THINKING & REASONING, 2004, *10* (2), 197–219

Developing reason

Deanna Kuhn, Jared B. Katz, and David Dean Jr.
Teachers College, Columbia University, NY, USA

We argue in favour of the general proposition that the nature of reasoning is best understood within a context of its origins and development. A major dimension of what develops in the years from childhood to adulthood, we propose, is increasing meta-level monitoring and management of cognition. Two domains are examined in presenting support for these claims—multivariable causal reasoning and argumentive reasoning.

Do reasoning skills develop? Putting the question this way, the answer, it seems, must be yes, because the alternatives are implausible. Sophisticated forms of reasoning, such as those required to address the celebrated four-card problem, do not emerge in full flower, nor are they transmitted from external sources in the same manner as factual information. We can debate the exact nature of the process, but few would refute the claim that it is through application and practice that reasoning skills improve. Moreover, the idea that one fully understands mature competencies only by studying their developmental origins has an impressive range of advocates, from traditional constructivists like Werner or Piaget to modern cognitive scientists (Keil, 1998).

It is in this context that we pose here the paradox that the study of adults' reasoning is conducted largely without reference to its development. The consequences range from a restriction in perspective to serious misinterpretation or factual error. One kind of error is practical, as when we assume that a real-life task is within the competence of all adults when in fact the skills necessary to complete it are only incompletely developed in many people. Developmental differences (in rate and endpoint) become the individual variation of adulthood. An example from our own earlier research is juror reasoning (Kuhn, Pennington, & Leadbeater, 1983; Kuhn, Weinstock, & Flaton, 1994).

Correspondence should be addressed to Deanna Kuhn, Teachers College, Columbia University, TC Box 119, New York, NY 10027, USA. Email: dk100@columbia.edu

http://www.tandf.co.uk/journals/pp/13546783.html DOI: 10.1080/13546780442000015

The other common error is conceptual, when we take for granted as intuitively given forms of reasoning that are in fact hard-won achievements which are years in the making. In this article, we discuss two such cases. One, in which there has been a good deal of empirical research, is multivariable causal reasoning. The other, which has been the subject of very little empirical research to date, is argumentive reasoning.

WHAT DEVELOPS?

Before delving into the specifics of either of these topics, it is well to begin with some preliminaries in the way of a model of development. We propose a general model of the sort depicted in Figure 1 (from Kuhn, 2001). As this model makes clear, more is developing than skills themselves. At a meta-level there is developing understanding of the process (meta-procedural understanding) and product (meta-declarative understanding) that is entailed in the exercise of intellectual skills. Although we do not discuss them here, we claim intellectual values to be a critical part of what develops, as they figure heavily in the disposition, as opposed to the competence, to apply skills (Kuhn, in press).

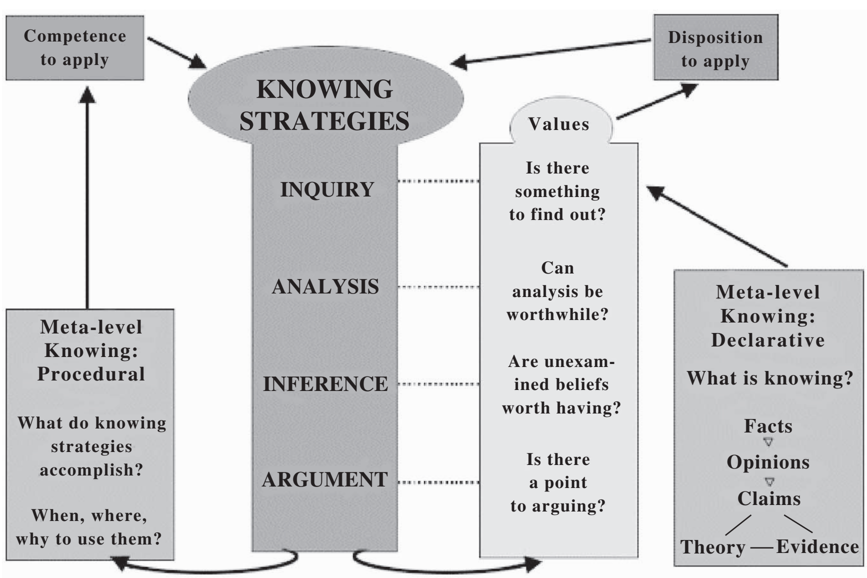


Figure 1. Knowing diagram. From Kuhn, D. (2001). How do people know? *Psychological Science*, *12*(1), 1–8.

Both inductive causal reasoning and argumentive reasoning can be characterised in the most general way as entailing the coordination of theory and evidence (Kuhn, 1989, 2001). Children from an early age construct theories as a means of understanding the world. These theories undergo revision as children interact in the world and encounter evidence bearing on their theories (Gelman & Wellman, 1998). However, in children's early years this process of theory – evidence coordination does not take place at a level of conscious awareness or control. We take gaining metacognitive control over this process to be a major dimension of cognitive development in the years from middle childhood to adolescence (Kuhn, 1989). In terms of the diagram in Figure 1, the meta-level components gradually assume a greater role in monitoring and management of procedural-level skills.

This meta-level of management assumes particular importance in light of the general finding from microgenetic research (Kuhn, Garcia-Mila, Zohar, & Andersen, 1995; Kuhn & Phelps, 1982; Siegler & Crowley, 1991) that individuals have available a range of strategies at any one time that might be brought to bear on a particular problem, implicating a meta-level operator that selects the strategy to be applied on a given occasion. This meta-level operator must also veto other available strategies as incorrect or less effective—a function that figures importantly in the dual-process models of cognitive development proposed by Klaczynski (this issue) and others.

MODELS OF CAUSAL REASONING

In this article we focus on the sample case of multivariable causal reasoning, a topic of great theoretical and practical significance and one that lies at the heart of the general topic of inductive inference. People beginning in early life and continuing throughout their life spans contemplate a wide range of phenomena that intersect in variable ways, and on the basis of this evidence draw conclusions regarding their interrelations as causes and effects. The task is a dual one of drawing valid inferences and inhibiting invalid ones. How do people do it? This question has been the subject of considerable research and theorising in the adult cognition literature. Although it is reasonable to suppose that these reasoning skills undergo development, it is the mature adult that has been the subject of theoretical models. Moreover, empirical work has relied almost entirely on college populations.

The model that received the most attention is Cheng and Novick's (1990, 1992) "probabilistic contrast" model of multivariable causal inference (MCI), later referred to by Cheng (1997) as the "causal power" theory. According to this theory, within the focal set of events regarded as theoretically relevant by the attributor, inferences of causality are based on estimated differences in the probabilities of the effect in the presence versus the absence of the potential cause. Factors yielding substantial differences across instances will be attributed as causes.

Although Cheng's causal power theory has received the lion's share of attention, several other modern theories of causal inference similarly implicate the constraining influence of theoretical belief and the computation of contrasts between conditions, and are generally consistent with the causal power theory. "Abnormal conditions" (ones absent in a comparison condition), for example, are the basis of the model of causal inference proposed by Hilton and Slugoski (1986). Similarly, models of counterfactual reasoning (Roese & Olson, 1996) rely on comparison of probabilities under two conditions (in which an event does or does not occur), while Bayesian net models (Glymour, 2001; Glymour & Cheng, 1998) also emphasise prior probabilities as constraining computations of causal power.

All of these models share a distinction emphasised by Mackie (1974) and others (Einhorn & Hogarth, 1986) between causes and enabling conditions. In Cheng and Novick's (1992) model, factors that are constant across instances will be either regarded as enabling conditions, if they are perceived as relevant, or dismissed as causally irrelevant (and hence excluded from the focal set). Note that the latter distinction rests entirely on theoretical belief. Covariation within a focal set of instances may provide the basis for a judgement of causality, but when this covariation is absent, theoretical belief offers the only basis for judging whether constant factors are causally relevant (as enabling conditions) or noncausal. This has not been an issue, since inferences of noncausality are in effect treated as non-inferences in such models and accorded little if any attention—a decision we will question here.

The conceptual advance represented by such models is in rejecting covariation alone as sufficient basis for causality and in specifying how knowledge beyond covariation serves to limit the number of covariates inferred to be causal. Theoretical knowledge of a possible causal mechanism appears necessary if a covariate is to be judged causal (Ahn, Kalish, Medin, & Gelman, 1995; Cheng, 1997; Lien & Cheng, 2000). Such knowledge admits a feature to the set considered to be causally relevant and allows the assessment of covariation between feature and outcome to be computed. The problem of causal inference might thus be seen as one of coordinating theoretical understandings of causal mechanism with empirical covariation data (Newsome, 2003; Rehder, 2003).

The goal of Cheng's current work in the MCI paradigm is a far-reaching one: to identify a set of universal inference rules that govern human (and even non-human) causal inference. In Cheng's (1997) words summarising her own effort towards this end, "... the theory I proposed presents a theoretical solution to the problem of causal induction first posed by Hume more than two and a half centuries ago. Moreover, the fact that this theory provides a simple explanation for a diverse set of phenomena regarding human reasoning and Pavlovian conditioning suggests that it *is* the solution adopted biologically by humans and perhaps other animals'' (p. 398, italics in original).

Although Cheng's own research based on her model is conducted with college students, there does exist some research following this general paradigm that is an exception to the exclusive use of college students as research subjects. A study by Harris, German, and Mills (1996) and a series of studies by Gopnik and colleagues (Gopnik & Sobel, 2000; Gopnik, Sobel, Schulz, & Glymour, 2001) have examined very young children and highlighted the respects in which their performance appears to conform to Cheng's model and thereby to reveal competent, adult-like causal reasoning.

The data we present here portrays a markedly different picture. Elsewhere (Kuhn & Dean, in press), we have examined in depth the methodological differences that might account for the picture of very early competence presented by Gopnik and Harris and their co-authors, and the picture of incompletely developed skills among older children and even adults presented here. We will not repeat that detailed analysis here, although interested readers may wish to refer to it. In brief summary, the emphasis in the research presented here is on describing and summarising individual patterns of reasoning over time. Research stemming from Cheng's MCI paradigm, in contrast, and consistent with its goal of identifying universal processes, has relied largely on more traditional quantitative methods that confine analysis to closed-ended responses and to the group rather than the individual as the unit of analysis.

Here, we begin by identifying several key attributes that one might expect to characterise a mature mental model of multivariable causality. We then examine data that bear on the extent to which the reasoning of various samples of children and adults conform to the model. Note that we are using the term ''mental model'' in a more generic manner than is customary (Kuhn, Black, Keselman, & Kaplan, 2000). Typically, the term is invoked to refer to a model of the way in which someone understands some particular phenomenon, such as electricity (Gentner & Gentner, 1983). We have argued that children similarly develop more generic mental models, such as a model of causality, that serve as means for their interpreting a wide range of phenomena and are susceptible to revision, as are more specific mental models.

The following are three key characteristics we propose as attributes of a mature mental model of multivariable causality:

1. Consistency. All else being equal, a cause that produces an effect on one occasion will produce the same effect on another occasion.

2. Additivity. More than one factor may operate on an outcome at a given time or on a given occasion. Normally, the effect of these factors combine, i.e., the total effect is the sum of all individual effects in operation.
3. Interactivity (non-additivity). In some contexts, such as necessary or sufficient or genuinely interactive causes, additivity may not apply and the total effect is not the sum of individual effects.

The mental model of multivariable causality that these attributes reflect, we claim, tends to be taken for granted as in place and operating in all individuals according to models of adult causal cognition such as Cheng's MCI model. We turn now to data that give us reason to question such an assumption.

EXAMINING ASSUMPTIONS OF THE MATURE MENTAL MODEL OF MULTIVARIABLE CAUSALITY

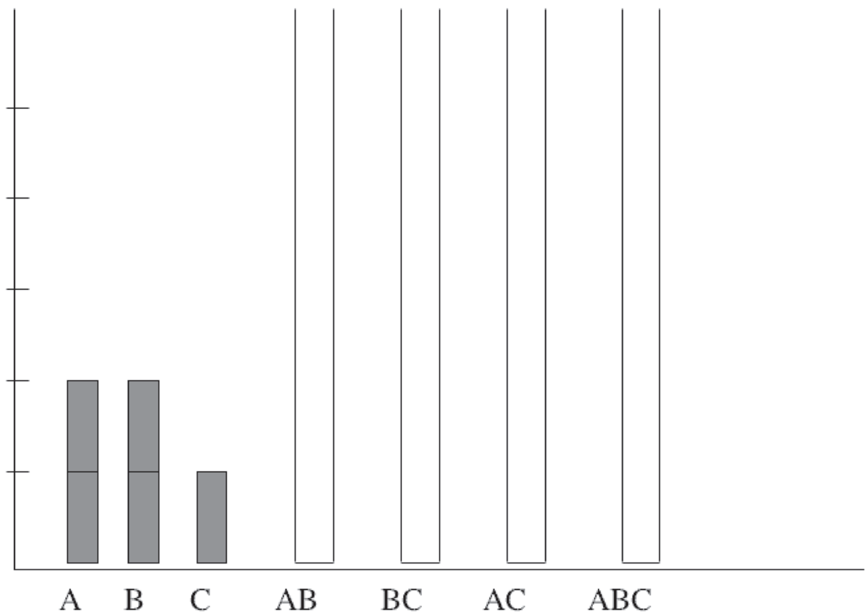
We begin with the third assumption, the distinction between additivity and interactivity, since the relevant data are the most straightforward. Consider college students' responses to the problems in Table 1 (Kuhn, unpublished). The wood-stacking problem (Table 1) is a straightforward one that we also administered to sixth graders and found them readily able to solve by simply adding the individual outputs to determine a joint output when individuals worked together. College students produced this same solution. When the content is transformed to one about chemical pollution, however, the solution becomes indeterminate:

Alt, Bot, & Crel are chemicals that pollute the air and make it dirty.
The first bar shows how much pollution Alt causes.
The second bar shows how much pollution Bot causes.
The 3rd bar shows how much pollution Crel causes.

Two pollutants together may not produce twice the level of pollution as they do individually (there may exist a ceiling on the total amount of pollution), or two together may produce more than twice the level of their individual effects (since in combination they are particularly harmful). However, no sixth graders, and only 3 of 33 college students (09%), recognised this indeterminacy. One college student predicted a particular form of interactive outcome, and the remaining 29 (88%) added the individual effects, producing a solution identical to the one given in the wood-stacking problem. These results are consistent with other findings (Dixon & Tuccillo, 2001; Wilkening, 1981) that even young children are able to predict outcomes based on the joint effects of two variables when they are asked explicitly to do so. Yet the findings

TABLE 1
Wood-stacking problem

Al, Bob, and Chris are stacking wood.
The first bar shows how much Al stacks; the 2nd bar shows how much Bob stacks; the 3rd bar shows how much Chris stacks.
Fill in the last 4 bars to show:
How much A&B together stack
How much B&C together stack
How much A&C together stack
How much A, B, & C together stack.



described here indicate that even adults tend not to differentiate additive and interactive causes when thinking about multiple factors affecting an outcome.

This lack of differentiation is less surprising when we note that there exist no natural language equivalents to distinguish the two cases. If we say for example, "Get a good night's sleep and eat a good breakfast and you'll do really well," we are neither required nor encouraged to distinguish between an additive and interactive model as the one we have in mind as applying in this situation.

At the same time, this example prompts us to ask whether it is really worth worrying very much about such a distinction—one that even graduate students in statistics are known to struggle with. Can it make all that much difference in everyday practical reasoning? As long as the individual causal agents are taken into account and integrated in some fashion, the result is arguably a good enough approximation to suffice.

We thus turn to the simpler case of multiple variables that act independently on an outcome and are additive in their effects. Keselman (2003) asked sixth graders to investigate and make inferences regarding the causal role of five variables that had been identified within a domain (variables affecting earthquake risk), as well as asking them to make outcome predictions for two new cases representing unique combinations of levels of variables within the domain. After each prediction, the question was asked, "Why did you predict this level of risk?". Three of the five variables had additive effects on the outcome and the remaining two had no influence. Those variables a student mentioned in his or her response were regarded as ones for which he or she had made an *implicit* judgement that the variable was causal. The variables students named earlier as causal (in announcing their post-investigation conclusions) were taken as *explicit* causal judgements.

Consistency between explicit and implicit causal judgements was low. Over half of the students justified one or both of their predictions by implicating a variable they had earlier explicitly concluded to be noncausal. More than 80% failed to implicate as contributing to the outcome one or more variables they had previously explicitly claimed to be causal. Overall, fewer features were implicated as contributory in the implicit attributions than were explicitly stated to be causal. Students showed low consistency not only between explicit and implicit causal theories, but also in the consistency of causal attribution across the predictions. Almost three-quarters of students failed to implicate the same variable(s) as having causal power across both prediction instances. Finally, roughly half of the students justified each of their predictions by appealing to the effect of only a single variable.

If we examine average adults, rather than the typical college students, they do somewhat better (Kuhn & Dean, in press). About half the members of a community choral group, representative of a broad cross-section of the adult population, showed inconsistency in causal attribution in the course of their successive interpretations of accumulating evidence, either at least once initially judging a variable as noncausal and later judging it to be causal, or initially judging a variable as causal and later judging it noncausal, or showing both inconsistencies. Similarly, over half showed inconsistency between implicit and explicit causal judgements. Almost half were inconsistent in causal attributions across three prediction questions. Like the sixth graders, these adults failed to implicate as causal in their implicit attributions as many variables as they needed to in order to yield correct predictions. Over a quarter appealed to the effect of only a single (usually shifting) variable in their prediction judgements, and over half appealed to the effect of only two of the four variables. All of these characteristics of an immature mental model of causality, in contrast, were infrequent in a college student population (Kuhn & Dean, in press).

We can thus point to an inadequate mental model of multivariable causality as a constraint on children's and even many adults' ability to predict effects of multiple variables on an outcome. Additional constraints come into play when individuals are required to bring new evidence to bear on their causal models. The most fundamental is a weakness in metacognitive awareness of new evidence, versus the prevailing mental model, as the basis for one's inferences. And when new evidence is in fact brought to bear on a claim, there emerges the further constraint of faulty inference rules. Notably, factors may be judged causal due simply to their association with the outcome (overattribution) or be judged noncausal because one or more other factors are assumed responsible for the outcome (underattribution, or discounting). In a series of studies (Kuhn, Schauble, & Garcia-Mila, 1992; Kuhn et al., 1995, 2000: Kuhn & Dean, in press), we have asked both children and adults to draw inferences of causality and noncausality when multiple factors are present in conjunction with an outcome, and we have found all of these phenomena to be common.

It is useful to see these weaknesses in a developmental framework. Before saying more about the performance of older participants, we therefore note the results of a study of 4–6-year-olds (Kuhn & Pearsall, 2000). It was hypothesised that children at this young age would fail to distinguish between theoretical explanations and evidence as a basis for their simple knowledge claims, in a parallel way to the confusion between theory and evidence as justifications for causal inferences that we observed in older children and adults. Children were shown a sequence of pictures in which, for example, two runners compete in a race. Certain cues suggest a theoretical explanation as to why one will win, e.g., one has fancy running shoes and the other does not. The final picture in the sequence provides evidence of the outcome, e.g., one of the runners holds a trophy and exhibits a wide grin. When children are asked to indicate the outcome and to justify this knowledge, 4-year-olds show a fragile distinction between the two kinds of justification—"How do you know?" and "Why is it so?"—in other words, the evidence for their claim (the outcome cue in this case) versus their explanation as to why it is plausible (the theory-generating cue). Rather, the two merge into a single representation of what happened, and the child tends to choose as evidence of what happened the cue having greater explanatory value as to why it happened. Thus, in the race example, young children often answered the "How do you know [he won]?" question not with evidence ("He's holding the trophy") but with a theory of why this state of affairs makes sense (e.g., "Because he has fast sneakers").

Similarly, in another set of pictures in which a boy is shown first climbing a tree and then down on the ground holding his knee, the ''How do you know [that he fell]?'' question was often answered, ''Because he wasn't holding on carefully''. These confusions between theory and evidence diminish sharply among 6-year-olds, who still make mistakes but the majority of the time distinguish the evidence for their event claim from a theoretical explanation that makes the claim plausible. Findings by other investigators support this characterisation of preschool children as having weak metacognitive control of their own knowing, for example failing to differentiate different sources of their own knowledge claims (Gopnik & Graff, 1988; Whitcombe & Robinson, 2000) and claiming that they had ''always known'' a piece of information they had just been given (Taylor, Esbensen, & Bennett, 1994).

When older children and adults are asked to coordinate new evidence with their existing mental models of a domain, similar indications are apparent of a fragile meta-level distinction between theory and evidence as the basis for one's inferences. In one of the problems posed to children and adults by Kuhn et al. (1995), for example, participants were asked to identify which of five variables (see Table 2) influenced the popularity of children's TV programmes.

The first programme that Geoff (a pseudonym) selected to examine had commercials but no music or humour, was 2 hours long, and on Tuesday, with a popularity rating of fair. Geoff interpreted this outcome as confirming his earlier expressed theories:

> You see, this shows you that the factors I was saying about, that you have to be funny to make it good or excellent, and the day doesn't really matter, and it's too long.

The second instance Geoff chose added humour and music, and changed the length to a half hour and the day to Wednesday, with an outcome of excellent. Now Geoff concludes, based on the two instances:

> It does make a difference when you put music and have commercials and the length of time and the humour. Basically the day is the only thing that doesn't really matter.

TABLE 2
Causal and noncausal effects in the TV problem (Geoff's problem)

Music (M or −)	Simple causal effect
Commercials (C or −)	Interactive causal effect (causal only in absence of music)
Length (0, 1, 2)	Curvilinear causal effect ($0 > 1 = 2$)
Day (t or w)	Noncausal
Humour (f or s)	Noncausal

For length variable, $0 = \frac{1}{2}$ hour, $1 = 1$ hour, $2 = 2$ hour. For day variable, t = Tuesday, w = Wednesday. For humour variable, f = funny, s = serious

Geoff thus utilised these two pieces of evidence as an opportunity to confirm all his theories. Three factors that covaried with outcome (music, humour, and length) he interpreted as causal. He also included commercials as causal even though this did not vary, but excluded day of the week, which did vary, as noncausal. He selects data for observation that he believes will "illustrate" the correctness of these theories. To the extent that the outcome data pose interpretive problems, he draws on a variable set of inference rules, applying to each variable those rules that are most protective of his theories. Presence or absence of commercials, for example, is implicated as causal based on its presence in just one successful outcome. When possible, however, in the case of the three other variables also believed causal, Geoff applies the more stringent covariation rule as the basis for inferring causality. As Geoff's reasoning illustrates, the explanatory burden shifts from one variable to another in a way that allows theories to be maintained.

It should be emphasised that in these studies we are not pitting individuals' prior knowledge against new information, asking them to forego the former in favour of the latter. The respondent is free to say, "Here are the implications of your data, but I don't find them convincing, and choose not to modify my theories based on them." Such an individual exhibits the meta-level awareness and management of their own cognition represented in Figure 1. It is the individual who is not aware of how prior beliefs and the presented information relate, because the two have not been represented as distinct entities, who is the cause for concern.

DEVELOPMENTAL COMPARISONS

Our earlier research on causal inferences has been situated in the contexts of scientific reasoning or knowledge acquisition, although we make the case that the forms of reasoning are in many respects identical (Kuhn & Dean, in press). In these studies (Kuhn et al., 1992, 1995, 2000), participants engage in investigation as well as inference, choosing from a database the cases they wish to examine. Although all of the errors that have been described are observed in both children and adults, and the patterns of change observed in microgenetic studies are similar (Kuhn, 1995), overall the performance of adults is superior to that of children.

In a line of current work described here, we have asked whether the same performance patterns can be observed in paper-and-pencil measures. Several different paper-and-pencil instruments have been tried. The most straightforward is the drugs problem in Table 3 and a more difficult one is the reading improvement problem in Table 4, since the former isolates each variable individually and the latter does not.

TABLE 3
Drugs problem

Researchers are trying three new drugs with AIDS patients to see if they improve patients' ability to avoid infections. The names of the drugs are ALON, BENA, and CREL. For a six-month period, some of the patients in the study took all three drugs, some took only two, some took one, and a final group didn't take any. Below are the results for each group. Analyze these results and then answer the questions.

Patients who took ALON, BENA, and CREL: Average frequency of infections: Low	Patients who took only ALON: Average frequency of infections: High
Patients who took ALON and BENA: Average frequency of infections: Medium	Patients who took only BENA: Average frequency of infections: Medium
Patients who took ALON and CREL: Average frequency of infections: Medium	Patients who took only CREL: Average frequency of infections: Medium
Patients who took BENA and CREL: Average frequency of infections: Low	Patients who took no drug: Average frequency of infections: High

Did the drug ALON have any effect on patients' ability to avoid infections?
YES NO UNSURE
How do you know?

Did they drug BENA have any effect on patients' ability to avoid infections?
YES NO UNSURE
How do you know?

Did they drug CREL have any effect on patients' ability to avoid infections?
YES NO UNSURE
How do you know?

Drugs problem causal structure.

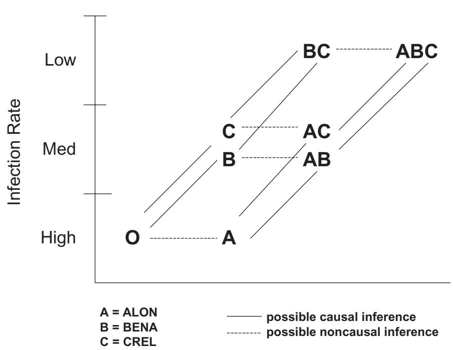


TABLE 4
Reading improvement problem

Which factors affect reading performance?

A school district is experimenting with new methods of improving beginning reading instruction. In different classrooms across the district, they are instituting a new reading curriculum, teacher aides, and reduced class size. Here are some preliminary results.

Type	Average reading performance
Regular classrooms	Poor
Classrooms with new curriculum and teacher aid	Greatly Improved
Classrooms with new curriculum and reduced class size	Improved
Classrooms with teacher aide and reduced class size	Improved
Classrooms with new curriculum, teacher aide, and reduced class size	Greatly Improved
Classrooms with teacher aide	Improved

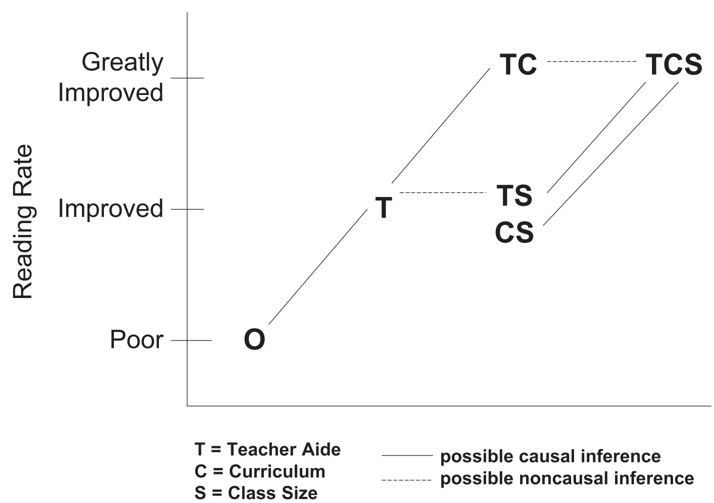
What conclusions do you draw from these findings? Justify your answers by referring to the data.

Is the new curriculum beneficial? How do you know?

Is the teacher aide beneficial? How do you know?

Is the reduced class size beneficial? How do you know?

Reading improvement problem causal structure.



In the reading improvement problem, community college students very rarely judge class size as having no causal effect (see Table 5). Instead, the overattribution illustrated earlier is the dominant response: Whatever factors are present in the context of an outcome contribute to that outcome. Hence a typical response regarding the new curriculum is:

> Yes it is [beneficial], because all the cases where a new curriculum has been applied the class has improved.

The reasoning is similar with respect to class size, despite its actual noncausal status:

> Class size is also beneficial because according to the data, improvement was evident.

Or,

> Yes, one case shows it [class size] greatly improved performance.

While not displayed at such high levels, underattribution of causality also occurs. The most common basis for it is to ignore the evidence entirely and resort to belief for justification. For example:

> A teacher aide is not beneficial because each teacher has their own method of teaching, so a teacher aide can create confusion.

Even where the data do support exclusion of a factor as having causal power, this ignoring of the data (despite explicit instruction to consider it) and exclusive reliance on one's prior beliefs are common. For example:

> Yes, reduced class size makes a difference because the numbers of children are small so they can learn better and faster.

Interestingly, even beginning graduate students in education have difficulty with this problem. A number decline to make inferences, citing the impossibility of examining each factor in isolation. They were unwilling to use a perfectly valid subtractive method, comparing outcomes for example, of CTS and CT, as a basis for inferring that S had no causal effect.

The pollution problem (Table 6) was introduced to explore the possibility that a visual representation of multiple factors influencing an outcome might facilitate reasoning. As evident in Table 5, at least in this form, it did not. Table 5 summarises performance on the various instruments by a number of different samples (although not all instruments are presented to all groups).

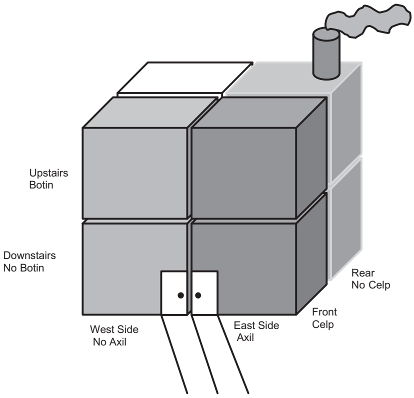
TABLE 5
Performance on causal reasoning problems by problem type and group

Group	*N*	*Evidence-based justification*			*Valid inclusion*			*Valid exclusion*		
		D	P	R	D	P	R	D	P	R
8th gr	72*	78.9%	58.3%		18.3%	2.8%		11.3%	0.0%	
CC	60*	82.1%		61.7%	46.2%		8.3%	7.7%		3.3%
Grad	216			93.1%			27.3%			10.6%
AD	84	94.0%	88.1%		48.8%	28.6%		20.2%	22.6%	

Group	*N*	*Invalid inclusion (theory-based)*			*Invalid inclusion (evidence-based)*			*Invalid exclusion (theory-based)*		
		D	P	R	D	P	R	D	P	R
8th gr	72*	5.6%	9.7%		29.6%	60.6%		19.7%	23.6%	
CC	60*	5.1%		31.7%	38.5%		56.7%	7.7%		8.3%
Grad	216			4.6%			52.3%			11.1%
AD	84	0.0%	0.0%		25.0%	32.1%		3.6%	6.0%	

D = Drugs problem, P = Pollution problem, R = Reading problem *Drugs *N*: 8th = 71, CC = 39. Percentages are based on total number of responses. CC = urban community college students. Grad = beginning graduate students in education. AD = community adults (members of a choral group).

TABLE 6 (below and opposite)
Pollution problem



A factory uses three different chemicals to make its product, gumball machines. In the different parts of the factory, different chemicals are needed for the work done in that section of the factory.

In all of the sections on the east side of the factory, Axil is used.

In all of the sections in the upstairs floor of the factory, Botin is used.

In all of the sections in the front of the factory, Celp is used. (See picture)

So, for example, someone who worked downstairs on the east side front section of the factory would be exposed to Axil and to Celp, but not to Botin.

The factory owners are worried about air pollution in the factory. As you can see from the picture, the pollution is worse (darker colour) in some sections of the factory and not as bad in other sections (lighter colours).

Based on what you see, answer the following questions.

Does the chemical **Axil** have any effect on the amount of pollution in the factory?

YES NO UNSURE

How do you know?

Does the chemical **Botin** have any effect on the amount of pollution in the factory?

YES NO UNSURE

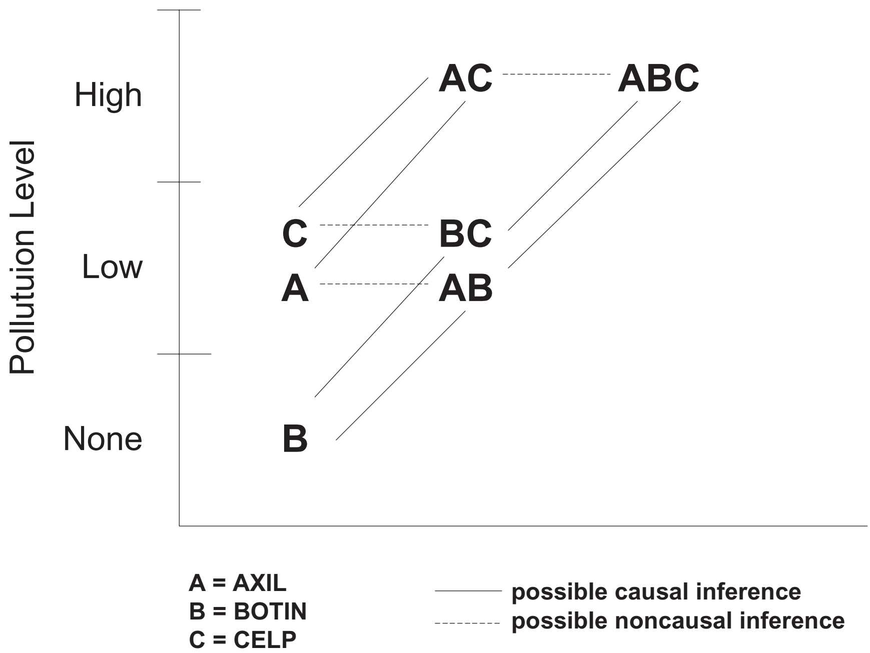
How do you know?

Does the chemical **Celp** have any effect on the amount of pollution in the factory?

YES NO UNSURE

How do you know?

Pollution problem causal structure



These data are consistent with the pattern suggested earlier. Developmental differences appear, although even adults continue to make errors.

IMPLICATIONS FOR UNDERSTANDING CAUSAL INFERENCE

The implicit assumption underlying research in the adult causal inference literature is that people's understanding of multiple causality reflects a standard scientific model: Multiple effects may contribute to an outcome in an additive manner—as long as background conditions remain constant, these effects are expected to be consistent, i.e., the same antecedent does not affect an outcome on one occasion and fail to do so on another, or affect the same outcome differently on one occasion than another. (A more complex model, encompassing interaction effects, presumes understanding of the simpler main effects of an additive model.) Indeed, all of science is predicated on such a model. It is not clear how the world would operate in the absence of these assumptions, and hence it is not surprising that research on human causal inference has implicitly adopted such a model as a starting point. Nonetheless, the findings described here suggest that it is a mistake to do so.

Recognition of the role of theoretical understanding of mechanism in causal inference has been seen by causality researchers such as Cheng as providing needed reduction in complexity of a general model of causal inference. By reducing the number of variables considered as causal candidates, theory may facilitate multivariable causal inference. The findings described here, we believe, are more consistent with a model of causal inference as entailing the coordination of the two components—theory and evidence—in a way that complicates, more than it simplifies, the inference process. An individual brings a repertoire of inference strategies or rules (of varying validity) to the task of interpreting the implications of evidence, and these may be drawn on selectively in the service of theory–evidence coordination. As a result, consistency in causal attribution is commonly absent, with different rules applied at different times for different causal candidates, as expedient in the goal of achieving coordination by protecting theories from discrepant evidence. At its extremes, the coordination effort may produce, on the one hand, failure to interpret accurately the implications of evidence, due to application of faulty rules, with consequent representation of evidence as demonstrating the correctness of theories for which the evidence in fact offers no support. At the other extreme, consistent interpretation of evidence is achieved through exclusive application of valid inference rules and implications for theory are recognised. In the former case, theory and evidence are not clearly distinguished. In the latter, they are perfectly coordinated (even if the individual decides theory modifications may not be warranted by the evidence). The majority of cases, however, are likely to be intermediate between these two extremes, reflected in intra-individual variability in inference rules used, but with some accurate interpretation of the implications of evidence.

Consistent with such a model is a more subtle and variable role of theoretical belief than the one postulated by Cheng's MCI model. Theory–evidence coordination is a complex, dynamic process, with the role of theory not confined to an initial phase in which variables are excluded from consideration on theoretical grounds. Our data reveal frequent instances of an individual's shift from an earlier declaration of a variable as noncausal to a subsequent claim that it is causal. Thus, variables that are initially excluded are not necessarily forgotten. Evidence may continue to be evaluated with respect to their causal status. Conversely, the frequently observed shift from an early declaration of a variable as causal to a subsequent judgement of noncausality indicates that judgements of noncausality are not in fact exclusively theory-based (Kuhn & Dean, in press; Kuhn et al., 1995).

Although intra-individual variability is common in intellectual functioning (Siegler, 1994) and need not require specific explanation, at least some intra-individual variability, we have proposed, is attributable to individuals' imperfect skills in coordinating theories and evidence. Confinement to a university population, as is typical in the adult cognition literature, may reduce—although it does not eliminate—this variability. The distribution of usage of strategies or rules of differing effectiveness within an individual may change over time, as it is seen to do in microgenetic studies (Kuhn, 1995; Kuhn et al, 1995; Siegler & Crowley, 1991). It is here especially that we see the value of a developmental framework, which allows both intra-individual and inter-individual variation among adults to be interpreted developmentally. In this light, we have as much to learn from adults as we do from children with respect to how reasoning skills develop. It is this perspective that has led us to devote so much attention to the thinking of adults in an article in which we are arguing for the value of a developmental perspective.

PARALLELS IN THE DEVELOPMENT OF ARGUMENTIVE REASONING

Space does not permit us to examine argumentive reasoning in as great detail as we have causal reasoning. A very brief review is nonetheless worthwhile to illustrate that the case we have made for a developmental framework is not limited to the topic of inductive causal inference. Although much less empirical research has been done on argument, the parallels are instructive.

Like causal inference, argument is widely accepted as a basic form of human reasoning that does not need to be learned. The term ''argument'' of course is used to refer to two quite different acts. An individual constructs an argument to support a claim. The dialogic process in which two or more people engage in debate of opposing claims can be referred to as argumentation or argumentive discourse to distinguish it from argument as product. Nonetheless, implicit in argument as product is the advancement of a claim in a framework of evidence and counterclaims that is characteristic of argumentive discourse, and the two kinds of argument are intricately related (Billig, 1987; Kuhn, 1991). Most empirical research on argument has been devoted to argument as product.

Paralleling the case of causal inference, rudimentary skills of argument, of both process and product types, have been widely taken for granted as within the competence of children as well as adults. Empirical research with children has focused on demonstrating the competence of very young children to appropriately justify their claims and even to engage in effective argumentation (Eisenberg & Garvey, 1981; Orsolini, 1993; Stein & Miller, 1993).

Research examining the arguments (as products) of adolescents and adults, in contrast, report serious weaknesses. In supporting a claim, respondents commonly fail to construct two-sided arguments or to distinguish evidence and explanation in support of their claims (Brem & Rips, 2000; Kuhn, 1991, 2001; Kuhn, Shaw, & Felton, 1997; Perkins, 1985; Voss & Means, 1991). They also show wide susceptibility to belief bias (Klaczynski, 2000). Relatively little research, however, has been devoted to the more complex skills that are involved when one undertakes to guide the process of competitively co-constructing an argument in the context of discourse.

According to Walton (1989), skilled argumentation has two goals. One is to secure commitments from the opponent that can be used to support one's own argument. The other is to undermine the opponent's position by identifying and challenging weaknesses in his or her argument. Drawing on Walton's analysis, Felton and Kuhn (2001) and Kuhn and Udell (2003) identify two potential forms of development in argumentive discourse skills. One is enhanced understanding of discourse goals and the other is application of effective strategies to meet these goals. These two forms of development can be predicted to reinforce one another. Progress in use of discourse strategies is propelled by a better understanding of discourse goals. At the same time, exercise of these strategies in discourse promotes more refined understanding of the goals of argumentive discourse.

To examine development in argumentive discourse skills, Felton and Kuhn (2001) conducted a cross-sectional comparison of the dialogues of young teens and community college young adults arguing about capital punishment. The results revealed striking differences between the two groups. Teens' discourse focused largely on the arguments supporting their own position, at the expense of addressing the arguments of their opponents. Teens appear to interpret the goal of argumentive discourse as prevailing over an opponent by superior presentation of one's own position. This objective, if successfully met, undermines the opponent's position, but without addressing the opponent's argument.

Adults, in contrast, in addition to advancing their own arguments, were more likely to address the opponent's argument, most often through counterargument. In undertaking to undermine their opponent's argument, as well as advance their own argument, adults' dialogues thus came closer to achieving the dual goals of argumentive discourse. These appear to be skills that need to develop during childhood and adolescent years. Deep-level processing of the opponent's argument, in addition to articulating one's own argument and negotiating the mechanics of discourse, may represent cognitive overload for the novice arguer.

Based on these findings, Kuhn and Udell (2003) undertook an experimental study with young adolescents. Following a several-month-long intervention designed to exercise and develop their argumentation skills, participants showed a decrease in the proportion of dialogue devoted to exposition, that is, articulation and clarification of one's own position and perspective. Furthermore, they showed an increase in the proportion of dialogue devoted to challenges that address the partner's claims and seek to identify weaknesses in them, reflecting understanding of Walton's (1989) second goal of argumentation.

This research has helped us to understand the cognitive skills involved in dialogic argumentation. In the present context, their significance lies in the case we have undertaken to make for a developmental perspective in the study of informal (or formal, for that matter) reasoning. The perspective is simply captured. To fully understand a mature competency, watch it develop.

REFERENCES

Ahn, W-k., Kalish, C. W., Medin, D. L., & Gelman, S. A. (1995). The role of covariation versus mechanism information in causal attribution. *Cognition*, *54*, 299–352.

Billig, M. (1987). *Arguing and thinking: A rhetorical approach to social psychology*. Cambridge: Cambridge University Press.

Brem, S. K., & Rips, L. J. (2000). Explanation and evidence in informal argument. *Cognitive Science*, *24*(4), 573–604.

Cheng, P. W. (1997). From covariation to causation: A causal power theory. *Psychological Review*, *104*(2), 367–405.

Cheng, P. W., & Novick, L. R. (1990). A probabilistic contrast model of causal induction. *Journal of Personality and Social Psychology*, *58*, 545–547.

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

Dixon, J., & Tuccillo, F. (2001). Generating initial models for reasoning. *Journal of Experimental Child Psychology*, *78*, 178–212.

Einhorn, H. J., & Hogarth, R. M. (1986). Judging probable cause. *Psychological Bulletin*, *99*, 3–19.

Eisenberg, A. R., & Garvey, C. (1981). Children's use of verbal strategies in resolving conflicts. *Discourse Processes*, *4*, 149–170.

Felton, M., & Kuhn, D. (2001). The development of argumentive discourse skill. *Discourse Processes*, *32*(2&3), 135–153.

Gelman, S. A., & Wellman, H. M. (1998). Enabling constraints for cognitive development and learning: Domain specificity and epigenesis. In D. Kuhn & R. S. Siegler (Eds.), *Handbook of child psychology: Cognition, language, and perception* (5th ed., Vol. 2). New York: Wiley.

Gentner, D., & Gentner, D. R. (1983). Flowing waters or teeming crowds: Mental models of electricity. In D. Gentner & A. L. Stevens (Eds.), *Mental models*. Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

Glymour, C. (1998). Learning causes: Psychological explanations of causal explanation. *Minds and Machines*, *8*, 39–60.

Glymour, C. (2001). *The mind's arrows: Bayes nets and graphical causal models in psychology*. Boston, MA: MIT Press.

Glymour, C., & Cheng, P. W. (1998). Causal mechanism and probability: A normative approach. In M. Oaksford & N. Chater (Eds.), *Rational models of cognition* (pp. 296–313). Oxford: Oxford University Press.

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*(5), 1366–1371.

Gopnik, A., & Sobel, D. M. (2000). Detecting blickets: How young children use information about novel causal powers in categorization and induction. *Child Development*, *71*(5), 1205–1222.

Gopnik, A., Sobel, D. M., Schulz, L. E., & Glymour, C. (2001). Causal learning mechanisms in very young children: Two-, three-, and four-year-olds infer causal relations from patterns of variation and covariation. *Developmental Psychology*, *37*(5), 620–629.

Harris, P. L., German, T., & Mills, P. (1996). Children's use of counterfactual thinking in causal reasoning. *Cognition*, *61*, 233–259.

Hilton, D. J., & Slugoski, B. R. (1986). Knowledge-based causal attribution: The abnormal conditions focus model. *Psychological Review*, *93*(1), 75–88.

Keil, F. C. (1998). Cognitive science and the origins of thought and knowledge. In R. M. Lerner (Ed.), *Handbook of child psychology: Theoretical models of human development* (5th ed., Vol. 1). New York: Wiley.

Keselman, A. (2003). Supporting inquiry learning by promoting normative understanding of multivariable causality. *Journal of Research in Science Teaching*, *40*(9).

Klaczynski, P. A. (2000). Motivated scientific reasoning biases, epistemological beliefs, and theory polarization: A two-process approach to adolescent cognition. *Child Development*, *71*(5), 1347–1366.

Kuhn, D. (1989). Children and adults as intuitive scientists. *Psychological Review*, *96*, 674–689.

Kuhn, D. (1991). *The skills of argument*. Cambridge: Cambridge University Press.

Kuhn, D. (1995). Microgenetic study of change: What has it told us? *Psychological Science*, *6*, 133–139.

Kuhn, D. (2001). How do people know? *Psychological Science*, *12*(1), 1–8.

Kuhn, D. (in press). *Education for thinking*. Cambridge, MA: Harvard University Press.

Kuhn, D., Black, J., Keselman, A., & Kaplan, D. (2000). The development of cognitive skills to support inquiry learning. *Cognition and Instruction*, *18*(4), 495–523.

Kuhn, D., & Dean, D. (in press). Connecting scientific reasoning and causal inference research. *Journal of Cognition and Development*.

Kuhn, D., Garcia-Mila, M., Zohar, A., & Andersen, C. (1995). Strategies of knowledge acquisition. In *Monographs of the Society for Research in Child Development* (Vol. 60).

Kuhn, D., & Pearsall, S. (2000). Developmental origins of scientific thinking. *Journal of Cognition and Development*, *1*, 113–129.

Kuhn, D., Pennington, N., & Leadbeater, B. (1983). Adult thinking in developmental perspective: The sample case of juror reasoning. In P. Baltes & O. Brim (Eds.), *Life-span development and behavior* (Vol. 5). New York: Academic Press.

Kuhn, D., & Phelps, E. (1982). The development of problem-solving strategies. *Advances in Child Development and Behavior*, *17*, 1–44.

Kuhn, D., Schauble, L., & García-Mila, M. (1992). Cross-domain development of scientific reasoning. *Cognition and Instruction*, *9*(4), 285–327.

Kuhn, D., Shaw, V., & Felton, M. (1997). Effects of dyadic interaction on argumentive reasoning. *Cognition and Instruction*, *15*, 287–315.

Kuhn, D., & Udell, W. (2003). The development of argument skills. *Child Development*, *74*(5), 1245–1260.

Kuhn, D., Weinstock, M., & Flaton, R. (1994). How well do jurors reason? Competence dimensions of individual variation in a juror reasoning task. *Psychological Science*, *5*, 289–296.

Lien, Y., & Cheng, P. W. (2000). Distinguishing genuine from spurious causes: A coherence hypothesis. *Cognitive Psychology*, *40*(2), 87–137.

Mackie, J. L. (1974). *The cement of the universe: A study of causation*. London: Oxford University Press.

Newsome, G. L. (2003). The debate between current versions of covariation and mechanism approaches to causal inference. *Philosophical Psychology*, *16*(1), 87–107.

Orsolini, M. (1993). "Dwarfs don't shoot": An analysis of children's justifications. *Cognition and Instruction*, *11*(3&4), 281–297.

Perkins, D. N. (1985). Post-primary education has little impact upon formal reasoning. *Journal of Educational Psychology*, *77*, 563–571.

Rehder, B. (2003). Categorization as causal reasoning. *Cognitive Science*, *27*, 709–748.

Roese, N. J., & Olson, J.M. (1996). Counterfactuals, causal attributions, and the hindsight bias: A conceptual integration. *Journal of Experimental Social Psychology*, *32*, 197–227.

Siegler, R. S. (1994). Cognitive variability: A key to understanding cognitive development. *Current Directions in Psychological Science*, *3*(1), 1–5.

Siegler, R. S., & Crowley, K. (1991). The microgenetic method: A direct means for studying cognitive development. *American Psychologist*, *46*(6), 606–620.

Stein, N. L., & Miller, C. A. (1993). The development of memory and reasoning skill in argumentative contexts: Evaluating, explaining, and generating evidence. In R. Glaser (Ed.), *Advances in instructional psychology*. Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

Taylor, M., Esbensen, B. M., & Bennett, R. T. (1994). Children's understanding of knowledge acquisition: The tendency for children to report that they have always known what they have just learned. *Child Development*, *65*(6), 1581–1604.

Voss, J., & Means, M. (1991). Learning to reason via instruction in argumentation. *Learning and Instruction*, *1*, 337–350.

Walton, D. N. (1989). Dialogue theory for critical thinking. *Argumentation*, *3*, 169–184.

Whitcombe, E. L., & Robinson, E. J. (2000). Children's decisions about what to believe and their ability to report the source of their belief. *Cognitive Development*, *15*(3), 329–346.

Wilkening, F. (1981). Integrating velocity, time, and distance information: A developmental study. *Cognitive Psychology*, *13*(2), 231–247.