COGNITIVE SCIENCE
A Multidisciplinary Journal

Cognitive Science 34 (2010) 826–862

ISSN: 0364-0213 print / 1551-6709 online
DOI: 10.1111/j.1551-6709.2010.01108.x

The Feasibility of Folk Science

Frank C. Keil

Department of Psychology, Yale University

Received 1 June 2009; received in revised form 4 March 2010; accepted 4 March 2010

Abstract

If folk science means individuals having well worked out mechanistic theories of the workings of the world, then it is not feasible. Laypeople's explanatory understandings are remarkably coarse, full of gaps, and often full of inconsistencies. Even worse, most people overestimate their own understandings. Yet recent views suggest that formal scientists may not be so different. In spite of these limitations, science somehow works and its success offers hope for the feasibility of folk science as well. The success of science arises from the ways in which scientists learn to leverage understandings in other minds and to outsource explanatory work through sophisticated methods of deference and simplification of complex systems. Three studies ask whether analogous processes might be present not only in laypeople but also in young children and thereby form a foundation for supplementing explanatory understandings almost from the start of our first attempts to make sense of the world.

Keywords: Intuitive theories; Expertise; Folk science; Cognitive development; Formal science; Concepts; Explanations; Causal reasoning

1. Introduction

All of us, from the most sophisticated adults to the youngest children, often engage in what is commonly called ''folk science,'' that is, certain ways of understanding the natural and artificial world that arise more informally and not as direct reflections of formal instruction in scientific principles (Carey, 1988). There is now extensive work on how children and adults have developed folk psychologies (Wellman, 1990), folk physics (Proffitt, 1999; Vosniadou, 2001), and folk biologies (Inagaki & Hatano, 2002), as well as some indications of folk sciences in such areas as the behaviors of materials and substances (folk chemistry; Au, 1994), the behaviors of heavenly bodies (folk cosmology; Siegal, Butterworth, &

Correspondence should be sent to Frank C. Keil, Department of Psychology, Yale University, 2 Hillhouse Avenue, New Haven, CT 06520. E-mail: frank.keil@yale.edu

Newcombe, 2004), and the nature of value transactions (folk economics; Lakshminaryanan, Chen, & Santos, 2008). Without explicit instruction in such areas, people seem to develop domain-specific ways of thinking about relatively bounded sets of phenomena such as the behavior of solid objects, living kinds, and the minds of others.

These domain-specific understandings have been referred to as ‘‘intuitive theories’’ or ‘‘naïve theories,’’ on the assumption that they reflect sets of beliefs that cohere in a manner that resembles, in important respects, scientific theories (Carey, 1985; Carey & Spelke, 1996; Slaughter & Gopnik, 1996). In particular, intuitive theories are assumed to have coherence, consistency, and predictive value. In addition, such intuitive theories are thought to specify ontological kinds. Ontological kinds, in turn, are thought to be a product of causal explanatory systems that posit certain kinds as foundational entities within each system. Preschoolers, and perhaps even infants, are thought to have senses of ontological kinds that arise from their intuitive theories or folk sciences (Carey, 1985, 2009; Heyman, Phillips, & Gelman, 2003). Although researchers may disagree on just how early different theoretical domains appear, there appears to be considerable consensus on a folk physics and folk psychology having roots going back to infancy (Carey, 1985; Leslie, 1994). In addition, for 25 years, ever since Carey’s (1985) pioneering work on conceptual change, researchers have asked whether conceptual change in children might be analogous to the ways in which concepts change in the formal sciences as a consequence of theory change. If children’s concepts are embedded in rich intuitive theories, perhaps it is useful to think of them as little scientists going through sequences of theoretical revolutions with corresponding conceptual changes.

At one level, this body of work is clearly right. From infancy onwards, humans seem to assess the domain of regularities they encounter and then reason in ways that reflect the causal regularities specific to that domain. For example, psychological entities seem to be governed by different causal principles (e.g., action at a distance) than the most typical physical mechanical ones and we all seem to constrain our beliefs in ways that honor these domain-specific principles. However, such domain-specific constraints may not be sufficient for something to count as an intuitive theory, especially given that aspects of them may operate in other species as well (Carey & Spelke, 1996).

Naïve theories can also lead people astray. Indeed, a large subspecialty of cognitive science focusing on misconceptions has documented how laypeople and children make systematic mistakes in their reasoning about physical mechanics (e.g., Caramazza, McCloskey, & Green, 1981; Bertamini, Spooner, & Hecht, 2004), biology (e.g., Inagaki & Hatano, 2002; Shtulman, 2006; Shtulman & Schulz, 2008), and psychology (e.g., Leslie, Friedman, & German, 2004; Malle, Knobe, & Nelson, 2007). Although such misconceptions may lead to misleading ‘‘deficit models’’ of cognitive development and instruction (Zimmerman, 2007), they certainly do exist and have been argued to serve as evidence for the coherence of folk science domains; coherence that drives people to make mistakes even in the face of real-world counter-evidence. Thus, misconceptions sometimes have been used as evidence for intuitive theories.

In short, there seems to be an emerging consensus about the existence of many folk sciences across all cultures that lead both to real successes at understanding the world and to misconceptions. Yet one critical problem seems largely unaddressed. If we look in more detail at the nature of intuitive theories, how feasible is it that they should suffice as vehicles for understanding? Are they sufficiently well articulated, and in the right ways, to enable laypeople and children to make sense of the world around them? Do they meet reasonable levels of coherence, consistency, and predictive power to count as theories? Here, I argue that the answer is not straightforward. If one looks at intuitive theories and folk science through one lens, namely their sufficiency as stand-alone explanatory systems, the matter seems quite hopeless and it is a real mystery as to how anyone gains explanatory insights into the workings of the world. That lens raises similar worries, however, when refocused on more formal science and, as a result, suggests a different, more socially embedded analysis of how formal science succeeds. This analysis and the reasons that formal science succeeds point toward a new way of looking at folk science as well, and a series of studies is then described showing that folk science is not only feasible but also that the foundations making it feasible are present at surprisingly early ages.

More specifically, this article considers the nature of the intuitive understandings in the folk sciences and documents their severe limitations when considered as mechanistic theories. They are remarkably incomplete and often contain inconsistencies and contradictions, and most laypeople grossly underestimate these limitations as well. Although these shortcomings may seem to cripple any lay ability to make sense of the world, a different alternative emerges if we consider more recent views of the formal sciences in which gaps and inconsistencies and other kinds of gross simplifications are also shown to be commonplace. Formal sciences surmount those limitations by developing and navigating a social infrastructure of knowledge in which there are divisions of cognitive labor and sophisticated mechanisms for recognizing appropriate experts and knowing when and how to defer to them.

This alternative view of formal science suggests a reanalysis of folk science as well, in which roughly comparable mechanisms enable laypeople to have a sense of relevant areas of expertise and of how to defer to them. Despite having very crude internal models of how the world works, laypeople may have sophisticated ways of navigating the terrain of expert clusters in their culture. Moreover, this ability may emerge early in development and may be how causal understanding is framed even by young children, that is, as embedded in larger social networks of knowledge resources. To illustrate this idea with one kind of example, three studies are described that document how children, even as they know very little about specific mechanisms, come to sense fertile vs. infertile areas of expertise. Finally, the implications of this alternative view of folk science are discussed both in terms of the development of explanatory understandings and in terms of the challenges that remain in understanding how folk science is used on a daily basis.

There have been many debates about the meanings of ''success'' or ''progress'' in the formal sciences and whether such meanings require a realist philosophy of science (Psillos, 1999, 2009). Here, it is simply assumed that the formal sciences are successful because they enable their practitioners to characterize aspects of the world in ways that allow them to intervene on elements of the world so as to come closer to desired goals (Woodward, 2003) and to test hypotheses in more rigorous ways (including cases where intervention is not possible). Sciences make progress when they become more successful over extended time periods (allowing for shorter term dead ends and detours).

A folk science would be considered feasible to the extent that it enables laypeople to characterize the world in ways that enable them to intervene more effectively and confidently to achieve goals and that enable them to test and justify hypotheses. One key question is whether folk science could be feasible in this sense if it is primarily understood in terms having internal mechanistic representations of the world. That is, are people's internal representations sufficiently articulated to directly enable them to have more effective and confident interventions and more powerful ways of testing and justifying hypotheses? I will argue that, in general, they are not, but that there are other facets to folk science that give it much more power.

1.1. Reasons for concern

There are several reasons to be concerned about the feasibility of folk science if it is to be understood as resting on the use of intuitive, mechanistic theories that enable informative interventions. Most of these reasons revolve around evidence, much of it quite recent, that laypeople seem to have extraordinarily weak representations of how things work and why phenomena occur. If intuitive theories imply richly articulated sets of beliefs that capture detailed mechanistic understandings of natural and artificial phenomena, then such theories may be largely absent in the minds of the vast majority of people. The mere presence of domain-specific patterns of reasoning may be too weak a basis on which to infer folk theories.

We have known for some time that people do not think in a manner that neatly follows the laws of classical deductive logic (Braine, 1978; Rips, 2002), but the problem goes far deeper than logical fallacies or failures to reason correctly with logical operations such as conjunctions (Tversky & Kahneman, 1983). The concepts and beliefs that make up a system of knowledge in a domain seem to be surprisingly impoverished. In particular, people have huge explanatory gaps, inconsistencies in their beliefs, and often outright contradictions. All of these problems are then compounded by the finding that people grossly underestimate these problems (Rozenblit & Keil, 2002), and that they routinely get false ''rushes'' of insight, when in fact they have not increased their real understanding at all (Weisberg, Keil, Goodstein, Rawson, & Gray, 2008).

First, there is the problem of gaps and holes. People are frequently unable to come up with complete explanations of mechanisms, even for surprisingly simple systems. For example, when people are asked to explain the workings of objects as simple as cylinder locks, flush toilets, and bicycles, they may only be able to come up with one or two correct components (Rozenblit & Keil, 2002). Even those few components may then be related incorrectly (Lawson, 2006). People are often initially unaware of these gaps until they emerge in their attempts to explain how a thing works.

Second, there are surprising inconsistencies. A person may have two different beliefs that cannot be coherently put together into a larger belief system. These inconsistencies may sometimes emerge after an inferential step or two, but they are nonetheless quite striking when they appear. For example, an adult may believe in fixed essences but also have some sense of evolutionary change (Shtulman & Schulz, 2008), or a child may believe that food gets transformed into energy as part of a naïve vitalism (Inagaki & Hatano, 2004), yet also can believe that food goes pretty much intact into various body cavities (Carvalho, Silva, Lima, Coquet, & Clement, 2004; Teixeira, 2000). Adults sometimes seem to make completely incompatible claims, such as believing that human behavior is best explained through a form of strict causal determinism (as opposed to free will), while also believing in moral responsibility (Nahmias, Coates, & Kvaran, 2007). Moreover, the determinists may not always show such incompatible beliefs when situations are made more concrete and emotion laden (Nichols & Knobe, 2007). Thus, they are not even consistent in having inconsistent beliefs.

Finally, people can have two explicit beliefs that are outright contradictions. They can seem to hold two beliefs simultaneously that cannot both be true, not because of some glitch in an inferential chain, but because they are directly at odds with each other. In such cases, people seem to have compartmentalized beliefs so that the conflict is not so obvious. A child might say definitely that all birds fly, yet be equally firm that a penguin is a bird and then say with certainty that penguins cannot fly. In the more logical realm, they are also able to embrace contradictions, declaring that ''P or not p'' is not necessarily true (Morris & Sloutsky, 1998; Osherson & Markman, 1975). In addition, even when students think they deeply comprehend a passage they have just read, they often have overlooked outright contradictions in those passages (Epstein, Glenberg, & Bradley, 1984; Glenberg, Wilkinson, & Epstein, 1982).

In short, intuitive theories, the seeming core of folk science, are a mess. They are full of gaps, and even the more intact fragments contain inconsistencies and contradictions. Moreover, these problems are not simply the follies of youth. Adults can show the same problems of fragmentation as well (di Sessa, 1993; di Sessa, Gillespie, & Esterly, 2004). How can people rely on such knowledge structures to make sense of their world when those structures are so simple as to be nearly devoid of specific content?

One way to address the weak nature of intuitive theories is to describe them as ''framework theories'' (Bang, Medin, & Atran, 2007; Carey, 1999; Wellman & Gelman, 1992, 1997). Wellman and Gelman (1992) compare framework theories to ''paradigms'' in the sense meant by Kuhn (1962) or ''research traditions'' in the sense meant by Laudan (1977). Framework theories are meant to identify large domains, general causal patterns in those domains, and foundational ontological kinds (Wellman & Gelman, 1992). More recently, they have been considered as interpretative systems that may vary across cultural groups, such as seeing the biological world in ecological terms that include humans as opposed to systems that set humans apart (Bang et al., 2007). However they are construed, framework theories alone will not suffice as a basis for a folk science, and indeed they have often been presented as requiring an accompanying specific intuitive theory (Gopnik & Wellman, 1992). Framework theories may achieve more internal coherence just because of their highly skeletal nature, but they seem to be more like constraining orientations than serviceable explanations. Beneath them, adults and children alike attempt more specific theories that are plagued with the problems of gaps, inconsistencies, and contradictions.

1.2. Formal science reconsidered

The limits of the folk sciences might be addressed by assuming that the folk sciences are completely different entities from the formal sciences, and that they have nothing in common except for the label ''science.'' They do not support understanding and are really nothing more than empty shells providing false explanatory comfort, and stand in stark contrast to the richly detailed and analytic formal sciences. The other alternative, which is favored here, is to consider formal science itself and realize that it shares limitations similar to those we just described for folk sciences but, nonetheless, is able to characterize the world in ways that enable more effective interventions and testing of hypotheses. If we assume that the ability to design successful interventions is one hallmark of deeper causal understanding (Woodward, 2003), advances in such areas as medicine and engineering can be seen as evidence that science ''works.''

In the formal sciences, we have begun to learn about the cognitive activities of individual scientists as well as of scientific communities as a whole. At the individual level, scientists do not normally function as deductive nomological machines, carefully setting up laws to act as axioms, the consequences of which are then deductively explored and then tested against experimental data. Although this view is appealing in its elegance and rigor (Hempel & Oppenheim, 1948; Popper, 1959), in practice it is not an effective way for individual scientists, or the field as a whole, to operate.

When cognitive scientists have acted as anthropologists in the laboratory, they have found that scientists' individual theories have many of the characteristics of folk science, albeit at a much more refined level of detail (Dunbar, 1995; Dunbar & Blanchette, 2001). They have gaps in their models of various phenomena, often falling back on schematic gists that gloss over unknown details. They use hunches and guesses frequently instead of using deductions or using a precise inductive routine, such as one honoring strict Bayesian principles. When they do achieve insights or plan new studies, they often use analogies in ways that are far from analytic (Gentner & Kurtz, 2006; Nersessian & Chandrasekharan, 2009). Most importantly, these techniques often work. Major discoveries have occurred in laboratories as cognitive scientists observe their activities (Dunbar, 1995). There may be fewer outright contradictions in scientists' reasoning compared with that of a layperson or of a child, but there can be inconsistencies that are sometimes only discovered quite late as a researcher thinks more deeply about the implications of two lines of reasoning that eventually end in conflict. Indeed, it is the dawning of awareness of such latent contradictions that was said to be the engine driving many scientific revolutions (Carey, 2000, 2009; Kuhn, 1962; Nersessian, 2002; Smith, Carey, & Wiser, 1985).

In short, formal scientists as individuals share many properties with individuals engaging in folk science. They use many heuristics to make a highly complex system more cognitively tractable, and they fall prey to many of the same illusions and biases that occur in folk science. They constantly make vast oversimplifications and may often do so in tacit ways that are well within the capacities of laypeople. They do use mental models but in ways that acknowledge these cognitive limitations both by being heavily simplified idealizations and by often being connected to external physical implementations of the models that greatly amplify their internal cognitive capacities (Nersessian, 2010).

1.3. Science communities

The story is more intricate with communities of scientists rather than lone individuals. Scientists, of course, almost invariably rely on each other in intricate and powerful ways. The idea of a lone wolf making great discoveries may have romantic appeal but is not true in the history of the modern sciences (Dunbar & Blanchette, 2001). A central part of science as a group enterprise involves the two facets of the divisions of labor that distinguish all mature sciences: patterns of deference and understanding the distribution and relations among areas of expertise. Although the empirical studies in this article focus on the emerging ability to sense causal patterns above the level of mechanism and not specifically on folk science communities, the ability to sense rich causal patterns is an essential part of functioning within those communities. In particular, the sensing of such causal patterns is critical to determining plausible and fruitful areas of expertise in a community and to choosing between competing experts.

Scientists routinely defer to colleagues for data, theories, or results that are essential to their own enterprise. They do so with partial, sometimes quite minimal, understandings of their colleagues' areas of expertise. Most people doing cognitive science research with functional magnetic resonance imagery (fMRI) have very modest technical understandings of how signals are processed so as to yield the final results. They do believe, however, that they have good reason to trust those who are experts in the physical and computational aspects of fMRI. Many molecular biologists may have to defer in a massively parallel manner to experts on X-ray crystallography, bioinformatics, bioassay chips, and organic chemistry, among many other areas, in order to conduct their own very focused research project. In turn, all those to whom they defer also have their own networks of deference, which can create a branching pattern of the sort shown in Fig. 1, in which a researcher on cell movement must rely on an ever widening web of experts. (For other examples of collaborations across disciplines, see Nersessian, 2010.)

The setting up of chains of deference draws on sophisticated cognitive processing. One certainly does not defer automatically to any one who makes a claim relevant to one's research. Cognitive neuroscientists, for example, will not generally pay attention to astrologists' claims about how the alignment of heavenly bodies influences the brain, no matter how earnest and persistent the astrologer is. In addition, scientists have to know when there is no need to defer. Some phenomena seem so transparent and self-evident that it is silly and wasteful to defer to ''experts.'' Thus, the cell biologist does not feel a need to defer to an expert on whether a cell is moving.

In other cases, scientists know that a domain does not have enough internal substance to require experts; it wears all its information ''on its sleeve.'' For example, if a dozen people contracted a new strain of influenza and all happened to be wearing striped shirts when they became ill, no one would seek out experts on people who wear striped shirts to understand the disease. However, if all the people who contracted the disease had diabetes, one might

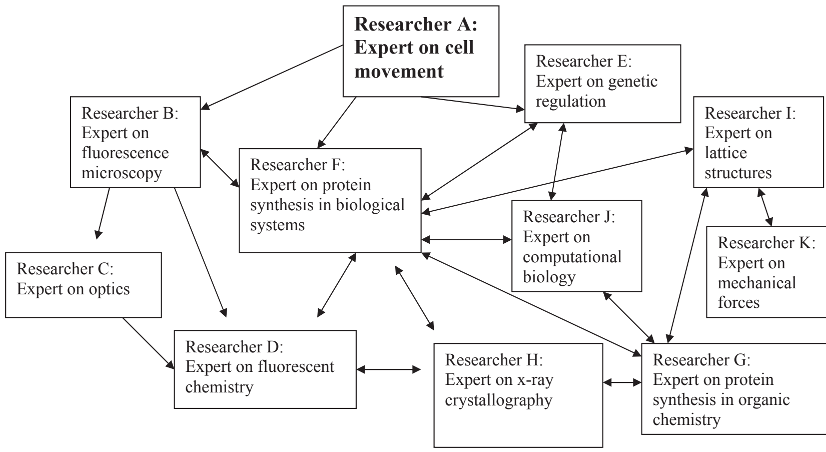

Fig. 1. Patterns of deference in the sciences. Any practicing scientist, or laboratory group, is part of a richly interconnected massive net of deference relations. Here, an expert on cell movement depends critically on experts in topics ranging from microscopy, to gene regulation, to protein synthesis to lattice structures. Each of these experts in turn defers to other experts in both asymmetrical and symmetrical patterns in an ever more branching net of which this figure only shows a small fragment.

consult a diabetes expert to see whether there was a causal pattern that linked diabetics to influenza (in fact there is, see Valdez, Narayan, Geiss, & Engelgau, 1999). Scientists make many decisions along these lines every day, decisions that critically influence how they amplify their own understandings. Even a scientist with a very modest knowledge of influenza would see the potentially greater usefulness of diabetes experts than alleged experts on people who wear striped shirts.

A related ability concerns skill at knowing something about knowledge clusters in areas other than one's own. Practicing molecular biologists, for example, not only need to know how to defer, they also need to know the layout of knowledge communities adjacent to their own domain. At least some of these intuitions seem to be driven by assumptions about the world as being causally heterogeneous (Cartwright, 1999; Keil, Stein, Webb, Billings, & Rozenblit, 2008), namely that there are unique causal patterns associated with various domains and that one can sense such differences without learning many details. For example, the casual patterns and principles associated with biochemistry might be seen as distinct from those associated with evolution. If a domain's principles are those governing physiological pathways and reactions, one might assume that an expert on one particular set of pathways and reactions will have much greater than average insights into other pathways and reactions than, say, an expert on natural selection. We assume that there are strong similarities between the principles governing one pathway and those of another. Such an assumption follows from the idea of local property clusters as governed by a common set of causal relations and principles.

Scientists may also discover legitimate experts by finding or creating common grounds through physical models where they and other experts have aspects of models that ''interlock'' and thereby form a kind of conduit between two areas of expertise (Nersessian, 2010). Thus, in one laboratory, a method was developed to record from multiple neurons in vivo such that the model system of neurons in culture ''interlocks concepts, methods, and materials from biology, chemistry, and electrical engineering'' (p. 19). In many cases, chains of deference may spread out from such model systems.

In summary, there are many social dimensions to science that allow it, as a collective enterprise, to vastly surpass the powers of any one mind. Practicing scientists know how to tap into social networks of knowledge so as to support their own gaps. Formal science works, and is feasible, just because of this aggregate community activity. Just as the division of physical labor added much more efficiency to economic systems (Durkheim, 1893/1997; Smith, 1776/1904), the division of cognitive labor in the sciences has enabled more efficient ways of doing scientific research. However, such divisions of cognitive labor depend on mechanisms of deference, evaluation, collaboration, and even competition. One major part of such mechanisms is having a sense of causal patterns above the level of mechanism and using that sense as a guide to the right areas of expertise.

1.4. Reconsidering folk science

Do any of those shifts in views of formal science have implications for how we should view folk science? We have seen that the social dimension of science has become a central topic in the philosophy of science (Kitcher, 1998, 2001), but it has only recently been seriously considered in the study of the folk sciences. For example, Paul Harris describes how the dominant view of children as junior folk scientists tended to see them as ''stubborn autodidacts,'' in which children supposedly develop understandings all on their own through direct perception and intervention on the world (Harris, 2002). This, of course, cannot possibly be true. Very young children learn a great deal through second or later hand sources (Gelman, 2009). They rely on the testimony of others to learn about dinosaurs, most animals, and all things microscopic. It is in fact striking that some of the first animals many children learn about are large African land mammals they have never seen first hand (D. L. Medin, personal communication).

There are certainly important differences between formal science and folk science concerning how knowledge is acquired, evaluated, stored, and disseminated (see also Carey & Smith, 1993). However, some of the challenges facing formal science are also challenges for folk science, and some of the solutions to these challenges may be similar as well. Thus, although formal science may differ from folk science in many ways reflecting its extraordinary complexity and power, many of the constraints on scientific knowledge are likely to be at least as strong as in the area of folk science. If there are gaps in formal science, there will be much larger gaps in folk science. If there are hunches in science, there will be much wilder guesses in folk science. And if there is a need to defer to others in science, that need may be larger in folk science as well.

Formal science and folk science share a common goal of explaining real-world phenomena and of making predictions in a manner that allows individuals to cope with massive complexity and potential overloads of information. Formal science has coped with this problem by developing a highly elaborated network of communities that distribute information in a manner that can be reliably accessed by others. The quality of that information is regulated by mechanisms that enable scientists to evaluate and trust information outside their own areas of expertise (Hardwig, 1991). Similarly, much of the power of folk science may lie in the ability to access, and rely on, information in other minds. Thus, the feasibility of folk science may rest in the ability of laypeople to leverage their own very fragmentary and partial understandings through methods that sense where appropriate clusters of expertise are situated and sense when a particular phenomenon needs resources beyond those available to a naïve observer.

One of the most useful ways of reconsidering folk science is to ask how it develops. If even young children engage in activities that in some ways resemble those practices that make formal science work, those activities can be taken as indications of how folk science gains some of its own explanatory power, as the activities should be easily available to all people. Moreover, if some abilities develop dramatically during childhood, these can be taken as indications of the ways that folk science might grow as a function of experience and exposure to a culture. The studies described here focus on the emergence of the ability to sense what phenomena would be likely areas for expertise to develop and what ways of thinking about those phenomena would be likely to exemplify real expertise as opposed to other less useful forms of knowledge. Although this is only a small part of the full story of what would make folk science feasible, it does help support the broader account. Consider, therefore, in more detail what these capacities might look like and what we know already about their development in children.

1.4.1. Sensing causal structure and when deference is required

To defer effectively, one needs to have some sense of what domains are feasible as areas of expertise. Explanatory expertise does not naturally build up around all real-world categories. Instead, it dwells only on those that have rich underlying causal structures. If a category coheres because of several causal relations that bind together category properties into a structured set of interrelationships, then it is a legitimate area of expertise. Even if a category has many salient members, it might not be appropriate for expertise unless it has a richly articulated causal structure. Consider, for example, the category of dogs with red collars. It is a large category (red collars are popular) and it may have salient exemplars, but there is nothing else to know beyond the single criterion of having a red collar. By contrast, the category of dogs that are used for hunting has a rich underlying structure. We assume that hunting dogs have an array of traits that make them better at hunting than nonhunters. Such traits can range from those concerning physical robustness and athleticism to a wide variety of behavioral, cognitive, and perceptual traits. There is, in addition, a hierarchical structure with subtypes of hunting dogs, such as terriers, hounds, gun dogs, and even ''cur type dogs.'' There can clearly be experts on hunting dogs, and there are many books and articles representing such an expertise (e.g., Lamb, 2006).

How do we know that it makes sense for there to be experts on hunting dogs but not on dogs with red collars? How do we know this even if we are largely ignorant of the details of what makes good hunting dogs? Many adults may not be able to list anything about hunting dogs vs. nonhunters but still be convinced that the category has legitimate experts. As adults, we at least have an ability to sense where there is causal complexity even when we know almost none of the details of that complexity, and we can do this both for natural kinds (living and not) and for artifacts. We seem to glimpse enough of the causal structure, either as a high-level gloss or in terms of a local fragment, to know whether it would be a reasonable area of expertise. It is less obvious that children are able to sense such differences. Perhaps adults are successful just because they have been taught science in high school and college and have gradually been taught certain science categories and given a sense of the causal complexities underlying those categories. They can ascertain the need for experts for novel categories by assessing the similarity of those novel categories to those that they learned through formal education. If, by contrast, children are driven by surface features and frequency of instances of categories, they may be blind to such complexities. Perhaps a sense of rich causal structure must be explicitly taught.

Children, however, are not acausal creatures merely noticing various contingencies. They appear to sense causal relations and structures from an early age, even in infancy, and use that information to make a wide range of inferences that go beyond correlation and frequency detection (Bullock, Gelman, & Baillargeon, 1982). In fact, they might seem so sophisticated in causal reasoning that their folk sciences would seem to have the fidelity and detail that is so clearly missing in adults. The reality is that they have sophistication and ability in just those ways that do not lead to carefully articulated mechanistic models. Instead, they seem to excel in ways that would be optimal for supporting expertise seeking, deference, and knowing how to leverage expertise in other minds.

Consider some recent studies of what young children and infants can do: They can detect local anomalies, causal ambiguities, and confounds, and then use that information to prompt further exploration (Gweon & Schulz, 2008; Schulz & Bonawitz, 2007), they can sense what kind of domain they are in, and what kinds of causal relations are likely to obtain between what kinds of ontological kinds (Carey, 2009), and they can infer an invisible causal microstructure that is responsible for surface features (Gelman, 2003; Sobel, Yoachim, Gopnik, Meltzoff, & Blumenthal, 2007). They can sense which variables are causally relevant vs. irrelevant to a task (Gopnik & Sobel, 2000; Gopnik et al., 2004), and they can tell whether a causal relation is likely or unlikely (Kushnir & Gopnik, 2005). Even before they can speak they can detect apparently subtle causal relations, such as that intentional agents are more likely causes of events that make a system change from disorder to order (G. Newman, F. C. Keil, V. Kuhlmeir, & K. Wynn, unpublished data), or that intentional agents are more related to nonrandom samples (Xu & Denison, 2009), or that intentional agents are more likely to be causes of motions of inert objects (Saxe, Tenenbaum, & Carey, 2005; Saxe, Tzelnic, & Carey, 2007).

At the same time, they can make striking errors on what animal and machine insides look like (Simons & Keil, 1995), on how microstructural essences are causally related to surface features (Newman & Keil, 2008), and on many aspects of cosmology (Straatemeier, van der Maas, & Jansen, 2008). They can display near-total ignorance of even the simplest mechanistic components of digestion, locomotion, or physical leverage (Inagaki & Hatano, 2002; Siegler, 1976). When asked, for example, about a biological process such as contamination, they may know that some kind of physical stuff mediates effects but be largely clueless on any of the detail (Legare, Wellman, & Gelman, 2009). It seems much easier for children to detect causal relevance (certain kinds of properties are likely to be causally relevant to broad domains) and causal powers (certain kinds of properties tend to produce certain kinds of effects) than it is to detect a particular physical instantiation of mechanism.

In short, children are sensitive to a wide range of causal relations and often surprisingly abstract patterns while at the same time showing even more ignorance about mechanistic details than adults do. This may be just the kind of causal information that is useful for figuring out how and when to defer to others. There are already indications that children can use these sparse patterns to sense similarities between kinds of knowledge and thereby infer the divisions of cognitive labor around them, starting with gross contrasts between the physical, mechanical, and biological areas in preschool (Lutz & Keil, 2002), moving on to more subtle contrasts during the elementary school years (Keil et al., 2008). Knowing how to cluster different phenomena into expertise categories such as biology and mechanics, however, may only scratch the surface of how children leverage their sparse causal understandings through access to others. One other aspect, explored here, asks whether even within one domain, such as biology, children can distinguish between causally rich phenomena or entities that are worthy of expertise and those that are not, and whether, for a given phenomenon, they can distinguish between forms of expertise that are more or less likely to be informative.

Thus, Studies 1 and 2 ask not how children cluster bits of knowledge into domains of expertise, but rather how they decide whether a category of things would constitute a plausible area of expertise. This is seen as a critical part of determining the feasibility of folk science. If children do not eventually master the ability to have some sense of where explanatorily rich phenomena lie, they would have immense difficulty knowing where to devote their energies both at trying to understand a phenomenon themselves and at knowing where there might be relevant experts to use to augment their understandings. Study 3 then focuses on domains that are legitimate areas of both folk science and formal science. Two alleged experts for each domain are described who have different ways of approaching it, one focusing on the sorts of property types and relations that are likely to be causally relevant to that domain and the other focusing on the sorts of property types and relations that are likely to be causally irrelevant. Children are then asked which expert is more likely to know what he or she is talking about.

The studies described here are therefore one small part of a larger story of how folk science might be feasible and useful even as it is so dreadful in the mechanistic details. Children's propensity to infer abstract causal relations and patterns may give them a sense of the expertise domains that exist all around them. They then may buttress this sense with their early-emerging abilities to evaluate the testimony of informants, a topic addressed further in the general discussion.

2. A preliminary study on plausible areas of expertise across domains

To develop a method for studying plausible areas, a preliminary study asked whether children come to distinguish causally empty categories from those having a causal richness worthy of expertise. It examined three different domains, living natural kinds, nonliving natural kinds, and artifacts to see whether the ability to distinguish plausible from implausible areas of expertise varied across domains. These three domains were chosen because they represent one of the largest cuts of distinct causal patterns in the natural and artificial worlds (Keil, 1979). Twenty-five adults and 77 children distributed across grades *K*, 2, and 4 were asked to rate, on a 4-point scale, the extent to which each of 12 phenomena required an expert. Half the phenomena represented causally empty categories (e.g., ''*dogs with red collars*'') and half represented causally rich ones (e.g., ''dogs that are good at hunting''). A second design constraint was that there should be ample members in both plausible and implausible categories so that participants would not distinguish them on the basis that some categories were populated by more instances than others (there are just as many dogs with red collars as there are dogs that are good at hunting). The two living kind pairs referred to ''…dogs that are good at hunting'' vs. ''*…just those dogs with red collars*'' and ''…just those flowers that bloom every year'' vs. ''*…just those flowers planted in purple pots.*'' The two artifact pairs referred to ''…just those cars that have electric (hybrid) engines'' vs. ''*…just those cars with more bugs on the right side than the left side of the windshield*'' and ''…just those books that are bought by teenagers'' vs. ''*…just those books that have green covers.*'' The two nonliving natural kind pairs referred to ''…just those clouds that are in the sky during thunderstorms'' vs. ''*…just those clouds that are shaped like teddy bears*'' and ''…just on forest fires'' vs. ''*…just those fires used to roast marshmallows.*''

In this preliminary study, the youngest children first detected the plausible vs. implausible category difference for artifacts, but not till second grade and older did children also detect the plausible vs. implausible difference for living kinds and nonliving natural kinds (all differences were significant at $p < .01$ in repeated measures ANOVAS and *t* tests with corrections for multiple tests). Fourth graders were close to adult levels at detecting the contrast between the plausible and implausible items. The ability to perceive plausible expertise categories therefore appears to be quite fragile in young children with dramatic improvements in the ability between kindergarten and fourth grade.

There were, however, several limitations to this preliminary study. Some implausible items may have been rejected because they seemed unlikely as opposed to causally empty (e.g., some children felt that it was implausible for clouds to be shaped like teddy bears). In addition, the youngest children might have performed at a higher level with more simplified instructions and with materials that reduced unnecessary processing loads. Finally, although a cross-domain difference was found, with only two items per domain, that pattern was only suggestive.

3. Study 1: Domains and plausible areas of expertise

The preliminary study suggested that the youngest school children have only a faint sense of how to tell plausible expertise domains apart from implausible ones in terms of the density of underlying causal structure. In addition, it suggested that the ability to make such a distinction might first appear for artifacts. Given that kindergartners were able to perform at above chance levels for artifacts, the question arises as to whether more simplified stimuli and procedures might make it possible for higher levels of success in younger children for other categories as well. To correct these limitations, Study 1 screened (through consensus judgments of three adults) all items for plausibility such that the existence of members of a category would not be judged implausible or silly even if the category was not legitimate for expertise. Richer pictorial diagrams were used to make sure that children understood the situations being described. Finally, Study 1 focused on just two domains, with new items and more instances in the artifact and living kinds categories, so as to better assess the generality and robustness of the domain difference found in the preliminary study.

3.1. Method

3.1.1. Participants

Seven male and 19 female university undergraduates participated in this study along with 77 children: 25 kindergartners, 26 second graders, and 26 fourth graders (*M* ages, 5 years 11 months; 7 years 6 months; 9 years 7 months). Adults were run individually with each session lasting approximately 15 min. Children were interviewed individually outside of their classroom with each session lasting approximately 30 min. There were roughly equal numbers of males and females in each child age group. All were from the greater New Haven area and reflected the general demographics of that population (approximately 75% White, 13% African American, 6% Asian, 6% other, with most children being of middle-class backgrounds).

3.1.2. Materials and procedure

3.1.2.1. Stimuli: Eight pairs of categories were designed with one member of each pair representing a causally rich legitimate area of expertise and one representing a causally empty category. Four such pairs were designed for living kinds, and four for artifacts. For each pair, the same subject term was used with different qualifiers indicating the subcategory. The full set of stimuli is shown in Table 1. The order of stimuli was constructed in a quasi-random order such that no two items with the same overall category were closer than three elements apart on the list. Half the participants received the list in a forward order; half received the list in the opposite order. All stimuli were different from those used in the preliminary study.

3.1.2.2. Procedure: Each participant was given all 16 stimuli items in one of the two orders. The stimuli items were preceded by a training protocol that explained the definition of an expert and also provided examples of experts. After training, participants were asked to rate the likelihood that there were experts on given stimuli topics. They used a 4-point rating scale that was depicted visually as a continuum running from 1 to 4 (1 = very unlikely that there are experts on this; 2 = kind of unlikely that there are experts on this; 3 = kind of likely that there are experts on this; 4 = very likely that there are experts on this). The training consisted of explaining to children what an expert was through the following instructions: ‘‘An expert is someone who knows a great deal about something, someone who

Table 1
Stimuli used in Study 1

Living kind pairs
How likely is it that there are people who are experts on just those dogs that can live in very cold places?
**How likely is it that there are people who are experts on just those dogs named Fido?*
How likely is it that there are people who are experts on just those trees that lose their leaves in the winter?
** How likely is it that there are people who are experts on just those trees that were cut down this morning?*
How likely is it that there are people who are experts on just those birds that eat insects?
**How likely is it that there are people who are experts on just those birds that hatched their eggs on weekends?*
How likely is it that there are people who are experts on just those flowers that grow in the desert?
**How likely is it that there are people who are experts on just those flowers that are planted on curvy roads?*
Artifacts
How likely is it that there are people who are experts on just those hammers that are used by miners?
** How likely is it that there are people who are experts on just those hammers that have red handles?*
How likely is it that there are people who are experts on just those cameras that can be used underwater?
** How likely is it that there are people who are experts on just those cameras bought on Thursdays?*
How likely is it that there are people who are experts on just those cars that are made to be driven on rough roads?
** How likely is it that there are people who are experts on just those cars that have dirty windshields?*
How likely is it that there are people who are experts on just those brushes used for painting?
**How likely is it that there are people who are experts on just those brushes used by people with beards?*

Note. An asterisk precedes all causally empty or implausible expertise categories.

understands almost everything about a special area. For example, a doctor is an expert on the human body and a car mechanic is an expert on cars.'' Adults were not trained on the idea of experts. Both groups were then told that they would hear about ''some different areas that it makes sense for people to study and be experts in and others that it does not make any sense.'' They were then given examples during which time they were also taught to use the 4-point scale. Adults were given one example and children were given two examples. Children were given feedback after the first example and then were given the second example, for which they were also given feedback. Adults did the task by reading examples accompanied by drawings depicting the situations. Children were presented the task orally with the same drawings presented by the experimenter. Children were asked frequently at random intervals to justify their responses so as to create an adequate data set of responses for coding.

3.2. Results

As there were no significant effects of order or gender, all data were collapsed across these categories for all age groups. Average scores were then computed for the plausible and implausible scores for living kinds and artifacts, creating four summary scores for each child. The mean overall scores and standard deviations are shown in Table 2.

A 4×4 repeated measures ANOVA was used to analyze the likelihood judgments of categories, with item type (plausible living kind, implausible living kind, plausible artifact, and implausible artifact) as the within-subject factor and age (*K*, 2, 4, Adults) as the

Table 2
Study 1 means (on a 4-point scale) and standard deviations (in parentheses)

	Living Kinds		Artifacts	
Grade	Plausible	Implausible	Plausible	Implausible
K	2.49 (.61)	2.48 (.65)	2.77 (.68)	2.08 (.64)
2	3.12 (.64)	1.66 (.43)	2.68 (.45)	1.51 (.58)
4	3.37 (.43)	1.52 (.45)	3.14 (.38)	1.35 (.40)
A	3.56 (.61)	1.36 (.47)	3.60 (.30)	1.31 (.54)

between-subject factor. Effect size estimates were computed using partial eta squared (η^2). The analyses indicated a main effect of item type, $F(3, 98) = 310.38$, $p < .001$, $\eta^2 = .76$, primarily because participants responded differently as a function of whether items were plausible or implausible. There was also a main effect of grade, $F(3, 98) = 3.44$, $p < .02$, $\eta^2 = .10$, reflecting a somewhat larger drop with age of implausible judgments relative to plausible ones. There was also a significant item type by grade interaction, $F(9, 98) = 27.91$, $p < .001$, $\eta^2 = .46$, indicating a greater contrast between the two implausible and the two plausible categories in older children and adults.

The pattern of developmental change is more easily visualized when difference scores are calculated by subtracting the implausible scores from the plausible scores for each category type (e.g., the plausible dog item minus the implausible dog item). To determine whether there were contrasts between the broad categories of living kinds and artifacts, new mean difference scores were created for each of these two aggregate categories and are shown in Fig. 2.

A 2 × 4 repeated measures ANOVA was used to analyze the likelihood judgment difference scores, with aggregate category type (living kinds, artifacts) as the within-subject factor and age (*K*, 2, 4, Adults) as the between-subject factor. The analyses found no main effect

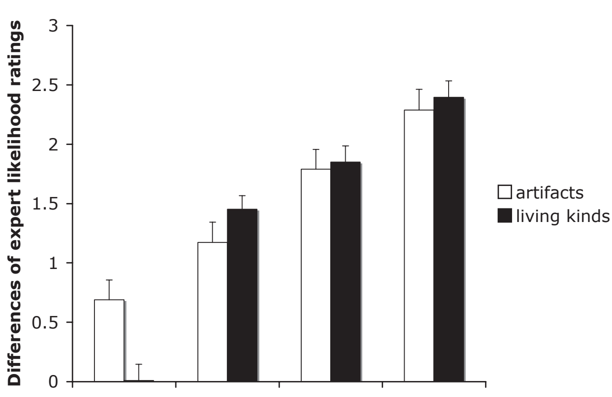

Fig. 2. Ratings of difference scores for likelihood of experts for different domains in Study 1 with standard errors shown. Younger children sensed the plausible vs. implausible domain difference earlier for the aggregate artifact category than living kind category. There was a dramatic rise from kindergarten to second grade in the ability to sense appropriate expert categories for living kinds.

of category type, $F(1,98) = .60$, n.s., $\eta^2 = .006$. There was a main effect of grade, $F(3, 98) = 36.54$, $p < .001$, $\eta^2 = .61$, reflecting larger difference scores in older children and adults. There was also a category type by grade interaction, $F(3, 98) = 7.78$, $p < .001$, $\eta^2 = .19$. In addition, t tests with Bonferroni corrections for multiple tests showed that the difference scores for each age were significantly different from those for all other ages (all at $p < .02$).

When difference scores were compared through t tests to a score of 0 (chance level), all scores but one were significantly different from 0 at the $p < .01$ level (with Bonferroni corrections for multiple t tests). The one nonsignificant case was the living kind difference score for kindergartners. This pattern held for every individual item as well. Thus, separate t tests revealed that each of the four artifact pair differences was significantly different from a population mean of 0 in the kindergarten age group (all ps <.04), while none of the living kind pair differences was significantly different from 0 in the kindergarten age group. All items were significantly different from 0 in the second and fourth graders and adults (all ps <.01). Similarly, all plausible scores showed significant rises across ages and all implausible scores showed significant drops across ages, as revealed by separate one-way ANOVAS (all ps <.001).

3.3. Discussion

Study 1 validated and extended the results of the preliminary study by showing that the youngest school children were unable to see the difference between plausible and implausible domains for living kinds but were well above chance on artifacts. As all the stimuli were new, these results document the robustness of the category difference. By simplifying the stimuli and using clearer graphics, Study 1 seemed to make the task easier for second graders than the preliminary tasks and thus the rise from kindergarten to second grade is more dramatic, but this simplification offered no help to kindergarteners in the realm of living kinds.

The justifications offered by children of different ages reinforced the idea that kindergarteners had more difficulty perceiving the difference between the two domains. For example, one second grader gave the following response to the implausible item for brushes (How likely is it that there are people who are experts on just those brushes used by people with beards?). ''Unlikely—because it does not matter if you have a beard or not a beard because girls do it…because it does not really matter.'' By contrast, kindergarteners, even when successful for artifacts, almost never gave such a reason. Instead, they often said they did not know or referred to an irrelevant relationship such as ''cause people could get messed up or messy.''

For living kinds, the change in justifications was similar to that for artifacts. For example, one second grader answered unlikely to ''How likely is it that there are people who are experts on just those trees that were cut down this morning?'' with the following response: ''Unlikely (kind of unlikely) cause they are just plain old trees.'' Similarly a fourth grader said: ''Unlikely (very) because there is nothing special about trees that were cut down just this morning than trees that were cut down yesterday morning or something.''

Kindergarteners very rarely gave such responses. In fact, of 56 kindergarten responses to implausible categories that were transcribed, only two referred to the arbitrary nature of the implausible category. By contrast, of 88 second grade responses that were coded, 42 referred to the arbitrary nature of the category: a more than 20-fold increase in this kind of justification. Thus, the nascent ability to detect implausible artifact domains in kindergarteners precedes the ability to explain such judgments, an ability that emerges sharply in the second grade and seems to coincide with the onset of the ability to see such contrasts for living kinds.

4. Study 2: Reducing the need to understand expertise

The preliminary study and Study 1 document a strong developmental shift between kindergarten and second grade in the ability to tell plausible expertise domains apart from implausible ones in terms of the density of underlying causal structure. In addition, they both pointed towards the ability to make such a distinction first with artifacts. Yet both studies have levels of complexity that may have created special difficulties for younger children and which may be unrelated to the ability to see implausible and causally empty domains, per se. In particular, both involved using a rating scale, which while certainly feasible with young children and even preschoolers (Mills & Keil, 2005), could pose additional cognitive loads. In addition, although younger children can be taught about experts (Lutz & Keil, 2002), it might reduce cognitive load to not have a training sequence on experts. To that end the procedure was changed in Study 2, but the same stimulus items as in Study 1 were used.

4.1. Method

4.1.1. Participants

Eleven male and 14 female university undergraduates participated in this study along with 75 children: 25 kindergartners, 25 second graders, and 25 fourth graders (*M* ages, 5 years 10 months; 8 years 2 months; 10 years 4 months). Adults were run individually with each session lasting approximately 15 min. Children were interviewed individually outside of their classroom with each session lasting approximately 30 min. There were roughly equal numbers of males and females in each child age group. All were from the greater New Haven area and reflected the general demographics of that population (approximately 75% White, 13% African American, 6% Asian, 6% other, with most children being of middle-class backgrounds).

4.1.2. Materials and procedure

4.1.2.1. Stimuli: The full set of stimuli are the same as those used in Study 1 and two new animals were added, insects and fish, so as to create four animals and two plant items to ensure that the living kind category was more fully sampled. The order of stimuli was constructed in a quasi-random order such that no two items with the same overall category were closer than two elements apart on the list. Half the participants received the list in a forward order; half received the list in the opposite order. The primary difference with Study 1 was that the stimuli were now presented as pairs and a child was asked which one of the two people in the pair they would like to ask to learn the most important things about the category and to be helped the most.

4.1.2.2. Procedure: Each participant was given all 10 stimuli items in one of the two orders. The training indicated that participants would be asked which of two people they should ask to get more useful information about a category. (i.e., ''When there is more than one person who can help you, you have to choose the person who you think would know the most important things and help you the most.'') They were then given two training examples. Two sets of items were created such that gender of consultants was matched to gender of the child.

By pitting the domains explicitly against each other and focusing on using two people as alternative expertise resources, this design asked whether such simplifications could make implausible domains more apparent to the kindergarten age group that had such difficulties in the preliminary study and Study 1.

4.2. Results

Each choice was scored as either a 0 (picking the implausible domain person) or a 1 (picking the plausible domain person). These scores were then added together and divided by the number of items in each category (four artifact pairs and six living kind pairs consisting of four animals and two plants). A 2×4 repeated measures ANOVA was used to analyze the likelihood judgments of categories, with category type as the within-subject factor and age (*K*, 2, 4, Adults) as the between-subject factor. Effect size estimates were computed using partial η^2. There was a main effect of category type ($F[1, 102] = 4.11$, $p < .05$, $\eta^2 = .039$). The grade by category type interaction was not significant ($F[3, 102] = 1.45$, $\eta^2 = .041$). There was a main effect for grade ($F[3, 102] = 6.70$, $p < .001$, $\eta^2 = .165$).

As seen in Fig. 3, participants tended to do better on the artifacts, especially the kindergarteners. Two-tailed *t* tests against a mean of .50 revealed that kindergartners were above chance levels for artifacts ($t = 3.39$, $p < .003$, d. f. = 24) and for living kinds ($t = 2.32$, $p < .03$, d.f. = 24). If a Bonferroni correction is made for multiple *t* tests, the kindergarteners remain well above chance levels on the artifacts and are not significantly above chance levels for living kinds.

4.3. Discussion

Taken together, these results fit closely with those of the preliminary study and Study 1. Despite making the task easier in several ways, the kindergartners showed only modest improvement for living kind judgments, passing chance levels at only a trend level and not at the customary .05 level when corrections for multiple *t* tests are included. In addition, although somewhat reduced in magnitude, there was still an early advantage for artifacts. When kindergarteners encounter categories, they have considerably more difficulty than second graders in seeing where true causal complexity and richness lies.

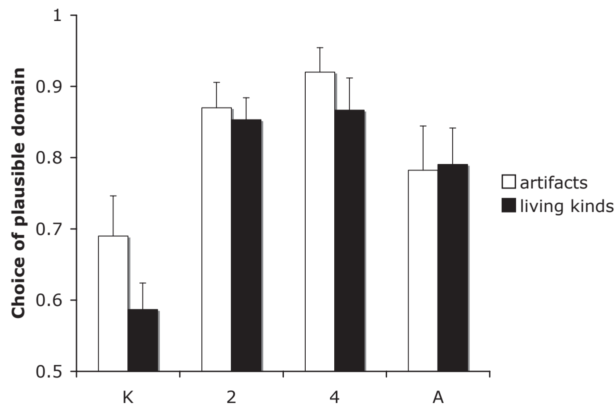

Fig. 3. Results from Study 2 in which children were asked to pick which of two people would be a more useful expert on the same category. Shown here are the average scores for all artifacts and for all living kinds {on a scale ranging from .5 (representing picking at random [50%] between the plausible and implausible domains) to a 1.0 (representing 100% choice of the plausible domain)}, with standard errors shown. As in the preliminary study and Study 1, kindergarteners were more able to discern the plausible domains for artifacts than for living kinds.

The artifact advantage may be linked more strongly to relatively complex artifacts with many parts. Across all three studies, the kindergarteners tended to perform particularly well on artifacts with many discernable moving parts, such as cameras and cars. Simple artifacts, such as brushes, were more difficult to judge. It may be easier early on to see causal complexity when there are several discrete parts with clearly different functions. Causal complexity may be more difficult to see in both simple artifacts and living kinds. For simple artifacts, the complexity lies in the ergonomics of its use, which may in turn require considerable awareness of how a simple artifact is embedded in cultural practices. For living kinds, parts and their functions may be more difficult to directly discern as well. It may also be that adults talk more about the causal relations and parts of complex artifacts than they do for living kinds and that these comments help children to see more causal complexity early on for artifacts. This does not seem likely for such items as cameras, but should be examined in future studies. Notions of causal density could also be explored by converging measures to better understand their nature. For example, children and adults alike might be able to list more distinctive features of the causally dense categories and especially features with causal connections. Similarly, the artifact/living kind difference might be explored by seeing whether adults and children list more part like properties, or perhaps more internal ones, when describing complex artifacts than when describing living kinds.

5. Study 3: A different sense of appropriate experts

Study 1 and the preliminary study consistently demonstrate that kindergarteners have considerable difficulty evaluating the plausibility of an expertise domain on the basis of causal density and structure. Although other studies do document an ability of even preschoolers to reason about experts, the possibility remains that younger children might generally have a much harder time thinking about the plausibility of someone being an expert and that a problem with thinking about expert plausibility is causing drops in performance rather than a problem in seeing causal density and structure. Study 2 addressed this issue by using a task that did not explicitly ask about who was the expert. Another manipulation check would be to use a different kind of task that did ask about plausible experts and on which kindergarteners were able to do quite well. Study 3 therefore used a different criterion from causal density, namely causal relevance, to see whether younger children might be able to choose between plausible and implausible experts.

Prior work has shown that kindergarteners are clearly sensitive to a dimension of causal structure known as ''causal relevance,'' namely knowing that certain types of properties tend to be causally important in specific domains (Keil, Smith, Simons, & Levin, 1998), a phenomenon that is well documented in adults (Medin & Shoben, 1988) and even in other primates, who weigh different property types as more central for tools vs. foods (Santos, Hauser, & Spelke, 2002). Adults and children alike tend to see color, surface markings, and precise number of parts as less causally relevant to understanding artifacts than shape, size, and fragility. For living kinds, these properties tend to have more equal levels of importance and often color, surface markings, and precise number of parts can outweigh overall shape, size, and fragility. Thus, a hot pink, striped hammer with a two-piece handle and a head is still judged a hammer if its overall shape, strength, and size are within certain bounds. However, a dog-like mammal that is naturally hot pink, striped, and naturally has eight legs, is a much more problematic dog. Causal relevance intuitions seem more basic than intuitions about the causal density and structure. One might not have much of a sense at all of the causal complexity that underlies the basis of being a particular species while still believing that color, number of parts, and surface patterns tend to matter more for biological species than for tools.

Study 3 capitalized on this ability to make causal relevancy judgments by asking children which of two alleged experts were more plausible as a function of domains. If even kindergarteners were able to make contrasting judgments about expertise for living kinds and artifacts by considering the kinds of properties an alleged expert emphasized, then we can see that the difficulties demonstrated by kindergarteners in the preliminary study and Study 1 are not simply reflecting problems thinking about experts. As reviewed earlier, even preschool children do seem to be able to quickly sense, even in novel artificial systems, what classes or types of interventions are likely to be linked to what properties. That ability may make them effective here in the ability to evaluate experts.

5.1. Method

5.1.1. Participants

Ninety-one children: 30 kindergartners, 32 second graders, and 29 fourth graders (*M* ages, 5 years 9 months; 7 years 6 months; 9 years 11 months) participated as subjects. All children were interviewed individually outside of their classroom with each session lasting approximately 30 min. There were roughly equal numbers of males and females in each child age group. All were from the greater New Haven area and reflected the general demographics of that population (approximately 75% White, 13% African American, 6% Asian, 6% other, with most children being of middle-class backgrounds).

5.1.2. Materials and procedure

5.1.2.1. Stimuli: Eight unfamiliar words were created. For half of the participants, four of these names were said to be tool names and four were said to be animal names. Children were simply told that two people each claimed to be an expert on a new kind of tool/animal. The label was then given and each ''expert'' was described as saying that certain properties were critical to understanding what kind of tool/animal it was. One ''expert'' then said that a particular color, surface marking pattern, and number of parts was most important while another said it was most important to know about size, fragility, and overall shape. Drawings were used to show particular colors, shapes, sizes, surface, markings, number of parts, or fragility in a manner that was abstracted away from any object in particular (e.g., a color patch or a rough outline). The order of stimuli was randomized. Half the participants received the list in a forward order; half received the list in the opposite order.

5.1.2.2. Procedure: Each participant was given all eight stimuli items in one of the two orders. The training indicated that the child would hear about two people who claimed to know a lot about a new kind of tool or a new kind of animal. They were then told that only one person ''really knew what they were talking about'' while the other person was ''sort of confused.'' They were given one practice example from a category unlike any in the test set, namely an expert on ''ice-cream sundaes.'' They were then given the eight stimulus pairs and asked to choose who really knew what they were talking about. The training protocol and the stimuli are shown in the Appendix.

5.2. Results

Each choice was scored as either a 0 (picking the person who emphasized domain inconsistent properties) or a 1 (picking the person who emphasized domain consistent properties). These scores were then added together and divided by the number of items in each category (four tool pairs and four animal pairs). A 2×3 repeated measures ANOVA was used to analyze which person was chosen, with category type as the within-subject factor and age (K, 2, 4) as the between-subject factor. Effect size estimates were computed using partial η^2. There was a main effect of category type ($F[1, 88] = 30.318$, $p < .001$, $\eta^2 = .256$). The grade by category type interaction was not significant ($F[2, 91] = 1.01$, $\eta^2 = .022$). There was a main effect of grade ($F[2, 88] = 7.27$, $p < .001$, $\eta^2 = .142$).

As seen in Fig. 4, even kindergarteners showed different response patterns for tools and animals. A t test of just the kindergarten age group showed that their responses for tools and animals were significantly different ($t = 2.38$, $p < .025$, d.f. = 29). The other two age groups showed more dramatic differences significant at the $p < .001$ level.

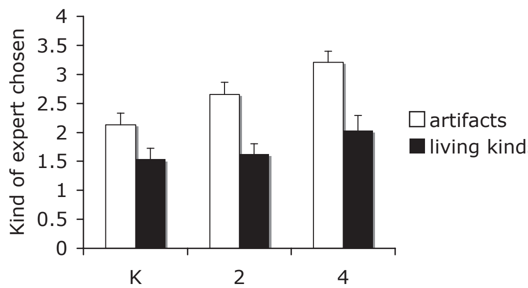

Fig. 4. Judgments of quality of experts as a function of property type by domain interactions, with standard errors shown. A value of 0 favors people who talked about color, surface markings, and precise number of parts. A value of 4 favors experts who talked about shape, size, and fragility.

5.3. Discussion

The results are compatible with an account arguing that, when causal relevancy is considered, even kindergarteners can sense which kinds of causal types should be stressed by an expert in a domain. For artifact judgments, the children at all ages favored an expert who stressed shape, size, and fragility and not color, surface patterns, or precise number of parts. For living kind judgments, all properties were more equally weighted. It could be argued, however, that the children were simply responding at chance levels for the living kinds rather than systematically favoring an even distribution of property types. This alternative would have to be examined in follow-up studies. Most importantly, when using a very different criterion from causal complexity, namely causal relevancy, even kindergarteners were using different criteria to judge experts in the domains of artifacts vs. living kinds.

In short, when children are asked to tap into earlier emerging intuitions about causal relevancy even kindergarteners show an ability to sense that for completely novel artifacts and animals, an expert who emphasizes shape, size, and fragility as opposed to color, surface marking, and number of parts is more plausible when doing so for artifacts than for living kinds. The ability seems to be present earlier on than intuitions about causal complexity. One limitation of this study is that there was a tendency for the ''correct'' experts on some of the artifact items to have more words in their reasons than the ''incorrect'' experts, which could have been used as cues to expertise. However, it is also the case that children did not pick the expert more often for those artifact items where the expert said many more words than for the case where the numbers of words were more closely matched (''nillard''); and there was no tendency on the living kind items for the children to pick the ''expert'' that used more words. It therefore seems unlikely that the total number of words used by the experts was the only factor driving judgments by younger children.

6. General discussion

If the feasibility of folk science rests on the extent to which individuals have fully detailed mechanistic models of the world around them, there is not much hope. If, however, folk science is more like the formal sciences in the extent to which individuals powerfully leverage their understandings by developing ways of relying on knowledge in other minds, then it may well be feasible. This is not to diminish the powerful cognitive processes and structures in the mind of each individual doing folk science. It is just that those processes and structures are quite different from intricate, fully articulated theories. Instead, there appears to be a diverse set of ways in which we build on the crude mechanistic fragments in our heads. Moreover, these ways of going beyond such fragments appear early in development, with some being present in kindergarteners and others rapidly emerging by the second grade. Given the rapid rise in the ability to detect where causal complexity lies, it could be argued that folk science is much more plausible in second graders than it is in kindergarteners, although a firm conclusion along those lines must await further studies asking whether other task variations might improve performance in younger children.

To be able to augment our imperfect mechanistic understandings, we need several cognitive components. First, we must sense causal and relational patterns in the world in concise, abstract ways that allow us to know how to link various phenomena to appropriate groups of experts or domains of expertise. This process would normally work at a more fine-grained level than the very weak constraints provided by ''framework theories,'' presumably by looking at patterns that are particularly diagnostic for particular domains. Second, we must have ways of evaluating the causal intricacy of phenomena in a domain without being swamped by the details. Third, we must have some heuristics for evaluating alleged experts, some of which rely on hunches about what kinds of properties and relations matter in a specific domain, and some of which rely on past behaviors or inferred motives of an expert (Harris, 2007; Mills & Keil, 2005, 2008). Finally, we must have a sense of the knowledge formation process itself so as to know what kinds of knowledge are probably learnable on one's own and what kinds almost surely require testimony by others given a particular environment. This wide range of capacities may not be an elaborate product of higher education, but instead may be manifested in ways in which all of us make sense of our world from an early age.

The studies presented here offer a glimpse into one part of the overall system that supports folk science: the ability to sense causally rich and fertile domains for a science vs. causally sterile ones. It is ironic, in light of older theories of children being hopelessly concrete, that the concrete details are often where children most need to defer to others. Most adults freely use words to refer with confidence to things that they cannot identify directly (such as the distinctions between beeches vs. elms, silver vs. platinum, and ferrets vs. weasels) because they have a sense of who does know about these things and of the chain that could connect them to that person. Similarly, young children seem to have a sense of how to find supporting details. This sense is often built on knowledge of more abstract patterns that are unique to specific domains and that can serve as ways of linking a phenomenon to an appropriate group of experts. Importantly, children who sense that ''dogs that hunt'' is a much more plausible category than ''dogs with red collars'' might nonetheless be unable to give any specific details about hunting dogs. Much of this competency also seems implicit in nature. Children often had compelling intuitions about plausible areas of expertise or plausible kinds of expert explanations, but then could provide almost no information about why. They sense relatively abstract concise patterns for domains in ways that allow them to make such judgments while knowing hardly any details.

What role do metacognitive skills play in the development of the ability to defer appropriately? It is well established that children's metacognitive understanding develops considerably in the school years (Kuhn, 2000; Schneider, 2008). Children become better at understanding the limits of their own cognitive processes and knowledge bases and at using strategies to cope with their limitations (Schneider, 2008). Their epistemological understanding also develops as they show marked increases in the extent to which they can appreciate that different people can hold equally valid but different ways of understanding a set of phenomena and that errors can be an intrinsic part of a science (Kuhn, 2009; Masnick & Klahr, 2003). They slowly come to appreciate that people can be biased in their interpretations without being aware of their bias (Mills & Keil, 2005, 2008; Pillow & Henrichon, 1996). Although many of these skills do show strong developmental changes during the same period when children come to see which areas of expertise are more implausible, the influence of metacognitive factors on these tasks is unclear. Metacognitive factors may be more involved in influencing the ability to evaluate one's own degree of understanding and when one needs to defer. By contrast, the main developmental change found in Studies 1–3 may be a gradually increasing appreciation of the causal complexity underlying some categories, especially natural kinds. Thus, decisions on when to defer will be more closely linked to metacognitive abilities than will decisions on whom to defer to.

Several developmental puzzles remain to be explained. One question concerns the role of the detection of local anomalies, inconsistencies, and causal ambiguities. We have seen that even adults tolerate considerable amounts of contradictions in larger belief systems that seem to be a somewhat motley collection of conceptual fragments with lots of theoretical holes and gaps. Yet inconsistencies and contradictions are also often said to be ''the engines that drive conceptual change'' in young children (Carey, 2009). These may primarily occur as local juxtapositions, perhaps often only transiently during a serendipitous coming together of two clashing ideas in explicit thought. Conceptual change may then be triggered in the more abstract schematic skeletal framework as well, perhaps through a heuristic that if a part has a contradiction or if some threshold of several contradictions is exceeded, there may be a problem with the framework as a whole.

What are the broader implications for folk science? In actual practice folk science often involves a ''gluing'' together of fragments as a local situation suggests, buttressed by notions of relevant supporting experts. For example, if adults are asked to explain how the body turns food into movement, they might have never before tried to come up with such an explanation and may have very little in terms of a preconceived systematic idea beyond the very broad constraints laid down by a ''framework theory.'' But as the question is posed, they start to assemble various relevant fragments (e.g., chewing, waste, things they have heard about nutrition, some ideas about muscles, etc.) and try to piece them together. That process, and the beliefs that drive it, are powerfully helped by hunches about where they are on thin ice vs. where there is a strong foundation of supporting knowledge, even if they do not know it themselves. This creation of explanations ''on the fly'' is now receiving more attention in the modeling approaches as well (e.g., Klenk, Friedman, & Forbus, 2008).

How do all the elements that make up a folk science fit together into a system that has some efficacy for its users? The suggestion here is that the array of elements may be far larger than it might appear at first and that such an array is needed to support the highly impoverished nature of local specific intuitive ''theories.'' Children and adults alike seem to have very broad and extremely schematic orientations towards domains and a sense of causal patternings that both suggest where expertise is needed and point a person towards the likely expert cluster. They also have abilities to evaluate sources of information as well as to reflect on their own ways of acquiring knowledge. Given the presence of all these abilities in young children and given how similar they are to many of the key features that make formal science succeed, it seems likely that such abilities are a foundational part of what makes folk science feasible.

Folk science in technological cultures may also carry within it notions of formal science that are absent in more traditional cultures. Thus, the average layperson in North America or Europe may have well-entrenched ideas about the discipline known as ''science'' itself. This may not make a difference in terms of how chains of deference are set up, but it is certainly a topic of interest for future studies. Perhaps in traditional cultures there are notions about legitimate areas of expertise and about experts as well, but those areas of expertise are organized in ways that reflect that culture's specific epistemological orientations.

We also know that individual scientists constantly make simplifying assumptions so as to prune back a causal briar patch into something more elegant and cognitively manageable. They seem to do so in several ways. First, they tend to stick to one level of a reductionist hierarchy, not diving down deeper to lower levels (Owens, 1989; Wilson & Keil, 1998). For example, some psychologists may try to envision how cognitive structures and processes are constrained by the biology of the nervous system, but they rarely if ever try to fully specify a mechanistic model of what is happening at the cellular level. Similarly, while attending to chemistry, the biologist also stops short of the sorts of details that occupy chemists. Thus, few biologists ask how quantum bond angles are relevant to an aspect of cellular metabolism. There is a skill here that is not well understood, partly because we do not have a reliable way of specifying levels of reduction and whether there really are clearly objective levels that apply in the same way across all the subspecialities of such major areas as psychology, biology, chemistry, and physics. Somehow, scientists in practice draw these lines all the time so as to make their task more manageable and they seem to do so in a way that allows their science to advance. Folk science must certainly also engage in a somewhat similar process, perhaps acknowledging the possibility of cross-level influences but trying to focus on one level at a time. Indeed, when two distinct levels are present, as in some views of disease causation in both biological and supernatural terms, children and adults alike often have the two levels work at different timescales so as not to mix them together (Legare & Gelman, 2008).

Scientists also make a science ''local'' by ruling out those tendrils of causal influence that, while technically present, can be ignored for the purposes of the scientific task at hand. Without such simplifications, one runs the risk of having to consider the ways in which almost any event can potentially have a causal influence on any other that follows later: a scientific version of the ''butterfly effect'' problem (Hilborn, 2004). Elga (2007) considers this problem in detail and suggests that scientists, and laypeople, see the world as localized nets of causal relations. Those nets form clumps and clusters that can be considered as ''stand-alone'' systems in which more remote influences are ruled out. When trying to understand the behavior of billiard balls on a table, we tend to consider just the forces at play between objects on the table, even though there are very small but real influences exerted by the moon, people in the room, and countless other objects and events such as sound waves. We may technically rule out effects that would have to travel faster than the speed of light to have an effect (things outside what physicists call the ''light cone;'' Elga, 2007); but there is still a vast array of other real influences that are simply below some threshold in almost all matters of scientific inquiry. Folk scientists engage in such local analyses with a vengeance, but the heuristics they use to prune away causal links to other domains have not been studied. Even in the heights of analytical science, the processes through which scientists zero in on certain causal processes and relations as most relevant is described as more of an art than an analytically driven procedure (Strevens, 2009).

A second facet of deferring involves heuristics for establishing trust in others. Two commonly used heuristics rely on prior accuracy and familiarity with an informant (Corriveau & Harris, 2009). Quite young children seem to be able to employ both of these heuristics, suggesting that both are also central parts of folk science. Thus, children are capable of judging that a person who is mistaken is likely to be mistaken again in the future, and they also use various cues of affiliation to judge testimony, such as whether they are familiar with a person or whether a potential informant tends not to dissent in general (Corriveau, Fusaro, & Harris, 2009). A tendency to discount dissenters may not always be the best strategy, but it is an easy heuristic to use.

Any account that discusses the social dimensions of scientific understanding must also at least address the vast literature on the social construction of science, the idea that cultures and groups within cultures can create lenses that distort how phenomena are viewed and understood. In the extreme, such approaches argue that there is no scientific ''truth'' and that all is socially constructed, often to serve political and economic agendas. The approach taken here instead assumes (following Boghossian, 2006; Kitcher, 1998, 2001) that, while complete fabrications and massive distortions are possible, most of science does not work that way. Scientific knowledge, of either the formal or folk variety, is not simply a relative matter. Science refers to real patterns in the world, usually of a causal nature, and is thereby constrained by them.

Kitcher (2001) suggests that the process is much like the making of maps for an area. Depending on one's goals and perspective, one might come up with a very different map of the same environment than another person would. Kitcher describes how the subway map of London distorts real distances and spatial layout in dramatic ways but is still a ''true'' map with respect to representing the ordered relations between stops and intersections in an efficient and clear way. A different map for walking routes might have a radically different appearance and set of spatial relations but would also be true. Sciences can each construct their own maps that leave out some informational dimensions and distort others; but they are rarely complete social constructions. In the same manner, we might consider folk science as constructing maps that are driven by real-world causal and relational patterns even as they simplify and distort information in various ways. Moreover, different cultural groups can construct very different maps that, while all true, result in sharply contrasting reasoning patterns (Atran & Medin, 2008).

A full account of folk science requires at least the following: It must specify the kinds of relevant knowledge in the minds of individuals. In particular, this means a specification of typical knowledge fragments and the levels of abstraction at which they are usually represented. It also must specify the ways in which people use that partial knowledge in situations when they are trying to explain a phenomenon, including the ways they stitch together fragments on the fly to create a more coherent description, guided by the constraints imposed a framework theory. A comprehensive account also requires a specification of the ways in which people evaluate their own knowledge. However, beyond these descriptions of what each person knows about how the world works and their understanding of their degrees of ignorance, another critical dimension is a specification of how individuals support their own impoverished knowledge by learning to outsource their understandings to others.

Just as users of a natural language confidently talk about natural kinds such as platinum and silver, even though they may have no individual means of telling them apart (Malt, 1990; Putnam, 1975), so also do our folk science explanations rely extensively on our senses of legitimate areas of expertise and how to defer to those areas. If even young children are well on their way to navigating such areas of expertise, that ability is one major form of support for the feasibility of folk science. Thus, if one has a sense of when one's own shaky understanding is likely to be resting on firm ground in other minds, one can then use that understanding more confidently to make other inferences. For example, even if I cannot tell silver from platinum, if I am confident that through various chains of deference experts would affirm that something was silver, I can feel more entitled to make other inferences about that thing by virtue of its being silver. The same sort of process can work in folk science where, for example, even if we do not know much about the molecular details that are responsible for the surface properties of a particular plant, we can make other inferences about that plant based on the secure knowledge there are experts who could explain such a causal pathway.

In summary, the finding that laypeople of all ages have extremely limited understandings at the level of mechanistic clockworks diagrams of how the world works does not mean that they are not deeply interested in causal patterns or in seeking out explanations, activities that are very much a part of folk science. Laypeople are in fact immensely sophisticated at tracking certain kinds of abstract causal patterns and using that information to figure out how to better ground their incomplete understandings; and this is an ability so basic that its roots can be found before children start schooling and as surging in influence by the time they reach the second grade. Folk science may be an intrinsically social process, but one that starts off that way even before children have assimilated most of the practices of their culture.

We may attribute much of the social infrastructure of formal science to explicitly taught social institutions that are acquired quite late, when in fact basic features of deference and expertise seeking are natural ways that children approach the process of understanding. Just as some anthropologists avoid the study of children because of a misleading assumption that all the interesting aspects of cultures are only gradually learned from long-term immersion in those cultures (Hirschfeld, 2002), it may be equally misleading to assume that all the interesting facets of the social embedding of both the formal and folk sciences are learned through long-term assimilation of cultural practices. Some of that embedding may be part of how children approach their very first explorations of how the world works.

Acknowledgments

The research described in this article was supported by NIH grant no. R37HD023922 to Frank C. Keil. Many thanks to Eric Smith for help in all aspects of running the studies described in this article. Thanks as well to Kara Gaughen and Jonathan Kominsky for help in preparing the manuscript and to Kristi Lockhart and three reviewers for comments on drafts of the manuscript.

References

Atran, S., & Medin, D. L. (2008). *The native mind and the cultural construction of nature*. Boston, MA: MIT Press.

Au, T. K. (1994). Developing an intuitive understanding of substance kinds. *Cognitive Psychology*, *27*, 71–111.

Bang, M., Medin, D., & Atran, S. (2007). Cultural mosaics and mental models of nature. *Proceedings of the National Academy of Sciences*, *104*, 13868–13874.

Bertamini, M., Spooner, A., & Hecht, H. (2004). The representation of naive knowledge about physics. In G. Malcolm (Ed.), *Multidisciplinary approaches to visual representations and interpretations* (pp. 32–43). Amsterdam: Elsevier.

Boghossian, P. (2006). *Fear of knowledge: Against relativism and constructivism*. Oxford, England: Oxford University Press.

Braine, M. D. S. (1978). On the relation between the natural logic of reasoning and standard logic. *Psychological Review*, *85*, 1–21.

Bullock, M., Gelman, R., & Baillargeon, R. (1982). The development of causal reasoning. In W. J. Friedman (Ed.), *The developmental psychology of time* (pp. 209–254). New York: Academic Press.

Caramazza, A., McCloskey, M., & Green, B. (1981). Naive beliefs in ''sophisticated'' subjects: Misconceptions about trajectories of objects. *Cognition*, *9*, 117–123.

Carey, S. (1985). *Conceptual change in childhood*. Cambridge, MA: Bradford Books, MIT Press.

Carey, S. (1988). Conceptual differences between children and adults. *Mind and Language*, *3*, 167–181.

Carey, S. (1999). Sources of conceptual change. In E. K. Scholnick, K. Nelson, S. A. Gelman, & P. Miller (Eds.), *Conceptual development*: *Piaget's leqacy* (pp. 293–326).

Carey, S. (2000). Science education as conceptual change. *Journal of Applied Developmental Psychology*, *21*, 13–19.

Carey, S. (2009). *The origin of concepts*. New York: Oxford university Press.

Carey, S., & Smith, C. (1993). On understanding the nature of scientific knowledge. *Educational Psychologist*, *28*, 235–251.

Carey, S., & Spelke, E. (1996). Science and core knowledge. *Philosophy of Science*, *63*, 515–533.

Cartwright, N. (1999). *The dappled world: A study of the boundaries of science*. London: Cambridge University Press.

Carvalho, G. S., Silva, R., Lima, N., Coquet, E., & Clement, P. (2004). Portuguese primary school children's conceptions about digestion: Identification of learning obstacles. *International Journal of Science Education*, *26*, 1111–1130.

Corriveau, K. H., & Harris, P. L. (2009). Choosing your informant: weighing familiarity and past accuracy. *Developmental Science*, *12*, 426–437.

Corriveau, K. H., Fusaro, M., & Harris, P. L. (2009). Going with the flow: Preschoolers perfer non-dissenters as in formants. *Psychological Science*, *20*, 372–377.

Dunbar, K. (1995). How scientists formally reason: Scientific reasoning in real-world laboratories. In R. Sternberg & J. Davidson (Eds.), *The nature of insight* (pp. 365–396). Cambridge, MA: MIT Press.

Dunbar, K., & Blanchette, I. (2001). The in vivo/in vitro approach to cognition: The case of analogy. *Trends in Cognitive Sciences*, *5*, 334–339.

Durkheim, E. (1997). *The division of labor in society* (G. Simpson, Trans.). New York: The Free Press (Original work published 1893).

Elga, A. (2007). Isolation and folk physics. In H. Price & R. Corry (Eds.), *Causation, physics, and the constitution of reality: Russell's Republic revisited* (pp. 106–119). Oxford, England: Oxford University Press.

Epstein, W., Glenberg, A. M., & Bradley, M. M. (1984). Coactivation and comprehension: Contribution of text variables to the illusion of knowing. *Memory & Cognition*, *12*, 355–360.

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

Gelman, S. A. (2009). Learning from others: Children's construction of concepts. *Annual Review of Psychology*, *60*, 115–140.

Gentner, D., & Kurtz, K. (2006). Relations, objects, and the composition of analogies. *Cognitive Science*, *30*, 609–642.

Glenberg, A. M., Wilkinson, A. C., & Epstein, W. (1982). The illusion of knowing: Failure in the self-assessment of comprehension. *Memory & Cognition*, *10*, 597–602.

Gopnik, A., Glymour, C., Sobel, D., Schulz, L., Kushnir, T., & Danks, D. (2004). A theory of causal learning in children: Causal maps and Bayes nets. *Psychological Review*, *111*, 1–31.

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

Gopnik, A., & Wellman, H. M. (1992). Why the child's theory of mind really is a theory. *Mind and Language*, *7*, 145–171.

Gweon, H., & Schulz, L. (2008). Stretching to learn: Ambiguous evidence and variability in preschoolers' exploratory play. *Proceedings of the 30th annual meeting of the Cognitive Science Society*.

Hardwig, J. (1991). The role of trust in knowledge. *Journal of Philosophy*, *88*, 693–708.

Harris, P. L. (2002). What do children learn from testimony? In P. Carruthers, S. P. Stich, & M. Siegal (Eds.), *The cognitive basis of science* (pp. 316–334). London: Cambridge University Press.

Harris, P. L. (2007). Trust. *Developmental Science*, *10*, 135–138.

Hempel, C. G., & Oppenheim, P. (1948). Studies in the logic of explanation. *Philosophy of Science*, *15*, 135–175.

Heyman, G. D., Phillips, A. T., & Gelman, S. A. (2003). Children's reasoning about physics within and across ontological kinds. *Cognition*, *89*, 43–61.

Hilborn, R. C. (2004). Sea gulls, butterflies, and grasshoppers: A brief history of the butterfly effect in nonlinear dynamics. *American Journal of Physics*, *72*, 425–427.

Hirschfeld, L. (2002). Why don't anthropologists like children? *American Anthropologist*, *104*, 611–627.

Inagaki, K., & Hatano, G. (2002). *Young children's naive thinking about the biological world*. New York: Psychological Press.

Inagaki, K., & Hatano, G. (2004). Vitalistic causality in young children's naive biology. *Trends in Cognitive Sciences*, *8*, 356–362.

Keil, F. C. (1979). *Semantic and conceptual development: An ontological perspective*. Cambridge, MA: Harvard University Press.

Keil, F. C., Smith, W. C., Simons, D. J., & Levin, D. T. (1998). Two dogmas of conceptual empiricism: Implications for hybrid models of the structure of knowledge. *Cognition*, *65*, 137–165.

Keil, F. C., Stein, C., Webb, L., Billings, V. D., & Rozenblit, L. (2008). Discerning the division of cognitive labor: An emerging understanding of how knowledge is clustered in other minds. *Cognitive Science*, *32*, 259–300.

Kitcher, P. (1998). A plea for science studies. In N. Koertge (Ed.), *A house built on sand: Exposing postmodernist myths about science* (pp. 32–52). New York: Oxford University Press.

Kitcher, P. (2001). *Science, truth and democracy*. Oxford, England: Oxford University Press.

Klenk, M., Friedman, S., & Forbus, K. (2008). Learning modeling abstractions via generalization. *The Proceedings of the 22nd International Workshop on Qualitative Reasoning*, Boulder, CO.

Kuhn, T. S. (1962). *The structure of scientific revolutions* (1st ed.). Chicago, IL: University of Chicago Press.

Kushnir, T., & Gopnik, A. (2005). Young children infer causal strength from probabilities and interventions, *Psychological Science*, *16*, 678–683.

Kuhn, D. (2000). Metacognitive development. *Current Directions in Psychological Science*, *9*, 178–181.

Kuhn, D. (2009). The importance of learning about knowing: Creating a foundation for development of intellectual values. *Child Development Perspectivies*, *3*, 112–117.

Lakshminaryanan, V., Chen, M. K., & Santos, L. R. (2008). Endowment effect in capuchin monkeys (*Cebus apella*). *Philosophical Transactions of the Royal Society B: Biological Sciences*, *363*, 3837–3844.

Lamb, V. (2006). *The ultimate hunting dog reference book: A comprehensive guide to more than 60 sporting breeds*. Guilford, CT: The Lyons Press.

Laudan, L. (1977). *Progress and its problems*. Berkeley, CA: University of California Press.

Lawson, R. L. (2006). The science of cycology: Failures to understand how everyday objects work. *Memory & Cognition*, *34*, 1667–1675.

Legare, C. H., & Gelman, S. A. (2008). Bewitchment, biology, or both: The co-existence of natural and supernatural explanatory frameworks across development. *Cognitive Science*, *32*, 607–642.

Leqare, C. H., Wellman, H. M., & Gelman, S. A. (2009). Evidence for an explanation advantage in native biological reasoning. *Congnitive Psychology*, *58*, 177–194.

Leslie, A. (1994). ToMM, ToBY and agency: Core architecture and domain specificity. In L. A. Hirschfeld & S. A. Gelman (Eds.), *Mapping the mind: Domain specificity in cognition and culture* (pp. 119–148). Cambridge, UK: Cambridge University Press.

Leslie, A. M., Friedman, O., & German, T. P. (2004). Core mechanisms in 'theory of mind.' *Trends in Cognitive Sciences*, *8*, 528–533.

Lutz, D. R., & Keil, F. C. (2002). Early understanding of the division of cognitive labor. *Child Development*, *73*, 1073–1084.

Malle, B. F., Knobe, J. M., & Nelson, S. E. (2007). Actor–observer asymmetries in explanations of behavior: New answers to an old question. *Journal of Personality and Social Psychology*, *93*, 491–514.

Malt, B. C. (1990). Features and beliefs in the mental representation of categories. *Journal of Memory and Language*, *29*, 289–315.

Masnick, A. M., & Klahr, D. (2003). Error matters: An initial exploration of elementary school children's understanding of experimental error. *Journal of Cognition and Development*, *4*, 67–98.

Medin, D. L., & Shoben, E. J. (1988). Context and structure in conceptual combination. *Cognitive Psychology*, *20*, 158–190.

Mills, C. M., & Keil, F. C. (2005). The development of cynicism. *Psychological Science*, *16*, 385–390.

Mills, C. M., & Keil, F. C. (2008). Children's developing notions of (im)partiality. *Cognition*, *107*, 528–551.

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

Nahmias, E., Coates, J., & Kvaran, T. (2007). Free will, moral responsibility, and mechanism: Experiments on folk intuitions. *Midwest Studies in Philosophy*, *31*, 214–242.

Nersessian, N. J. (2002). Kuhn, conceptual change, and cognitive science. In T. Kuhn & T. Nichols (Eds.), *Contemporary philosophers in focus series* (pp. 178–211). Cambridge, MA: Cambridge University Press.

Nersessian, N. J. (2010). How do engineering scientists think? Model-based simulation in biomedical engineering laboratories, *Topics in Cognitive Science*, *1*, 730–757.

Nersessian, N. J., & Chandrasekharan, S. (2009). Hybrid analogies in conceptual innovation in science. *Cognitive Systems Research Journal, Special Issue: Integrating Cognitive Abilities*, *10*, 178–188.

Newman, G., & Keil, F. C. (2008). ''Where's the essence?'': Developmental shifts in children's beliefs about the nature of essential features. *Child Development*, *79*, 1344–1356.

Nichols, S., & Knobe, J. (2007). Moral responsibility and determinism: The cognitive science of folk intuitions. *Nous*, *41*, 663–685.

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

Owens, D. (1989). Levels of explanation. *Mind*, *98*, 59–79.

Pillow, B. H., & Henrichon, A. J. (1996). There's more to the picture than meets the eye: Young children's difficulty understanding biased interpretation. *Child Development*, *67*, 802–819.

Popper, K. (1959). *The logic of scientific discovery*. London: Hutchinson.

Proffitt, D. R. (1999). Naive physics. In R. A. Wilson & F. C. Keil (Eds.), *The MIT encyclopedia of the cognitive sciences* (pp. 577–579). Cambridge, MA: MIT Press.

Psillos, S. (1999). *Scientific realism: How science tracks truth*. London: Routledge.

Psillos, S. (2009). *Knowing the structure of nature*. London: Palgrave/MacMillan.

Putnam, H. (1975). The meaning of ''meaning''. In K. Gunderson (Ed.), *Language, mind, and knowledge* (Vol. 2, pp. 131–193). Minneapolis, MN: University of Minnesota Press.

Rips, L. J. (2002). Circular reasoning. *Cognitive Science*, *26*, 767–795.

Rozenblit, L. R., & Keil, F. C. (2002). The misunderstood limits of folk science: An illusion of explanatory depth. *Cognitive Science*, *26*, 521–562.

Santos, L. R., Hauser, M. D., & Spelke, E. S. (2002). Domain-specific knowledge in human children and non-human primates: Artifact and food kinds. In M. Bekoff, C. Allen, & G. Burghardt (Eds.), *The cognitive animal* (pp. 205–216). Cambridge, MA: MIT Press.

Saxe, R., Tenenbaum, J. B., & Carey, S. (2005). Secret agents: Inferences about hidden causes by 10- and 12-month-old infants. *Psychological Science*, *16*, 995–1001.

Saxe, R., Tzelnic, T., & Carey, S. (2007). Knowing who dunnit: Infants identify the causal agent in an unseen causal interaction. *Developmental Psychology*, *43*, 149–158.

Schneider, W. (2008). The development of metacognitive knowledge in children and adolescents: Major trends and implications for education. *Mind, Brain and Education*, *2*, 114–121.

Schulz, L., & Bonawitz, E. B. (2007). Serious Fun: Preschoolers play more when evidence is confounded. *Developmental Psychology*, *43*(4), 1045–1050.

di Sessa, A. (1993). Towards an epistemology of physics. *Cognition and Instruction*, *10*, 105–225.

di Sessa, A., Gillespie, N., & Esterly, J. (2004). Coherence versus fragmentation in the development of the concept of force. *Cognitive Science*, *28*, 843–900.

Shtulman, A. (2006). Qualitative differences between naive and scientific theories of evolution. *Cognitive Psychology*, *52*, 170–194.

Shtulman, A., & Schulz, L. (2008). The relation between essentialist beliefs and evolutionary reasoning. *Cognitive Science*, *32*, 1049–1062.

Siegal, M., Butterworth, G., & Newcombe, P. A. (2004). Culture and children's cosmology. *Developmental Science*, *7*, 308–324.

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

Simons, D. J., & Keil, F. C. (1995). An abstract to concrete shift in the development of biological thought: The insides story. *Cognition*, *56*, 129–163.

Slaughter, V., & Gopnik, A. (1996). Conceptual coherence in the child's theory of mind: Training children to understand belief. *Child Development*, *67*, 2967–2988.

Smith, A. (1904). *An inquiry into the nature and causes of the wealth of nations* (5th ed.). London: Methuen & Co. (Original work published 1776).

Smith, C., Carey, S., & Wiser, M. (1985). On differentiation: A case study of the development of the concepts of size, weight, and density. *Cognition*, *21*, 177–237.

Sobel, D. M., Yoachim, C. M., Gopnik, A., Meltzoff, A. N., & Blumenthal, E. J. (2007). The blicket within: Preschoolers' inferences about insides and causes. *Journal of Cognition and Development*, *8*, 159–182.

Straatemeier, M., van der Maas, H., & Jansen, B. (2008). Children's knowledge of the earth: A new methodological and statistical approach. *Journal of Experimental Child Psychology*, *100*, 276–296.

Strevens, M. (2009). *Depth: An account of scientific explanation*. Cambridge, MA: Harvard University Press.

Teixeira, F. M. (2000). What happens to the food we eat? Children's conceptions of the structure and function of the digestive system. *International Journal of Science Education*, *22*, 507–520.

Tversky, A., & Kahneman, D. (1983). Extension versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review*, *90*, 293–315.

Valdez, R., Narayan, K. M., Geiss, L. S., & Engelgau, M. M. (1999). Impact of diabetes mellitus on mortality associated with pneumonia and influenza among non-Hispanic black and white US adults. *American Journal of Public Health*, *89*, 1715–1721.

Vosniadou, S. (2001). On the nature of naive physics. In M. Limon & L. Mason (Eds.), *Reframing the processes of conceptual change* (pp. 61–76). Dordrecht, The Netherlands: Kluwer Academic Publishers.

Weisberg, D. S., Keil, F. C., Goodstein, J., Rawson, E., & Gray, J. (2008). The seductive allure of neuroscience explanations. *Journal of Cognitive Neuroscience*, *20*, 470–477.

Wellman, H. M. (1990). *The child's theory of mind*. Cambridge, MA: Bradford Books/MIT Press.

Wellman, H. M., & Gelman, S. A. (1992). Cognitive development: Foundational theories of core domains. *Annual Review of Psychology*, *43*, 337–375.

Wellman, H. M., & Gelman, S. A. (1997). Knowledge acquisition in foundational domains. In D. Kuhn & R. S. Siegler (Eds.), *Handbook of child psychology*, Vol. 2 (pp. 523–573). New York: Wiley.

Wilson, R. A., & Keil, F. C. (1998). The shadows and shallows of explanation. *Minds and Machines*, *8*, 137–159.

Woodward, J. (2003). *Making things happen: A theory of causal explanation*. Oxford, England: Oxford University Press.

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

Zimmerman, C. (2007). The development of scientific thinking skills in elementary and middle school. *Developmental Review*, *27*, 172–223.

Appendix: Training and stimuli for Study 3

Training

Introduction

Today we are going to talk about a bunch of new things and we are going to play a kind of detective game. I am going to tell you about two different people. Now, each of these people think they really know a lot about a something, but only one of them really knows what they are talking about; the other one is a little confused. That is where I need you to help me out. I need you to be the detective and figure out which one really knows more. How does that sound?

Training

Ok. Now we are going to talk about a lot of things today, so I have got some pictures here to help you remember what we are talking about. I bet you have seen one of these before (show picture of glove). This is an example of an object that we might be talking about today; we are going to talk about some animals and some tools and some other stuff. Now each of these two people are going to tell you three different things about the object. To help you remember those, I have got these pictures here (lay out pictures above glove). Each of these boxes tells you something about the object.

This one here (point to red square) tells you about the color, and I bet you can figure out what this one (point to polka dot square) tells you. Right, it tells you that the glove has spots on it. Good job.

Now this one here (point to number) tells you about the number of parts. See, the glove has five places to put your fingers. This one here (point to apple) tells you just about how big the object is. The glove is about the size of an apple; that makes sense right, as you can hold an apple in your hand.

This one (point to hand outline) tells you about the shape. This tells you that it is shaped like a hand. Which makes sense as we wear gloves on our hands, right. Now this last one (point to vase) is a little silly. This one tells us about how breakable or fragile the object is. You can see here that the vase has fallen off the table and cracked; so this tells us that the glove is probably a little fragile and we should be careful.

So each one of these squares tells you something about the object we are going to be talking about. Ok? All right, now I think you have got the hang of this. Why do not we try one to get warmed up. I bet you have had an ice cream sundae before. So both of these guys (place pictures of people in front of child) really think they know all about ice cream sundaes, but really only one of them does. Now, this first guy said that to really know all about ice cream sundaes, you have to know two things, he said you have to know that they are made out of ice cream and that they have hot fudge and whipped cream on top. But this other guy, he said no way! He said to really know all about ice cream sundaes you have to know that they come in a green bowl and that you have to eat them with a metal spoon. Now which one of these guys do you think really knows more about ice cream sundaes? Good job. This is exactly what we are going to do with the rest of the things we are going to do today. Does that sound Ok? Great, let us get started!!

Stimuli

Before we get started, I just want to let you know that there are no right or wrong answers here, ok. I just want you to tell me who *you* think really knows more.

1. The first thing we are going to talk about today are called calipers. Now, calipers are a special kind of tool that carpenters (people who build houses and furniture) use.

This first guy, he said to really know all about calipers, you have to know that:

- They have lines on them, just like this
- They are blue
- They have two parts on top

The second guy said no way. He said to really know all about calipers you have to know that:

- They are shaped kind of like an ‘F’
- They have to be very strong so they do not break when you use them
- They have to be a size so that you can hold them in your hand

Which one of these two guys do you think ***really*** knows more about calipers?

2. The next thing we are going to talk about today is called a katsura. A katsura is a kind of tree.

   This first lady, she said that to really know all about katsura trees, you have to know that:

   - Their bark is brown
   - Each leaf has four points on it
   - Their bark has wavy lines in it

   This other lady said no way. She said to really know all about katsura trees you have to know that:

   - They are as tall as a house with three floors
   - They are shaped kind of like a stalk of broccoli
   - Their branches are very strong so they do not break off

   Which one of these two ladies do you think ***really*** knows more about katsura trees?

3. The next thing we are going to talk about is called a nillard. A nillard is a type of scientific equipment they use in the doctors office.

   This first guy said to really know all about nillards you have to know that:

   - Their knobs are held on very tightly
   - They are about as big as a lunch box
   - They are shaped kind of like a box with a bulge in front for the screen

   The other guy said no way. He said to know all about nillards you have to know that:

- They have six parts that all work together inside
- They have stripes on the outside
- They are gray

Which one of these two guys do you think ***really*** knows more about nillards?

4. The next thing we are going to talk about is called a verbena. A verbena is a kind of flower. They are very pretty.

   Now this first lady said to really know all about verbena flowers you have to know that:

   - They are pink
   - Their petals have lines coming out of the middle like this
   - They have one row of seeds on the inside

   This other lady said no way. She said that to really know all about verbena flowers you have to know that:

   - They are about the size of a little girls hand
   - Their petals are held on tightly so they do not blow away in the wind
   - They are shaped kind of like a lollipop

   Which one of these two ladies do you think ***really*** knows more about verbena flowers?

5. The next thing we are going to talk about is called a rubra. A rubra is a kind of insect.

   This first guy said to really know all about rubras you have to know that:

   - They have six legs
   - They have stripes on their legs
   - They are green

   The other guy said that to really know all about rubras you have to know that:

   - They are about the size of a quarter
   - They are shaped kind of like a pear
   - Their legs are very strong

   Which one of these two guys do you think ***really*** knows more about rubras?

6. This next thing we are going to talk about is called a serval. A serval is a kind of animal.

   The first lady said to really know all about servals you have to know that:

- They have five tufts of hair on their head
- They have spots on their backs
- Their fur is brown

The other lady said no way. She said to really know all about servals you have to know that:

- They are about the size of a microwave
- Their fur is held on really tightly
- They are shaped kind of like a football

Which one of these two ladies do you think ***really*** knows more about servals?

7. This next thing is called a skidder. A skidder is a kind of big machine they use on farms.

   Now, this first lady said to really know all about skidders you have to know that:

   - They are yellow
   - That they have four wheels
   - That they have brown spots on them

   But the other lady said to really know all about skidders you have to know that:

   - They are as long as a whale
   - They are shaped kind of like a box with wheels
   - That they have to be very strong so they do not break when you use them

   Which one of these two ladies do you think ***really*** knows more about skidders?

8. The last thing we are going to talk about is called a julick. A julick is a kind of furniture. It is kind of like a chair.

   Now this first guy said to really know all about julicks you have to know that:

   - They are brown
   - They have a crisscross on their seat
   - They have 12 bars on the back

   But the other guy said to really know all about julicks you have to know that:

   - They are shaped kind of like an ‘H’ with a seat
   - They are about as big as a big mailbox
   - They have to be very strong so they do not break when you sit on them

   Which one of these two guys do you think ***really*** knows more about julicks?