? Did he or she really have one, kind of have one, or not have one at all?; and (6) How smart was what he or she did—very smart, kind of smart, or not smart at all? In addition, directly after rating the amount of thinking in each scenario, participants were asked to explain their ratings (e.g., “Why did you say he was thinking a lot?”). Questions about amount of thinking and mental effort were used because these are dimensions on which adults strongly distinguish between reasoning and non-reasoning cases, and because they reference concepts

that even young children understand. Additionally, a question about length of time tested whether children might understand that reasoning typically takes longer than nonreasoning, even if they could not distinguish them by thinking processes per se. The final two questions tested whether children distinguished the cases on the basis of two salient non-process features—presence of a problem, and smartness of the response. It was expected that reasoning cases would receive high ratings on both of these dimensions. Explanations yielded additional data for interpreting developmental findings.

An 8" × 10" drawing showing three circles of increasing size was used to help children respond to questions along a 3-point scale (scored 0–2 for analyses). The researcher explained the scale by telling children to point to the small circle to say *not very much or none at all*, to the medium circle to say *a medium amount*, and to the large circle to say *a whole lot*. For the first few trials (and then on any subsequent trials where children did not readily respond to questions), the researcher pointed to and labeled each circle with its corresponding response option. To control for possible order effects, the options were given in reverse order for half the children in each group. Two warm-up items trained children how to use the scale. The researcher first asked children to imagine that she had eaten 100 hamburgers for lunch yesterday, asking "How hungry was I—very hungry, kind of hungry, or not hungry at all?" and encouraged them to use the scale to respond. She then had them imagine that she had eaten only one piece of lettuce, again asking how hungry she was. None of the children had difficulty with the warm-up items.

Cognitive word task. This task, based on Astington and Olson (1990), tested children's knowledge of words describing various cognitive activities, both to ensure comprehension of key terms used in the reasoning/not-reasoning task and to provide a more global measure of metacognitive knowledge. Children were shown a drawing of a girl, "Sarah," and were asked to hear about "some different ways Sarah can use her mind" and to say "which way you think she is using it." There were 12 multiple-choice items that the researcher read aloud to the child, each consisting of a brief description of a cognitive activity followed by three terms (one target term and two distractors) from which the child could select (see the Appendix for items). The task began with an easy warm-up item (*listen*) to introduce the format. In the few cases where children failed the warm-up item, it was repeated. No child needed it repeated more than once.

On the basis of research documenting children's acquisition of cognitive words (see Astington & Olson, 1990; Booth & Hall, 1994, 1995), the 12 items reflected a range of difficulty levels—some easy for even the youngest children (e.g., *remember*), some difficult for even the oldest children (e.g., *interpret*), and the rest intermediate. Several items (*remember*, *compare*, *wonder*, and *decide*) were used in the reasoning/not-reasoning task itself. Distractor terms were chosen so that one term was somewhat similar to the target term and the other was quite dissimilar (e.g., distractor terms for *conclude* were *predict* and *wish*). The position of the target term in the choice set was counterbalanced across items. Item order was randomized within age groups.

Explanation coding. Participants' explanations for their amount of thinking ratings in the reasoning/not-reasoning task were coded for whether they cited target features of the characters' thinking process or other aspects of the scenario. For reasoning scenarios, responses were credited with citing target thinking processes if they explicitly mentioned features of the cognitive process that was used (e.g., "because he wrote down everything and compared it"), or otherwise expressed that the person was thinking hard to solve the problem (e.g., "because she really wanted to figure it out" or "he didn't know what it was, so he really had to think"). For shortcut problem-solving cases, credit was given for responses that downplayed the characters' thinking, emphasizing that they were *just* pressing buttons, flipping a coin, or remembering the answer (e.g., "the coin was doing the thinking"; "she was just pressing random buttons"; "he was using memory, not thinking"). For the memory scenario, it also seemed appropriate to give credit for stating that remembering inherently involves thinking (e.g., "he had to think to remember"). For automatic action scenarios, credit was given for responses that mentioned instincts, automatic responding, or described characters as acting without thinking (e.g., "he just saw the cookies and went yum"). Other responses did not receive credit. Two independent raters coded the data using written transcripts that concealed participants' age, race, and gender. Inter-rater agreement on 20% of the data was 92%.

Results

Reasoning/Not-Reasoning Task

Analyses tested children's ability to make the two key contrasts between reasoning and shortcut problem solving and between reasoning and automatic

action. In each case, overall age effects were examined with a series of 4 (grade) × 2 (scenario type: reasoning or nonreasoning) ANOVAs on each of the five rating dimensions (amount of thinking, mental effort, length of time, problem, and smartness). Scenario type was a within-subjects factor. In addition, planned paired *t* tests assessed whether individual age groups successfully made the contrasts.

There were no significant effects of gender or question order (i.e., whether the response scale was presented as increasing or decreasing), and no significant interactions with grade, on task performance. Thus, all analyses collapse across gender and question order. Preliminary analyses found that ratings for one automatic action scenario, describing a boy who impulsively eats some cookies, differed significantly from ratings for the other two, which described people reacting to dangerous situations. Child (but not adult) participants rated the cookie scenario significantly lower on amount of thinking, $M = 0.23$, $SD = 0.54$, $t(59) = 9.98$, mental effort, $M = 0.28$, $SD = 0.58$, $t(59) = 8.29$, and smartness, $M = 0.23$, $SD = 0.53$, $t(59) = 18.27$, all $ps < .001$. Children's explanations revealed that many thought the boy was misbehaving, an interpretation that was unintended. This scenario was therefore excluded from subsequent analyses.

Figure 1 displays participants' dimension ratings by scenario type and grade. (Note that the maximum score on each dimension is 2.) Responses to the initial question for each scenario (asking simply whether the character was thinking or not) were redundant

with the amount of thinking ratings and were not analyzed separately.

Reasoning versus shortcut problem-solving. As shown in Figure 1, all age groups successfully distinguished between reasoning and shortcut problem solving in their ratings. Main effects for scenario type were found for amount of thinking, $F(1,76) = 244.6$; mental effort, $F(1,76) = 231.3$; length of time, $F(1,76) = 137.6$; presence of a problem, $F(1,76) = 53.2$; and smartness of response, $F(1,76) = 221.5$, all $ps < .001$, with reasoning cases receiving significantly higher ratings on all measures. Paired *t* tests confirmed that these effects held for all age groups, $ps < .05$. Significant Grade × Scenario Type interactions were found only for length of time, $F(3,76) = 4.12$, $p < .01$, and smartness judgments, $F(3,76) = 2.96$, $p < .05$. In both cases, these simply reflected an increasing divergence of ratings with age. To reiterate, even first graders significantly distinguished between reasoning and shortcut problem solving on these dimensions.

Reasoning versus automatic action. In contrast, there were clear age differences in distinguishing reasoning from automatic action. Specifically, Significant Grade × Scenario Type interactions were found for both amount of thinking, $F(3,76) = 7.77$, and mental effort, $F(3,76) = 9.97$, $ps < .001$. Paired *t* tests revealed that whereas older children and adults rated reasoning cases as involving more thinking and mental effort than automatic action cases, first graders rated them as involving similarly high amounts of thinking, $t(19) = 1.22$, and mental effort, $t(19) = 0.62$, $ps > .05$. By third grade, children appropriately rated

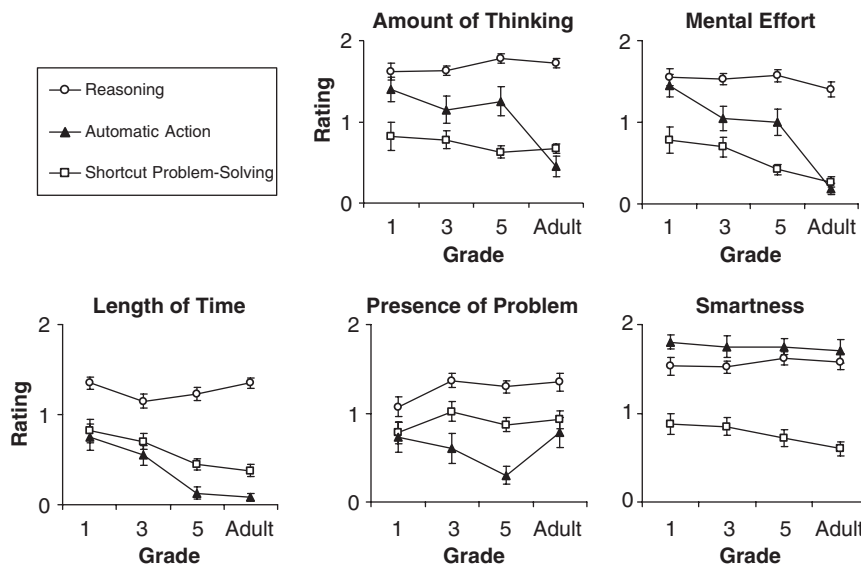


Figure 1. Dimension ratings for reasoning/not-reasoning task scenarios, Study 1. Note. Error bars are standard errors of the means.

automatic action cases lower on both amount of thinking, $t(19) = 2.86$, and mental effort, $t(19) = 3.12$, $ps < .05$.

There was a Significant Grade \times Scenario Type interaction for length of time ratings; older ages reported larger time differences between scenario types, $F(3,76) = 6.83$, $p < .001$. However, even first graders appropriately rated reasoning cases as involving more time, $t(19) = 3.37$, $p < .01$. Thus, first graders could distinguish the cases by length of time, although not by amount of thinking or mental effort, a finding consistent with the expectation that children would more readily distinguish the cases by overt behavioral features.

For problem and smartness judgments there were no significant interactions with grade. However, paired t tests indicated that although all age groups gave higher problem ratings to reasoning cases than to automatic action cases, $ts(19) > 3.05$, $ps < .01$, this difference did not reach significance for first graders, $t(19) = 1.74$. Also, while first graders showed the adult pattern of rating both reasoning and automatic action scenarios high on smartness, their ratings for automatic action cases were significantly higher than those for reasoning, $t(19) = 2.57$, $p < .05$, whereas older groups' ratings did not differ.

In sum, as expected, adults and older children characterized reasoning cases as involving extended, effortful thinking, the presence of problems to solve, and smart responses. In most instances, they rated nonreasoning cases much lower on these dimensions. (Automatic action cases were also viewed as involving smart responses, albeit ones achieved without much thinking.) Even the youngest children endorsed similar ideas about key contrasts between reasoning and nonreasoning, as they successfully distinguished between reasoning and shortcut problem-solving cases. Notably, children departed from the adult pattern not by mischaracterizing reasoning, but by overgeneralizing its cognitive features to automatic actions.

Explanations for amount of thinking judgments. Participants' explanations were coded for whether or not they referenced target thinking process features described in the scenario. Table 1 reports the percentage of process-based explanations in each age group. As shown in the table, references to target process features increased with age for all three scenario types.

For reasoning cases, most older children gave appropriate process-based explanations for their ratings, although only about half of first graders did. The rest gave various other responses, including overly vague explanations (e.g., "because he was

Table 1
Participants' References to Target Thinking Process Features in "Amount of Thinking" Explanations, Study 1

Scenario type	Grade			
	1	3	5	Adult
Reasoning	55	95**	95	100
Shortcut problem solving	55	80 [†]	95	90
Automatic action	0	35**	45	90**

Note. Values represent the percentage of participants referring to target thinking processes features in explanations for at least two out of three scenarios per type. Because one automatic action scenario was dropped from analyses, values for automatic action cases reflect the percentage of participants who referred to target thinking processes for at least one of two total scenarios. Significance was determined by $\chi^2(1, N = 40)$ against the next younger age group.

[†] $p < .10$. ** $p < .01$.

thinking a little," "he was only thinking about bikes") and references to other story events (e.g., "his parents should have been there"), with no consistent pattern to these responses. (One exception was the wilting plant story; some children said the boy was thinking a lot because he was helping his plants.) Similar results held for the shortcut problem-solving cases, where about half of the first graders reported that the characters were actively trying to solve their problems (e.g., "because she had to figure out what time" or "he couldn't decide which one to go to so he flipped a coin").

For automatic action cases, none of the first graders referred to instincts, automatic responding, or mentioned that the characters acted quickly without thinking. Such responses were more frequent among third and fifth graders, but were common only among adults. Most children instead focused on the idea that the characters' actions were smart or necessary (e.g., "if she didn't move her hand, she would get burnt"). Some even described the characters as having such thoughts before acting. Adults rarely gave high ratings to these cases, although some believed that thinking of some kind had occurred (e.g., "she had to process how close she was to the truck"). To illustrate how frequently younger children referenced the smartness of people's actions as evidence of thinking, across the two automatic action stories, roughly 95% of first graders, 68% of third graders, 60% of fifth graders, but only 8% of adults' explanations mentioned the smartness or necessity of people's actions. Thus, young children's tendency to report high amounts of thinking and mental effort in automatic action cases was closely related to their belief that these actions were smart or appropriate to the situation.

Cognitive Word Task

The cognitive word task measured children's understanding of specific cognitive words used in the reasoning/not-reasoning task (*decide*, *remember*, *wonder*, and *compare*) as well as additional words of varying difficulty. Initial analyses showed no significant main effects for gender, and no significant interactions with grade, on cognitive word task performance. As expected, performance improved with age, $F(3, 76) = 35.49$, $p < .001$. On average, first graders got 7.1 of 12 items correct ($SD = 2.5$), third graders got 9.5 items correct ($SD = 1.6$), fifth graders got 11.1 items correct ($SD = 1.0$), and adults got 11.8 items correct ($SD = 0.4$).

Table 2 shows performance on individual items by grade. Note that even first graders correctly understood most of the terms from the reasoning/not-reasoning task. (Some first graders had difficulty with *compare*, which was used in the "bike" reasoning scenario. However, children did not show lower performance on this scenario relative to others.) Nearly all third and fifth graders understood all the terms from the task, as well as additional, more difficult items. Thus, it seems unlikely that developmental differences on the reasoning/not-reasoning task were due to younger children's difficulty with any of the specific cognitive terms used.

For children, better performance on the cognitive word task was correlated with greater differentiation of reasoning and nonreasoning on the reasoning/not-reasoning task (where amount of differentiation

was calculated by subtracting children's mean "amount of thinking" ratings for nonreasoning cases from their mean ratings for reasoning cases), $r(60) = .34$, $p < .01$. Age was also correlated with performance on both the cognitive word task $r(60) = .70$, and the reasoning/not-reasoning task, $r(60) = .33$. Linear regression results indicated that age alone accounted for roughly 10% of the variance in reasoning/not-reasoning task performance, $F(1, 58) = 7.13$, $p < .01$, and cognitive word task performance did not significantly increase the model fit, $F(1, 57) = 1.60$, $p > .05$. Thus, development of the reasoning/nonreasoning distinction was generally associated with metacognitive developments assessed by the cognitive word task; however, performance on the reasoning/not-reasoning task was not directly attributable to differences in children's metacognitive word knowledge, after controlling for age.

Discussion

Study 1 found developmental differences in children's ability to distinguish reasoning from nonreasoning mental activities. Specifically, first graders did not respect a basic reasoning/nonreasoning distinction that is critical to adults—that reasoning (i.e., deliberate, effortful thinking) does not occur during quick, automatic responding. Not until third grade did children begin to recognize this. Evidence from children's explanations suggested that a primary issue was younger children's confusion of appropriate responding with extended, effortful thinking: young children cited the smartness of characters' actions to justify statements that characters were thinking a lot. (Also suggestive is the fact that the excluded cookie story, which children viewed as a case of misbehavior, and therefore *not* smart, was the only automatic action case rated low on thinking and mental effort.) These results indicate qualitative differences in children's understanding of when thinking is reasoning-like.

Children did successfully distinguish reasoning from shortcut problem solving, and did not assume that people were reasoning whenever they faced problems to solve. Thus, even the first graders acknowledged some basic distinctions between reasoning and nonreasoning. Previously, Flavell, Green, and Flavell (1993, 1995b) found that first-grade children readily attribute ongoing thinking to people presented with problems to solve (although they are less likely to do so when people lack an obvious target for cognition). The current results further demonstrate that children are aware of important thinking process features that underlie problem

Table 2
Performance on Cognitive Word Task Items by Grade, Study 1

Item	Grade			Adult
	1	3	5	
Decide	90*	100*	100*	100*
Figure out	80*	90*	95*	95*
Guess	80*	95*	95*	100*
Remember	75*	95*	100*	100*
Discover	75*	95*	100*	100*
Wonder	70*	85*	100*	95*
Understand	65*	95*	90*	100*
Compare	60*	85*	100*	100*
Get a hunch	40	80*	100*	100*
Predict	25	75*	90*	90*
Conclude	30	15	65*	100*
Interpret	15	45	70*	100*

Note. Values represent percentages of correct responses by grade ($N = 20$). Percentages marked with an asterisk are significantly above the chance expectation of 33% according to one-tailed binomial test.

* $p < .05$.

solving (e.g., mental effort, length of time) and can use these features to distinguish various thinking strategies. In particular, children knew that reasoning requires more mental activity than flipping a coin, using random trial-and-error, or recalling answers from memory. These results, along with those from Flavell et al., indicate that children's knowledge about problem solving is considerably more developed than their knowledge about uncontrolled, unconscious mental functioning (e.g., stream-of-consciousness, reflexes)—perhaps because thinking during problem solving involves more conscious control and is more available to introspection.

It is also noteworthy that first graders associated more and more effortful thinking with smarter responses, as prior work has shown that young children link thinking effort with task success (e.g., Heyman, Gee, & Giles, 2003). Further work is needed to clarify the connections young children are making between extended thinking, smart responses, and "good" behavior. For example, in Study 1 it is unclear whether children viewed extended thinking and appropriate behavior as *causally* related or as merely associated by moral valence. In addition, a more stringent test of children's understanding might have controlled for the effects of story content; in the current study, specific stories and the problems they described covaried with scenario type. This raises the possibility that children were responding to story elements other than the target thinking process features. However, including three different stories for each scenario type helps to allay concerns about the effects any one particular story might have had on the final results.

The present age effects are generally consistent with accounts of later theory of mind development given by Flavell et al. (1999) and by Schwanenflugel et al. (1994, 1998), which describe increasing differentiation in children's ontologies of conscious/unconscious and constructive/nonconstructive mental processing during middle childhood. That task success was associated with a measure of metacognitive knowledge also suggests that the reasoning/non-reasoning distinction is connected to broader developments in theory of mind, although it must be noted that the effects of metacognitive knowledge in the current study were not separable from those of age.

Study 2: Development of Beliefs about "Good" and "Bad" Reasoning

Study 2 assessed developmental changes in children's ability to distinguish good reasoning pro-

cesses from bad reasoning processes. Performance was assessed under two conditions: (1) when scenarios included only information about thinking process and no information about outcome and (2) when scenarios included information about both process and outcome, but outcome valence was inconsistent with process quality. The first condition tests whether children are sensitive to basic quality distinctions in thinking processes at all, whereas the second offers a stronger test of the belief (endorsed by adults) that the features of a person's thought process, more than the final outcome, determine whether it is judged "good" or "bad."

Method

Participants

A total of 120 children (40 in each of first, third, and fifth grades) and 40 adults from a small Midwestern city participated. Children were public school students recruited through letters to parents (~42% acceptance rate). Adults were college students who participated to fulfill an Introductory Psychology course requirement. Mean ages for first, third, fifth grade, and adult groups, respectively, were 6.2 years ($SD = 0.3$), 8.3 years ($SD = 0.3$), 10.6 years ($SD = 0.3$), and 18.7 years ($SD = 0.8$). There were 10 males and 10 females per age group and condition. The sample was 68% Caucasian, 14% Asian, 8% African American, 3% Hispanic, and 7% unknown or other backgrounds. No direct measures of socioeconomic status were obtained, but the populations were generally middle to upper-middle class.

Measures and Procedures

Children were tested individually in a quiet area of their school, receiving the good/bad thinking task followed by the cognitive word task. Sessions lasted 20–25 min and were audio-taped. Adults received the tasks as a written questionnaire.

Good/bad thinking task. Participants received 14 scenarios describing people thinking in everyday situations and were asked to rate whether they were "using good thinking or bad thinking" and to explain why (see the Appendix for scenarios). Scenarios were intended to portray everyday situations that all ages would understand (e.g., deciding whether or not to get a pet, figuring out who ate all the candy). Content was based on theoretical criteria for good and bad thinking (see Baron, 2000), as well as beliefs about good and bad reasoning that adults in prior

research had endorsed (Amsterlaw, in preparation; Lizotte & Amsterlaw, 2001). There were four types of scenarios: (1) *strategy* scenarios contrasted using a thoughtful decision strategy with using an arbitrary selection method (“eenie–meenie”), (2) *alternatives* scenarios contrasted considering alternative possibilities with jumping to an immediate conclusion, (3) *evidence* scenarios contrasted gathering evidence with acting on a hunch, and (4) *pros–cons* scenarios contrasted considering both pros and cons of a choice with considering only the pros. A *good version* and a *bad version* for each scenario type yielded eight focal scenarios. Versions were matched on peripheral features, such as subject matter and length, so that the critical difference was in the thinking process. (Subtle differences minimized repetitiveness; pilot testing with adults and children confirmed that these did not affect ratings.) In addition, good and bad versions of a *control scenario* were used to verify that children could follow the basic task format. The control scenario contrasted an obviously bad problem-solving strategy of not looking, not listening, and not trying with an obviously good strategy of looking, listening, and really trying.

Participants were randomly assigned to either the no outcome or the mismatch task condition. In the no outcome condition, scenarios included only information about people’s thinking processes, and did not say whether the final outcome was good or bad (i.e., whether people solved problems correctly or were happy with their choices). In the mismatch condition, a final line was added to each scenario to describe the outcome. Critically, for the eight focal scenarios, outcome valence was pitted against thinking process quality; good processes were paired with bad outcomes and bad processes were paired with good outcomes. (Control scenarios were presented with matched outcomes in the mismatch condition and with no outcomes in the no outcome condition.) In addition, four *match scenarios*, one for each scenario type, where good processes led to good outcomes and bad processes led to bad outcomes, acted as filler scenarios for the mismatch condition. Although not of direct interest, these were included to ensure that the task did not seem overly contrived to participants. (match scenarios were also presented, without outcome information, in the no outcome condition to maintain similarities across conditions.) Scenario orders were randomized within age groups; the two constraints were that scenarios of the same type did not follow each other and that the task always began with a match scenario to provide a warm-up to the task format.

For children, a 5" × 8" drawing of a child’s face was presented with each scenario. The researcher read each scenario aloud and asked the child to respond verbally to the key question, “Was X using good thinking or bad thinking?” (To control for possible order effects, the order of “good” and “bad” was counterbalanced between subjects.) After the child responded, the researcher asked the follow-up, “Was it very–very good/bad or just pretty good/bad?” Responses were later converted to a 0 (*very bad*) to 3 (*very good*) scale. Children were also asked to explain their ratings (e.g., “Why do you think it was very bad thinking?”).

Adults received the task in written form. After reading each scenario, they responded to “Was X using good thinking or bad thinking?” on a rating scale with points labeled 0 (*very bad*), 1 (*pretty bad*), 2 (*pretty good*), and 3 (*very good*). They were also asked to explain their ratings.

Cognitive word task. As before, the cognitive word task was included to test children’s comprehension of key terms used in presenting the good/bad thinking task (key terms were *decide*, *figure out*, *wonder*, *compare*, and *get a hunch*) and to provide a more global measure of metacognitive knowledge. The task was administered as in Study 1.

Explanation coding. Participants’ explanations for their ratings on the good/bad thinking task were coded for appropriate reference to the thinking process features targeted in each scenario. For example, for strategy scenarios, explanations for the good version had to mention that it was good that the character made a list, compared, or looked for what she wanted, and for the bad version that it was bad that the character used “eenie–meenie,” did not think about what he wanted, or should have done something else to decide, such as test out the items. Responses that did not clearly reference target-thinking processes (see below) were not counted as evidence of understanding. Two raters independently coded the data using transcripts that concealed participants’ age, race, and gender. Inter-rater agreement was 95%; disagreements were resolved through discussion. (One adult gave no explanations and so was excluded from the explanation analyses. All other participants had complete or near-complete data.)

For the mismatch condition, additional analyses assessed whether participants focused on thinking processes, outcomes, or both, in justifying their quality judgments. Explanations were coded into one of three categories: (1) *outcome* explanations cited only outcome information as a basis for the quality judgment (e.g., “He really liked the bike he got”);

(2) *process* explanations cited only process information, regardless of whether the process aspect described was the target process feature for that scenario type or if it was some other process feature (e.g., “She didn’t think about the possibilities that could have happened”); and (3) *mixed* explanations mentioned both outcomes and processes, seemingly without privileging one over the other (e.g., “He thought about the good things and the bad things, but he turned out not to like it”). Some explanations that mentioned both outcomes and processes were considered process-focused. These included explicit statements that thinking processes were central regardless of the final outcome (e.g., “Even if he got it right, it was still bad thinking because . . .”), and responses where an inconsistent outcome was used to infer attributes of the thinking process (e.g., “She must not have tested it thoroughly if she picked the wrong thing”). A fourth category included “I don’t know,” off-topic, or unintelligible responses, as well as responses that did not clearly refer to either outcomes or thinking processes. Uncodable responses were most frequent among first graders, totaling 41 out of 160 responses (26%), and were less frequent among third graders (13%), fifth graders (2%), and adults (1%). Again, two independent coders rated the explanations. Inter-rater agreement was 96%; disagreements were resolved through discussion.

Results

Good/Bad Thinking Task

For each condition, developmental changes in distinguishing good and bad versions of each scenario type were examined with a 4 (grade) \times 2 (version: good or bad) ANOVA, where version was a within-subjects factor. Preliminary analyses for effects of question order—whether the question “Was X using good thinking or bad thinking?” was asked in that order or with the options reversed—found no effects, and no significant interactions with grade, on task performance in either condition. There were also no main or interaction effects for gender in either condition. Thus, all analyses collapse across these variables.

No outcome condition. As expected, no significant age effect was found for performance on the control scenarios, $F(3,76) = 2.24$, $p > .05$. All age groups appropriately distinguished the good version ($M = 2.72$, out of a maximum score of 3, $SD = 0.48$) from the bad version ($M = 1.09$, $SD = 0.33$), $F(1,76) = 1416.9$, $p < .001$ (confirmed for all age groups with paired t tests, all $ps < .001$). The top

panel of Figure 2 displays participants’ mean thinking quality ratings for the focal “good process” and “bad process” scenario versions in the no outcome condition. The larger graph on the left gives the means (and standard errors) collapsing across scenario type, and the four smaller graphs on the right show results for each type separately. Analyses on the collapsed means found a Significant Grade \times Version interaction, $F(3,76) = 29.3$, $p < .001$, indicating increasing differentiation of good and bad thinking processes with age. This same effect was observed for all four scenario types when tested separately: for strategy scenarios, $F(3,76) = 32.08$; evidence scenarios, $F(3,76) = 7.16$; alternatives scenarios, $F(3,76) = 12.60$; and pros–cons scenarios, $F(3,76) = 7.71$, all $ps < .001$.

Specifically, for strategy scenarios, first graders did not differentiate between a thoughtful decision process and a chance-based one, giving them equally high ratings, $t(19) = 0.90$, $p > .05$. By third grade, children rated the bad process significantly lower than the good one, $t(19) = 7.31$, $p < .001$. Similarly, for alternatives scenarios, first graders did not differentiate between a good thinking process that considered alternatives from a bad one that involved going with one’s very first thought, $t(19) = 1.02$, $p > .05$. By third grade, however, children rated the process that involved considering alternatives as better than the one involving immediate conclusions, $t(19) = 12.25$, $p < .001$. First graders did distinguish between good and bad versions of the pros–cons scenarios, rating a process that considered both sides of the issue higher than one considering a single side, $t(19) = 3.16$, $p < .01$. They also appeared to distinguish evidence scenarios, rating a hypothesis-testing process as better than using a hunch, $t(19) = 4.29$, $p < .001$, with these tendencies becoming more pronounced with age.

Individual response patterns showed similar trends. For strategy scenarios, 25% of first graders rated the good version higher than the bad version, whereas 90% of third graders, 95% of fifth graders, and 100% of adults did so. For alternatives scenarios, 45% of first graders rated scenarios in the expected direction, whereas 90–100% of third graders, fifth graders, and adults did so. The figures were 60%, 90%, 95%, and 100% for pros–cons scenarios, and 65%, 75%, 80%, and 95% for evidence scenarios. All ages were near ceiling (95–100%) for control scenarios.

Children’s tendency to explain their thinking quality ratings by referencing target thinking process features increased with age, particularly between first and third grades (see Table 3). Considering

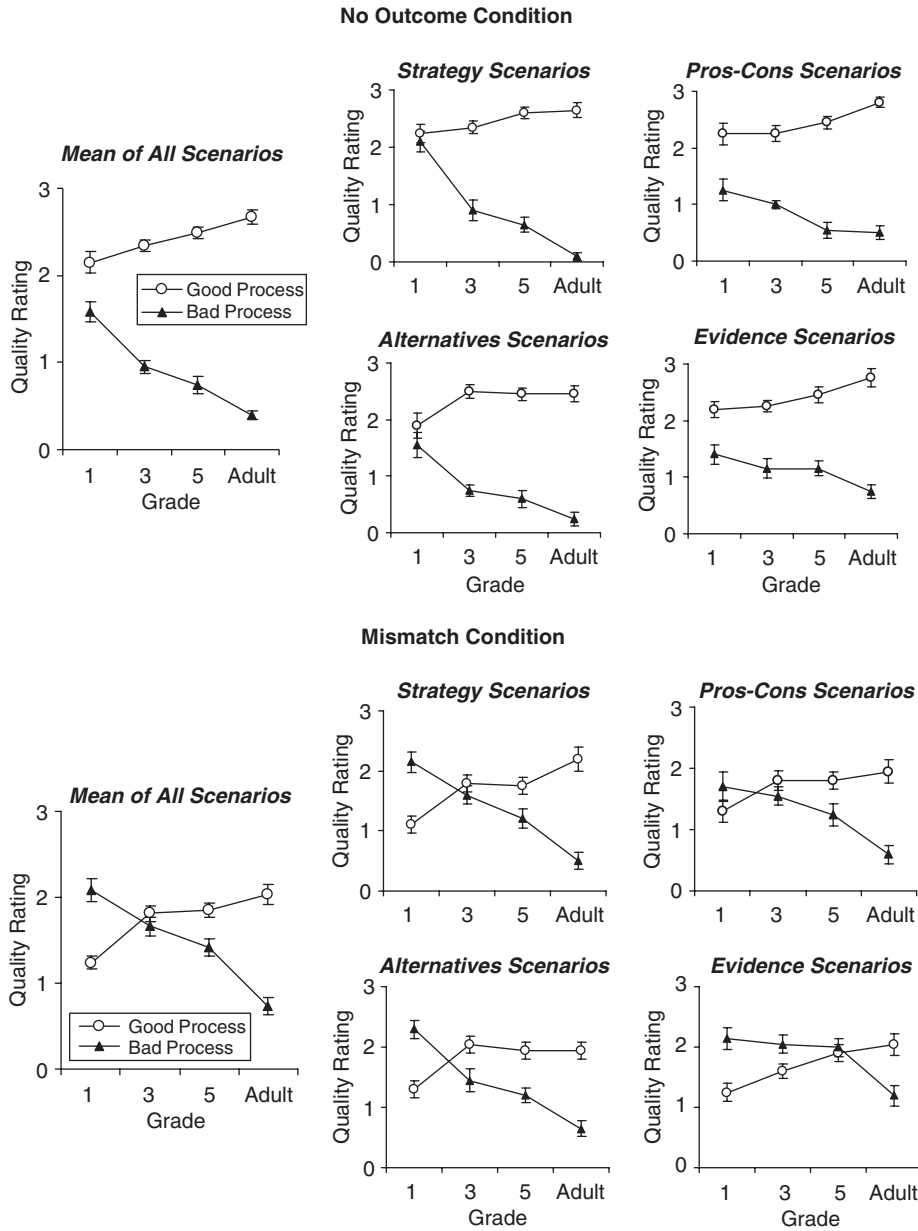


Figure 2. Thinking quality ratings for “good process” and “bad process” versions of scenarios from the good/bad thinking task, Study 2. Note. Error bars are standard errors of the means.

whether children gave appropriate process-based explanations to at least one of two versions (either good or bad) for each scenario type, only about half of first graders gave appropriate process-based explanations for ratings on strategy and alternatives scenarios, with somewhat better performance on pros-cons scenarios and poorer performance on evidence scenarios. In contrast, nearly all third graders, fifth graders, and adults appropriately referenced target thinking processes in their explanations for all four scenario types.

The explanations given by children who did not reference target thinking processes varied by scenario. For good strategy scenarios, most of these were nonspecific endorsements of the character’s thinking or the purchase itself (e.g., “because that’s a good way to do it”; “she can take pictures with a camera”), and for bad versions that eenie-meenie was a good way to decide. For pros-cons scenarios, most other responses to the good version said it was bad not to ask one’s parents about getting a pet or mentioned possible bad outcomes (e.g., pet getting

Table 3
Participants' References to Target Thinking Process Features in Thinking Quality Explanations, Study 2

Scenario type	Grade			
	1	3	5	Adult
No outcome condition				
Strategy	55	95**	100	100
Alternatives	55	100**	90	100
Pros-cons	70	100**	95	100
Evidence	15	85***	85	100 [†]
Control	100	100	100	100
Mismatch condition				
Strategy	25	75**	90	100
Alternatives	20	95***	95	100
Pros-cons	35	90***	85	90
Evidence	10	35 [†]	65 [†]	95*
Control	90	95	100	100

Note. Values represent the percentage of participants referring to target thinking process features in explanations for at least one of two scenario versions (either good or bad) per type. Significance was determined by $\chi^2(1, N = 40)$ against the next younger age group.

[†] $p < .10$. * $p < .05$. ** $p < .01$. *** $p < .001$.

sick), and for the bad version that it was good to have a pet. For alternatives scenarios, responses to both versions typically indicated agreement with the character about who had committed the crime (e.g., "because older brothers do that"). Similarly, responses to evidence scenarios indicated agreement or disagreement with the character about the source of the problem (e.g., "if a plant is drooping it doesn't need to be trimmed").

Because ratings and explanations were coded independently, children could get credit for appropriate explanations even if their ratings did not distinguish between good and bad thinking, and vice versa. However, in most cases, children who gave higher ratings to good scenario versions also gave appropriate process-based explanations (see Table 4). The only exception was for evidence scenarios, where only 3 of 13 first graders who gave correct ratings also gave process-based explanations. This confirms a general correspondence between task success and knowledge of key process features for most scenario types. (It also indicates that for evidence scenarios, first graders' rating data likely overestimated their understanding of the contrast.)

Mismatch condition. As before, there was no significant age effect for performance on the control scenarios, $F(3,76) = 1.78, p > .05$. Participants of all ages rated the good version ($M = 2.76, SD = 0.46$) higher than the bad version ($M = 0.14, SD = 0.38$),

Table 4
Performance on Thinking Quality Explanations as a Function of Performance on Thinking Quality Ratings, Study 2 (No Outcome Condition)

Scenario type	Grade			
	1	3	5	Adult
Strategy	4 (5)	18 (18)	19 (19)	19 (19)
Alternatives	8 (9)	20 (20)	18 (18)	19 (19)
Pros-cons	10 (12)	18 (18)	19 (19)	19 (19)
Evidence	3 (13)	13 (15)	15 (16)	19 (19)
Control	19 (19)	20 (20)	20 (20)	18 (18)

Note. Values represent the number of participants referring to target thinking processes in their quality explanations, out of the total number (shown in parentheses) who rated the "good-process" version higher than the "bad-process" version.

$F(1,76) = 1369.1, p < .001$ (confirmed for all age groups by paired t test, all $ps < .001$). Similarly, all ages rated good versions of the match scenarios ($M = 2.38, SD = 0.48$) higher than bad versions ($M = 1.00, SD = 0.50$), $F(1,76) = 354.5, p < .001$. There was also a significant Grade \times Version interaction, indicating increasing divergence of ratings with age, $F(3,76) = 6.50, p < .01$.

The lower panel of Figure 2 displays participants' thinking quality ratings for good and bad process versions in the mismatch condition. (Recall that for the focal scenarios in this condition, good process versions were always presented with bad outcomes and bad process versions with good outcomes.) As indicated in the figure, there was a significant Grade \times Version crossover interaction, $F(3,76) = 29.8, p < .001$, with first graders showing a reversed rating pattern from that of adults. This interaction occurred for all four scenario types when tested separately: for strategy scenarios, $F(3,76) = 22.8$; evidence scenarios, $F(3,76) = 12.4$; alternatives scenarios, $F(3,76) = 22.3$; and pros-cons scenarios, $F(3,76) = 7.56$, all $ps < .001$.

Specifically, for strategy scenarios, first graders rated the bad process version higher than the good process version, $t(19) = 4.01, p < .01$, and third graders rated them similarly, $t(19) = 0.81, p > .05$. Only fifth graders showed the adult pattern of rating the good version significantly higher than the bad version, $t(19) = 2.77, p < .05$. For alternatives scenarios, first graders also rated the bad process version higher than the good process version, $t(19) = 4.87, p < .001$. By third grade, however, children rated the good version higher, $t(19) = 2.45, p < .05$. For pros-cons scenarios, it was not until fifth grade that children rated the good version higher than the bad version, $t(19) = 2.46, p < .05$. First graders rated them

similarly, $t(19) = 1.45$, as did third graders, $t(19) = 0.93$, $ps > .05$. Finally, only adults distinguished between good and bad evidence scenarios, $t(18) = 3.92$, $p < .01$. Third graders still rated the bad version higher than the good version, $t(19) = 2.44$, $p < .05$, while fifth graders rated them similarly, $t(19) = 0.42$, $p > .05$.

Individual rating patterns reflected these same trends. Adults overwhelmingly rated good-process/bad-outcome scenarios higher than bad-process/good-outcome scenarios; 90% did so for strategy scenarios, 85% for alternatives and pros-cons scenarios, and 63% for evidence scenarios. When deviating from this pattern adults rarely rated scenarios in the opposite direction (4 of 160 total judgments), instead tending to rate scenarios equally. By contrast, first graders' dominant response was to rate bad-process/good-outcome scenarios higher, with 70% doing so for strategy scenarios, 65% for alternatives scenarios, 45% for pros-cons scenarios, and 65% for evidence scenarios. Third graders and fifth graders tended to rate the scenarios either equally or according to the adult pattern.

Although adults continued to see thinking quality as a function of underlying thinking processes even when outcomes were mismatched, outcome information did affect their ratings, as earlier outcome bias research has shown (e.g., Baron & Hershey, 1988), particularly in bad outcome cases. Relative to the no outcome condition, adults ratings for good process scenarios in the mismatch condition were 0.63 points lower on average when outcomes were bad, $t(38) = 4.30$, $p < .001$, and ratings for bad-process scenarios were 0.34 points higher on average when outcomes were good, $t(38) = 3.06$, $p < .01$. When performance across the two conditions was directly compared, all age groups, even adults, were more likely to rate good thinking processes higher than bad thinking processes in the no outcome condition. Assigning each participant a score (0–4) for the number of scenarios on which they showed the expected pattern, first graders averaged 1.95 ($SD = 1.40$) expected ratings in the no outcome condition, but only 0.40 ($SD = 0.82$) in the mismatch condition, $t(38) = 4.28$, $p < .001$. Third graders averaged 3.55 ($SD = 0.61$) expected ratings in the no outcome condition, compared with 1.40 ($SD = 1.31$) in the mismatch condition, $t(38) = 6.65$, $p < .001$, and fifth graders averaged 3.60 ($SD = 1.00$) versus 2.10 ($SD = 1.33$), $t(38) = 4.03$, $p < .001$. Adults showed the smallest margin of difference—3.95 ($SD = 0.22$) in the no outcome condition versus 3.15 ($SD = 0.80$) in the mismatch—but the difference was still significant, $t(38) = 3.22$, $p < .01$.

As in the no outcome condition, children's tendency to explain their ratings by citing target thinking processes increased with age, with children showing clear improvements between first and third grades on most scenario types (see Table 3). Additional results on the frequency of process- versus outcome-focused explanations, presented in Figure 3, showed that for all four scenario types, exclusively process-focused explanations increased significantly between first and fifth grades, $\chi^2(3, N = 160) > 30.1$, all $ps < .001$, whereas exclusively outcome-focused explanations significantly decreased in three of four cases, $\chi^2(3, N = 160) > 17.8$, all $ps < .001$. (For pros-cons scenarios, outcome-focused explanations were similarly low among all age groups, $\chi^2(3, N = 160) = 3.60$, $p > .05$.) Adults focused almost exclusively on process features in explaining their ratings for all scenario types.

Cognitive Word Task

Initial analyses showed no significant main effects for gender or condition (no outcome vs. mismatch), and no significant interactions of either variable with grade, on cognitive word task performance, all $ps > .05$. As before, performance improved with age, $F(3, 156) = 112.4$, $p < .001$. On average, first graders got 6.9 of 12 total items correct ($SD = 2.0$), third graders got 10.1 items correct ($SD = 1.1$), fifth graders got 10.8 items correct ($SD = 0.8$), and adults got 11.8 items correct ($SD = 0.4$). Performance on individual items was comparable with results from Study 1. Most children correctly understood all the cognitive terms used in the good/bad thinking task. The majority (72% or more) of first graders knew the terms *decide* and *figure out*, whereas 60% knew *wonder*, 50% knew *compare*, and 38% knew *get a hunch* (performance on all except *hunch* was significantly above the chance expectation of 33% correct, according to one-tailed binomial tests, $ps < .05$). Most (80% or more) third and fifth graders understood all the terms used in the task, as well as additional, harder items.

For children, better performance on the cognitive word task was associated with greater differentiation of good and bad thinking on the good/bad thinking task (where amount of differentiation was calculated by subtracting children's mean ratings for bad process versions from their mean ratings for good process versions) in both the no outcome, $r(60) = 0.67$, and mismatch conditions, $r(60) = 0.57$, $ps < .001$. Age was also correlated with performance on the good/bad thinking task, $r(57) = 0.58$ (no outcome) and $r(59) = 0.61$ (mismatch), as well as with

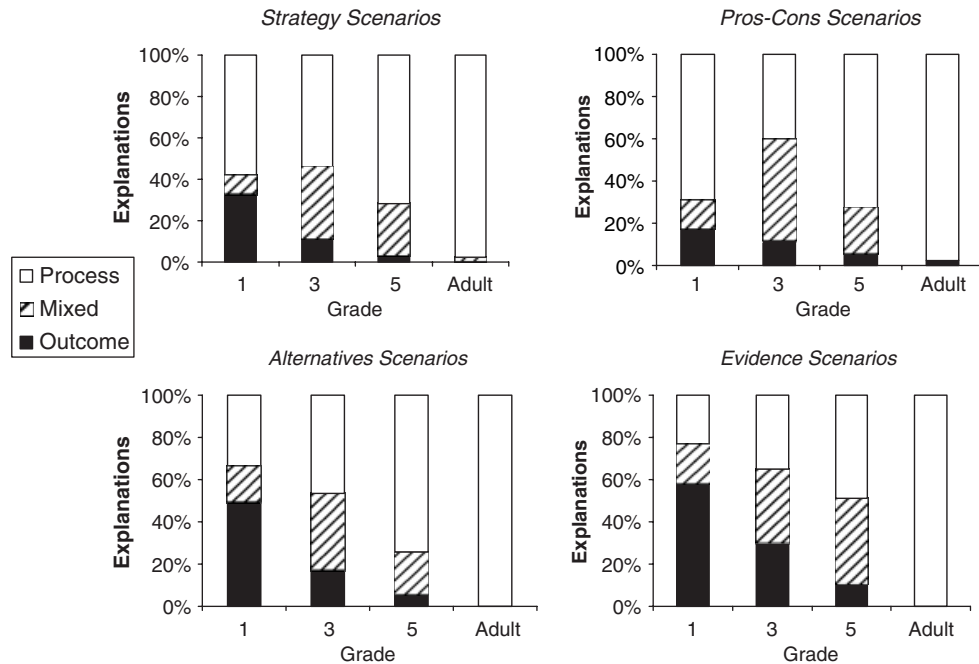


Figure 3. Frequency of outcome, process, and mixed explanations types in the mismatch condition of Study 2.

performance on the cognitive word task, $r(116) = 0.69$, $ps < .001$. Linear regression results indicated that age alone explained 33% of the variance in task performance in the no outcome condition, $F(1,55) = 27.93$, $p < .001$. Cognitive word task performance accounted for an additional 14%, $F(1,54) = 15.68$, $p < .01$, for a total of 47% variance explained, $F(2,54) = 25.53$, $p < .001$. In the mismatch condition, age accounted for 36% of the variance, $F(1,57) = 34.26$, $p < .001$. Cognitive word task performance did not add to this significantly, $F(1,56) = 4.00$, $p > .05$. Thus, for basic distinctions between good and bad thinking (no outcome condition), both being older and having better metacognitive word knowledge contributed to task success, whereas for prioritizing processes over outcomes (mismatch condition), only being older explained better performance. These results may mean that outcome bias declines with age independent of other metacognitive developments. Alternatively, more sensitive metacognitive assessments may be necessary to test the relationship.

Discussion

Study 2 observed developmental changes in children's differentiation of good and bad reasoning, both in their basic ability to discriminate good and bad processes and in their prioritization of process versus outcome information in global evaluations of

thinking quality. In the no outcome condition, first graders were sensitive to quality distinctions for pros-cons and control scenarios only. For other scenarios, it was not until third grade that children reliably distinguished good and bad thinking processes. In the mismatch condition, only fifth graders consistently showed the adult pattern of privileging process over outcome information for most cases. First graders tended to privilege outcomes over processes, whereas many third graders appeared to weight them equally. These results identify important age-related changes in children's ability to evaluate thinking processes vis-à-vis adult cultural norms of rationality.

Previously, a single study by Perkins et al. (2000) found that sixth graders have trouble detecting thinking flaws in everyday thinking, such as neglecting alternatives and failing to seek reasons on both sides of a case. The current work demonstrates that even first graders have some basic intuitions about the difference between good and bad thinking, and that third and fifth graders' knowledge about these topics is in fact fairly well developed. These two studies may have reached different conclusions because of methodological differences: Perkins et al.'s (2000, p. 276) arguably more stringent assessment used a paper-and-pencil measure where children provided written explanations that were later evaluated for elements such as "elaboration, creativity, and the presentation of a range of ideas,"

whereas this study used a verbal format, included a ratings-based assessment as well as explanations, and restricted evaluations of children's responses to conceptual content, rather than additional features.

In the current study, similar patterns of development were observed across multiple types of scenarios, which suggests the possibility of a basic developmental change in children's ability to evaluate thinking processes. One exception worth noting is that children's ability to distinguish evidence-based conclusions from plausible hunch-based conclusions developed more slowly than did other process-based distinctions. This echoes results from prior research on children's scientific reasoning skills (e.g., Kuhn, Amsel, & O'Loughlin, 1988), showing that the distinction between theory and evidence can be especially difficult for children.

Results from the cognitive word task showed that children who had more sophisticated metacognitive word knowledge were better able to distinguish between good and bad thinking. This suggests that developments in children's metacognition about reasoning are related to broader metacognitive changes, or at least to increased comprehension of mental terms. It also raises the question of whether younger children performed more poorly on the task simply because they had difficulty with specific terms. Although it is possible that this issue affected some younger children, it cannot account for the larger age effects. Some first graders did have difficulty with the terms *compare* and *get a hunch* (used in strategy and evidence scenarios, respectively); however, first graders who understood these terms were no more likely to distinguish between good and bad versions of the relevant scenarios than were children who did not, $\chi^2(1, N = 20) < 0.1, ps > .05$. Further, age effects were similar across multiple scenario types, including those (such as alternatives scenarios) that contained no problematic terms. Finally, third graders, who showed excellent comprehension of *all* terms from the good/bad thinking task, still differed from older ages in their ratings, most notably in the mismatch condition. Some deeper conceptual difference must explain these effects.

A likely reason for outcome biases in children's evaluations of thinking is simply that they lack the metacognitive knowledge necessary for making process-based judgments. (Indeed, results from the no outcome condition show that first graders were often unable to distinguish between good and bad thinking based on only process information.) If this is the case, we might expect first graders to show rating reversals in the mismatch condition only for

scenario types they failed to distinguish in the no outcome condition (strategy, evidence, alternatives), and not for those they did distinguish (pros-cons). The current study used a between-subjects design and so offers only an indirect test of this relationship; however, the data do show this pattern. Pros-cons scenarios are the only type for which first graders had no significant rating reversal.

This suggests that a main contributor to outcome bias is children's limited ability to discriminate good and bad thinking processes. That is not the whole story, though, because even when children can appropriately distinguish thinking processes, younger children still overweight outcomes in their quality judgments. (Third graders, for example, distinguished between good and bad thinking in the no outcome condition, but failed to privilege processes over outcomes in the mismatch condition.) Further research is necessary to determine the sources of this effect. One possibility is that outcome bias declines in older children and adults not only because they are more knowledgeable about specific thinking processes and strategies, but also because they are better able to distinguish between outcomes that are causally attributable to actors' strategy choices and those that are not (e.g., see Nicholls & Miller, 1985). When outcomes are inconsistent with process quality, older children and adults may discount them as due to factors beyond the actor's control (e.g., bad luck, lack of available information) while younger children hold actors equally accountable for both strategy choices and outcomes.

General Discussion

These two studies reveal meaningful developmental changes in how children conceptualize reasoning and evaluate its quality. Between the ages of about 6 years to about 10 years, children's categories of reasoning versus nonreasoning became increasingly differentiated, their ability to distinguish between good and bad reasoning processes improved, and their concepts of thinking quality became increasingly process-based.

These developments can be interpreted in the context of broader metacognitive changes taking place during middle childhood. Prior research has described the "interpretive" or "constructive" theory of mind (Carpendale & Chandler, 1996; Schwanenflugel et al., 1994) as a key developmental achievement of this period. That is, children come to recognize that subjective mental processes, such as interpretation and inference, intervene between the world and our representations of and responses to it.

This emphasis on human cognitive processes as sources of knowledge seems to entail an increasing sensitivity to a variety of process-based distinctions in children's concepts of thinking and knowing, including distinctions between states of consciousness and unconsciousness (Flavell et al., 1999), constructive and nonconstructive cognitive processing (Schwanenflugel et al., 1998), and sources of more and less certain knowledge (Pillow, Hill, Boyce, & Stein, 2000). The current results indicate that these changes are equally reflected in children's developing concepts of reasoning, as they learn to distinguish between reasoning and other kinds of thinking and develop ideas about the effectiveness of various reasoning practices.

Some theories posit that increasing metalevel knowledge about thinking contributes to developmental change in reasoning (Kuhn, 2000b; Moshman, 1998). Applied to the present findings, children's use of appropriate reasoning practices, such as gathering evidence to inform conclusions, may crucially depend on their having knowledge of them in the first place. In theory, such relationships make sense. When children encounter problems, surely they must appeal to some metalevel knowledge about how to proceed. If they did not, it is unclear how they would regulate their reasoning attempts at all. (To the extent that children do try to regulate their thinking, then, we might assume they have some minimal sense of standards to apply.) Nonetheless, the relationship between metacognition and reasoning is not straightforward. First, despite obvious metacognitive improvements of the sort demonstrated here, children's reasoning performance does not always improve with age (see Jacobs & Klaczynski, 2002). Although knowledge of appropriate strategies and standards might be considered necessary for effective reasoning, it may not be at all sufficient. (Adults' failure to live up to standards of normative judgment is a case in point.) "Dispositional" factors that describe individuals' tendencies to deploy their knowledge and skills in specific situations may prove just as important as metacognitive knowledge (Perkins et al., 1993a, 1993b). Second, there is accumulating evidence that development in children's reasoning knowledge and skills is not limited to the acquisition of *adaptive* practices and strategies. Even as children are learning strategies that enable them to reason appropriately, it seems they are also acquiring heuristics and shortcuts that can lead to biased or incorrect judgments, actually causing some aspects of reasoning performance to *decline* with age (e.g., Jacobs & Potenza, 1991; Klaczynski & Narasimham, 1998; Reyna &

Ellis, 1994). This raises the possibility that children's (as well as adults') metacognitive knowledge includes maladaptive, as well as adaptive, beliefs. Developmental theories must take these factors into account.

What leads to developmental change in children's knowledge about reasoning? Two likely mechanisms are (1) children's introspective reflection about their reasoning experiences and (2) social learning in interaction with adults (Moshman, 1994). With regard to experience and reflection, children's reasoning strategies and their metastrategic knowledge have both been found to improve during extended engagement with reasoning tasks (Kuhn & Pearsall, 1998), presumably because children learn from task feedback. Young children's heightened attention to outcomes in reasoning situations may thus serve them well, by bootstrapping knowledge about which strategies reliably obtain desired outcomes. At the same time, children's earliest forays into problem-solving and decision-making occur in social collaboration with more expert individuals (e.g., parents and teachers), and much of the specific knowledge children acquire, such as beliefs about good and bad reasoning practices, is likely transmitted to them via teaching, both direct (e.g., instruction about relevant concepts and strategies) and indirect (e.g., modeling practices, providing feedback, talking about thinking). (Of course, faulty beliefs and heuristics may also be acquired this way.)

Previous research has focused almost exclusively on children's understanding of reasoning in formal contexts, such as scientific experimentation and solving logic problems. However, given young children's limited experience with such cases, beliefs about everyday informal reasoning may well be the place where their earliest and most developed knowledge exists. The current work asks new questions about children's ideas about reasoning in everyday contexts (e.g., making choices about purchases, assigning blame to others), and finds that in responding to such cases, children readily recognize the legitimate problems these situations present, are interested in questions about the status of various problem-solving approaches, and have an emerging set of important insights about these issues. Relative to prior work, these results speak more directly to the kind of metalevel knowledge children might invoke when they encounter reasoning situations in their real lives. A crucial next step in this research path will be to evaluate developmental relationships between children's metacognition about reasoning and their reasoning behavior in naturalistic settings.

References

- Amsterlaw, J. (in preparation). Naïve rationality: Folk theories of everyday reasoning.
- Astington, J. W., & Olson, D. R. (1990). Metacognitive and metalinguistic language: Learning to talk about thought. *Applied Psychology: An International Review*, *39*, 77–87.
- Baron, J. (2000). *Thinking and deciding* (3rd ed.). New York: Cambridge University Press.
- Baron, J., & Hershey, J. C. (1988). Outcome bias in decision evaluation. *Journal of Personality and Social Psychology*, *54*, 569–579.
- Baron, J. B., & Sternberg, R. J. (Eds.) (1987). *Teaching thinking skills: Theory and practice*. New York: W. H. Freeman.
- Booth, J. R., & Hall, W. S. (1994). Role of the cognitive internal state lexicon in reading comprehension. *Journal of Educational Psychology*, *86*, 413–422.
- Booth, J. R., & Hall, W. S. (1995). Development of the understanding of the polysemous meanings of the mental-state verb know. *Cognitive Development*, *10*, 529–549.
- Carpendale, J. I., & Chandler, M. J. (1996). On the distinction between false belief understanding and subscribing to an interpretive theory of mind. *Child Development*, *67*, 1686–1706.
- DeLoache, J. S., Miller, K. F., & Pierroutsakos, S. L. (1998). Reasoning and problem solving. In D. Kuhn & R. Siegler (Eds.), *Handbook of child psychology: Vol. 2. Cognition, perception, and language development* (5th ed., pp. 801–850). New York: Wiley.
- Flavell, J. H., & Green, F. L. (1999). Development of intuitions about the controllability of different mental states. *Cognitive Development*, *14*, 133–146.
- Flavell, J. H., Green, F. L., & Flavell, E. R. (1993). Children's understanding of the stream of consciousness. *Child Development*, *64*, 387–398.
- Flavell, J. H., Green, F. L., & Flavell, E. R. (1995a). The development of children's knowledge about attentional focus. *Developmental Psychology*, *31*, 706–712.
- Flavell, J. H., Green, F. L., & Flavell, E. R. (1995b). Young children's knowledge about thinking. *Monographs of the Society for Research in Child Development*, *60* (1, Serial No. 243).
- Flavell, J. H., Green, F. L., & Flavell, E. R. (1998). The mind has a mind of its own: Developing knowledge about mental uncontrollability. *Cognitive Development*, *13*, 127–138.
- Flavell, J. H., Green, F. L., Flavell, E. R., & Lin, N. T. (1999). Development of children's knowledge about unconsciousness. *Child Development*, *70*, 396–412.
- Flavell, J. H., & Miller, P. H. (1998). Social cognition. In D. Kuhn & R. S. Siegler (Eds.), *Handbook of child psychology: Vol. 2. Cognition, perception, and language development* (5th ed., pp. 851–898). New York: Wiley.
- Galotti, K. M. (1989). Approaches to studying formal and everyday reasoning. *Psychological Bulletin*, *105*, 331–351.
- Gopnik, A., & Graf, P. (1988). Knowing how you know: Young children's ability to identify and remember the sources of their beliefs. *Child Development*, *59*, 1366–1371.
- Heyman, G. D., Gee, C. L., & Giles, J. W. (2003). Preschool children's reasoning about ability. *Child Development*, *74*, 516–534.
- Jacobs, J. E., & Klaczynski, P. A. (2002). The development of judgment and decision making during childhood and adolescence. *Current Directions in Psychological Science*, *11*, 145–149.
- Jacobs, J. E., & Potenza, M. (1991). The use of judgment heuristics to make social and object decisions: A developmental perspective. *Child Development*, *62*, 166–178.
- Joyner, M. H., & Kurtz-Costes, B. (1997). Metamemory development. In N. Cowan (Ed.), *The development of memory in childhood* (pp. 275–300). Hove, England: Psychology Press.
- Keil, F. (1989). *Concepts, kinds, and cognitive development*. Cambridge, MA: MIT Press.
- Klaczynski, P. A. (2001). Analytic and heuristic processing influences on adolescent reasoning and decision-making. *Child Development*, *72*, 844–861.
- Klaczynski, P. A., & Narasimham, G. (1998). Representations as mediators of adolescent deductive reasoning. *Developmental Psychology*, *34*, 865–881.
- Kreutzer, M. A., Leonard, C., & Flavell, J. H. (1975). An interview study of children's knowledge about memory. *Monographs of the Society for Research in Child Development*, *40* (1, Serial no. 159).
- Kuhn, D. (1996). Is good thinking scientific thinking? In D. R. Olson & N. Torrance (Eds.), *Modes of thought: Explorations in culture and cognition* (pp. 261–281). New York: Cambridge University Press.
- Kuhn, D. (2000a). Metacognitive development. *Current Directions in Psychological Science*, *9*, 178–181.
- Kuhn, D. (2000b). Theory of mind, metacognition, and reasoning: A life-span perspective. In P. Mitchell (Ed.), *Children's reasoning and the mind* (pp. 301–326). Hove, England: Psychology Press.
- Kuhn, D., Amsel, E., & O'Loughlin, M. (1988). *The development of scientific reasoning skills*. Orlando, FL: Academic Press.
- Kuhn, D., Katz, J., & Dean, D. (2004). Developing reason. *Thinking & Reasoning*, *10*, 197–219.
- Kuhn, D., & Pearsall, S. (1998). Relations between metastrategic knowledge and strategic performance. *Cognitive Development*, *13*, 227–247.
- Lizotte, D., & Amsterlaw, J. (2001). *Reasoning about reasoning: A descriptive study of young adults' everyday theories of cognition*. Poster presented at meeting of the Cognitive Development Society, Virginia Beach, VA.
- Montgomery, D. E. (1992). Young children's theory of knowing: The development of a folk epistemology. *Developmental Review*, *12*, 410–430.
- Morris, A. K. (2000). Development of logical reasoning: Children's ability to verbally explain the nature of the

- distinction between logical and nonlogical forms of argument. *Developmental Psychology*, 36, 741–758.
- Moshman, D. (1994). Reasoning, metareasoning, and the promotion of rationality. In A. Demetriou & A. Efklides (Eds.), *Intelligence, mind, and reasoning: Structure and development* (pp. 135–150). Amsterdam: North-Holland/Elsevier.
- Moshman, D. (1998). Cognitive development beyond childhood. In D. Kuhn & R. S. Siegler (Eds.), *Handbook of child psychology: Vol. 2. Cognition, perception, and language development* (5th ed., pp. 947–978). New York: Wiley.
- Moshman, D., & Franks, B. (1986). Development of the concept of inferential validity. *Child Development*, 57, 153–165.
- Nicholls, J. G., & Miller, A. T. (1985). Differentiation of the concepts of luck and skill. *Developmental Psychology*, 21, 76–82.
- O'Neill, D. K., & Gopnik, A. (1991). Young children's ability to identify the sources of their beliefs. *Developmental Psychology*, 27, 390–397.
- Perkins, D. N., Jay, E., & Tishman, S. (1993a). Beyond abilities: A dispositional theory of thinking. *Merrill-Palmer Quarterly*, 39, 1–21.
- Perkins, D. N., Jay, E., & Tishman, S. (1993b). New conceptions of thinking: From ontology to education. *Educational Psychologist*, 28, 67–85.
- Perkins, D., Tishman, S., Richart, R., Donis, K., & Andrade, A. (2000). Intelligence in the wild: A dispositional view of intellectual traits. *Educational Psychology Review*, 12, 269–293.
- Pillow, B. H. (1999). Epistemological development in adolescence and adulthood: A multidimensional framework. *Genetic, Social, & General Psychology Monographs*, 125, 413–432.
- Pillow, B. H., Hill, V., Boyce, A., & Stein, C. (2000). Understanding inference as a source of knowledge: Children's ability to evaluate the certainty of deduction, perception, and guessing. *Developmental Psychology*, 36, 169–179.
- Reyna, V. F., & Ellis, S. C. (1994). Fuzzy-trace theory and framing effects in children's risky decision-making. *Psychological Science*, 5, 275–279.
- Schwanenflugel, P. J., Fabricius, W. V., & Alexander, J. (1994). Developing theories of mind: Understanding concepts and relations between mental activities. *Child Development*, 65, 1546–1563.
- Schwanenflugel, P. J., Henderson, R. L., & Fabricius, W. V. (1998). Developing organization of mental verbs and theory of mind in middle childhood: Evidence from extensions. *Developmental Psychology*, 34, 512–524.
- Sodian, B., & Wimmer, H. (1987). Children's understanding of inference as a source of knowledge. *Child Development*, 58, 424–433.
- Stanovich, K. E., & West, R. F. (2000). Individual differences in reasoning: Implications for the rationality debate? *Behavioral and Brain Sciences*, 23, 645–726.
- Tunmer, W. E., Nesdale, A. R., & Pratt, C. (1983). The development of young children's awareness of logical inconsistencies. *Journal of Experimental Child Psychology*, 36, 97–108.
- Wellman, H. M. (1990). *The child's theory of mind*. Cambridge, MA: MIT Press.
- Wellman, H. M. (2002). Understanding the psychological world: Developing a theory of mind. In U. Goswami (Ed.), *Blackwell handbook of childhood cognitive development* (pp. 167–187). Malden, MA: Blackwell.
- Wellman, H. M., & Gelman, S. A. (1998). Knowledge acquisitions in foundational domains. In D. Kuhn & R. S. Siegler (Eds.) *Handbook of child psychology: Vol. 2. Cognition, perception, and language development* (5th ed., pp. 523–573). New York: Wiley.

Appendix

Scenarios for Reasoning/Not-Reasoning Task (Study 1)

Reasoning Scenarios

On Sylvia's birthday, somebody sent her some beautiful flowers and a birthday card. But the person didn't sign their name; it just said "from your secret friend." Who could it be? Whoever sent the flowers knew that Sylvia loved daisies. Sylvia thought about which friends of hers might know that. "I was just telling Chris how much I like daisies," she said. "It's probably Chris who sent them to me."

Kyle's parents are getting him a bike for his birthday this year. So Kyle went to the bike store to pick one out. There were lots of different bikes to choose from. Kyle wrote down a list of the important things he wanted in a bike, and then he compared the bikes in the store to his list. He looked for the bike that was the closest match to the things on his list.

Jeremy has a garden. One day Jeremy saw that some plants in his garden were wilting and turning brown. He wondered what was going on. Jeremy tried changing different things, like how much water and plant food the plants got, and he kept track of what helped the plants and what did not. On the basis of that, he decided they needed more plant food.

Shortcut Problem-Solving Scenarios

Hugh was excited because this summer he got to go to summer camp. There were two different camps he could go to and Hugh's parents said he could pick the one he wanted. So Hugh got out a quarter and he said, "If it's heads it's this camp, and if it's tails it's that camp." Hugh flipped the quarter and it landed on tails. So he went to the second camp.

One night Fabio and his family were watching TV. They were watching a certain game show where the

people on the show have to answer lots of questions. One of the questions was a big math problem with lots of parts. It was pretty tricky! But Fabio had seen this same question on the show before. He just remembered that the answer was 15.

Bianca got a new alarm clock, but she didn't know how to set the time. There were all these funny buttons on the back of the clock. Bianca didn't know what they were for. Without really watching what she was doing, Bianca started pressing different buttons. She pressed a few over here, and she pressed a few over there, just pressing buttons. After that, she saw that the clock was on the right time.

Automatic Action Scenarios

Henry loves chocolate. It's his favorite flavor. One day Henry's mom made a big batch of chocolate chip cookies. She left them sitting on the kitchen counter to cool. A little while later Henry got home from school. He saw the cookies sitting on the counter. And as soon as he saw them, he walked right over and ate three of them.

Lois likes to help her mom and dad cook dinner. One night they were making spaghetti for dinner. Lois was stirring the pot of spaghetti when she accidentally put her hand down on the stove. The stove was very, very hot! As soon as she touched the hot stove, Lois jerked her hand away.

When Marcy walks to school, she has to cross a very busy street with lots of traffic. One day Marcy started to cross the street when suddenly a huge truck came speeding down the road. At first she did not see it coming, but at the last minute she looked up. The truck was almost right to her! Marcy jumped back to the side of the road.

Items for Cognitive Word Task (Study 1 and Study 2)

Listen (warm up). Let's say Sarah is on the telephone. She has the phone right up to her ear. It's her mom on the other end. Sarah's mom is telling her a story. Is Sarah *looking* at her mom, *listening* to her mom, or *touching* her mom?

Decide. Let's say Sarah just got out of school at the end of the day and now she is thinking about what to do next. She can either go to the playground with her sister or she can go to the library and read books. She can't do both, so she has to think about which one she would rather do. Is Sarah *dreaming*, *learning*, or *deciding*?

Figure out. Let's say Sarah is leaving her classroom at the end of the day when she sees somebody's notebook lying on the ground. There is no name on the

front, only the initials AJ. Whose is it? Sarah thinks of the names of all the people in her class. There is only one name with those initials—it's Andrew James. Sarah thinks it must be his notebook. Was Sarah *remembering it*, *making it up*, or *figuring it out*?

Guess. One day Sarah's dad played a game with Sarah. He put his hands behind his back so Sarah couldn't see and he said, "I have some candy in one of my hands, but I'm not going to tell you which one. If you can pick which hand it is, I'll give you the candy." Sarah picked one, but she didn't know where the candy really was. Was Sarah *guessing*, *forgetting*, or *planning*?

Remember. Let's say Sarah's mom sent her to the grocery store to get some things. She didn't give Sarah a list. She just told her what things to get. Now Sarah is at the store and she's thinking about what things her mom wanted her to get. Is Sarah *learning*, *wishing*, or *remembering*?

Discover. Let's say Sarah is riding her bike over to her friend Joe's house. It usually takes her a while to bike over to Joe's. This time, when Sarah is riding her bike, she sees a little path on one side of the road that she never saw before. It turns out to be a secret shortcut over to Joe's house. Did Sarah *interpret* the path, *imagine* the path, or *discover* the path?

Wonder. Let's say one day there was a big snowstorm. When Sarah sees all the snow, she starts asking herself some questions in her head. "Where does snow come from?" she thinks "How is it made up in the clouds?" Is Sarah *guessing*, *wondering*, or *deciding*?

Understand. Sarah is in her science class and her teacher is giving them directions about how to do the homework assignment. Sarah is following what her teacher says, and she feels that she knows exactly what to do. Is Sarah *understanding*, *wondering*, or *exploring*?

Compare. Let's say Sarah is at the zoo and she is looking at some different animals. She is looking at an elephant and a rhinoceros, and she is thinking about how they are alike and how they are different. Is Sarah *listening*, *comparing*, or *understanding*?

Get a hunch. Let's say Sarah is driving along in her car when the car starts making a funny noise. Sarah doesn't know for sure why the car is making a funny noise, but she has a feeling that it has something to do with the fan. Is Sarah *getting a hunch*, *making a promise*, or *having a dream*?

Predict. Sarah and her friend Michelle are planning to go on a picnic. They want to choose a nice day. One morning they wake up early. Michelle says, "Should we go today?" Sarah looks out the window and she says, "Yes. It will be sunny at lunchtime today." Was Sarah *pretending*, *predicting*, or *concluding*?

Conclude. Sarah woke up one morning and looked outside. She saw that the ground was all wet. There were puddles everywhere. "It must have rained last night," she said. Did Sarah *predict* that it rained, *conclude* that it rained, or *wish* that it rained?

Interpret. Sarah has a cat named Fluff. When Sarah went away on vacation for a few weeks, she left Fluff at home with one of her friends. When Sarah came back and saw Fluff for the first time, Fluff started meowing and meowing at her. "He's telling me he missed me," Sarah said. Did Sarah *interpret* the meowing, *discover* the meowing, or *calculate* the meowing?

Scenarios for Good/Bad Thinking Task (Study 2)

Note that information about outcomes (shown below in parentheses) was presented in the mismatch condition but omitted in the no outcome condition.

Control Scenarios

Good version. One day Lucy was at school in math class when her teacher gave the class a subtraction problem to work on. The teacher told the class how to start the problem and then he said, "OK, now figure out the answer and write it down in your notebook." So Lucy had to figure out the answer to the problem. Here's what Lucy did. Lucy tried really hard to figure out the answer. She really wanted to do it right! She was listening when her teacher talked to the class, she looked at the problem very carefully, and she really thought about it. So that's how Lucy came up with her answer. She looked, she listened, and she really tried to figure it out. (It turned out that Lucy got the answer right.)

Bad version. One day Shawn was at school in math class when his teacher gave the class a multiplication problem to do. The teacher told the class how to start the problem and then she said, "OK, now figure out the answer and write it down on your worksheet." So Shawn had to figure out the answer to the problem. Here's what he did. Shawn didn't even try to figure out the answer. He didn't listen when his teacher was talking. He didn't even look at the problem. Shawn just wrote something down for an answer and he didn't care if it was right. So that's what Shawn did. He didn't listen, he didn't look, and he didn't even try to figure it out. (It turned out that Shawn got the answer wrong.)

Strategy Scenarios

Good version. Christina got some money for her birthday to buy herself a camera. So she went to the

store to look for one. When she got there, there were many different kinds to choose from. So Christina had to figure out which camera to get. This is what she did. Christina wrote down a list of the important things that she wanted in a camera. Then she compared the cameras at the store to her list to see which one was the closest match. So that's how she picked which camera to get. She compared them to her list. (It turned out that Christina didn't really like the camera she got. It didn't work very well.)

Bad version. Eric got some money for his birthday to buy himself a new bike. So he went to the bike store to look for one. When he got there, there were many different kinds to choose from. So Eric had to figure out which bike to get. This is what he did. Eric went down the row of bikes and he said "eenie meenie miney moe, catch a tiger by the toe." Whichever one his finger landed on was the bike he got. So that's how he picked which bike to get. He used "eenie meenie." (It turned out that Eric really liked the bike he got. It worked really well.)

Match version. Joyce got some money for her birthday to buy herself a CD player. So she went to the store to look for one. When she got there, there were many different kinds to choose from. So Joyce had to figure out which CD player to get. This is what she did. Joyce wrote down a list of the important things that she wanted in a CD player. Then she compared the CD players at the store to her list to see which one was the closest match. So that's how she picked which CD player to get. She compared them to her list. (It turned out that Joyce really liked the CD player she got. It worked really well.)

Alternatives Scenarios

Good version. Marvin had a nice big box of chocolates. But one day, Marvin came home after school to find that somebody had eaten all of his chocolates. Oh no! Who was it? So Marvin had to figure out who had eaten his chocolates. Here's what he did. Marvin's very first thought was that his sister did it. But he did not go blame it on her right away. No, Marvin tried to think if there was any way anybody else could have eaten them. So Marvin didn't just go with his first idea. He tried to think of all the possibilities, and that's how he decided it was his sister. (It turned out that Marvin was wrong. His brother really ate all the chocolates.)

Bad version. Trisha got some special colored pencils for her art class. One day, Trisha came home to find that somebody had been using her colored pencils and now some of them were broken. Oh no! Who was it? So Trisha had to figure out who had used her pencils. Here's what she did. Trisha's very first thought was

that her brother did it. Trisha did not try to think if there was any way anybody else could have done it. She just went with her first thought, and that's how she decided it was her brother. (It turned out that she was right. Her brother had been using her pencils.)

Match version. Erin loves oatmeal for breakfast. But one morning she woke up to find that someone had eaten up the last of the oatmeal. Oh no! Who ate it? So Erin had to figure out who had eaten the oatmeal. Here is what she did. Erin's very first thought was that her little sister did it. Erin did not try to think if there was any way anybody else could have done it. She just went with her first thought, and that's how she decided it was her sister. (It turned out that Erin was wrong. Her mother really ate all the oatmeal.)

Pros–Cons Scenarios

Good version. James was sitting on his front porch one afternoon when a girl walked by with some cute little kittens in a basket. "My cat had kittens," she said to him, "and now the kittens are getting big and they are ready to be adopted. Would you like to have a pet kitten?" So James had to figure out whether or not he should get a kitten. This is what he did. James thought about all the things that would be fun about having a pet kitten. He also thought about all the hard work it would be to take care of the kitten. So James thought about both things—the good things and the bad things, and that's how he decided to get a kitten. (It turned out that James wasn't so happy he got the kitten. It was a lot of work.)

Bad version. Ruby was walking home one day when she saw a sign in front of her neighbor's house. It said "Free Puppies." She went up to the house and there were all these cute, furry little puppies playing in the yard. The man asked Ruby: "Do you want to take a puppy home with you?" So Ruby had to figure out whether or not she should get a puppy. This is what she did. Ruby thought about all the things that would be fun about having a puppy. She didn't think about any things that would be hard about taking care of a puppy. So Ruby only thought about the good things, and that's how she decided to get a puppy. (It turned out that Ruby was really happy she got the puppy. It was a lot of fun.)

Match version. Tom was walking by the pet store one day when he saw a sign that said "Lizards for Sale." Tom went into the pet store and there were some little green lizards jumping around in a big tank. The lady at the store asked Tom if he wanted to get one. So Tom had to decide whether or not he should get a lizard. This is what he did. Tom thought

about all the things that would be fun about having a pet lizard. He also thought about all the hard work it would be to take care of the lizard. So Tom thought about both things—the good things and the bad things, and that's how he decided to get a lizard. (It turned out that Tom was really happy he got the lizard. It was a lot of fun.)

Evidence Scenarios

Good version. Liza planted some tomato plants in her garden this year, but they weren't doing very well. They started wilting and turning all yellow. Liza wondered what was going on—why were her plants getting all yellow like that? She thought of a few different things it could be: Maybe they needed more sunshine, maybe they needed more water, or maybe they needed some plant food. So Liza had to figure out what was wrong with her plants. This is what she did. Liza tried changing first one thing and then the other and she kept track of what helped the plants and what didn't, and that's how she decided the plants needed more plant food. So Liza used changing things and keeping track to decide. (It turned out that she was wrong. What the plants really needed was more sun.)

Bad version. Robert has a pet fish that lives in a big fish tank. One day Robert saw some yucky green stuff growing inside the fish tank. Why was that stuff growing in there?, he wondered. He knew of a few different things it could be: Maybe the water in the tank was too warm, maybe it was getting too much sunlight, or maybe he needed to clean the filter. So Robert had to figure out why that green stuff was growing in his fish tank. This is what he did. Robert had a hunch. He just had a feeling that the tank shouldn't be getting so much sun. So that's how he decided. Robert used a hunch. (It turned out that he was right. The tank was getting too much sun.)

Match version. Adam has a lilac bush planted in his yard. One day Adam saw that the lilac bush was looking kind of funny. The leaves didn't look very healthy and the flowers were drooping. Adam wondered what was going on. He knew of a few different things it could be: Maybe the lilac bush needed more water, maybe it needed to have its branches trimmed, or maybe it needed fertilizer. So Adam had to figure out what was wrong with his lilac bush. This is what he did. Adam had a hunch. He just had a feeling that the lilac bush needed to have its branches trimmed. So that's how he decided. Adam used a hunch. (It turned out that he was wrong. What the lilac bush really needed was more fertilizer.)