Published in final edited form as:
J Exp Psychol Gen. 2010 February ; 139(1): 1–15. doi:10.1037/a0018319.

A Bump on a Bump? Emerging Intuitions Concerning the Relative Difficulty of the Sciences

Frank C. Keil, **Kristi L. Lockhart**, and **Esther Schlegel**
Department of Psychology, Yale University

Abstract

In 4 studies, the authors examined how intuitions about the relative difficulties of the sciences develop. In Study 1, familiar everyday phenomena in physics, chemistry, biology, psychology, and economics were pretested in adults, so as to be equally difficult to explain. When participants in kindergarten, Grades 2, 4, 6, and 8, and college were asked to rate the difficulty of understanding these phenomena, children revealed a strong bias to see natural science phenomena as more difficult than those in psychology. The perceived relative difficulty of economics dropped dramatically in late childhood. In Study 2, children saw neuroscience phenomena as much more difficult than cognitive psychology phenomena, which were seen as more difficult than social psychology phenomena, even though all phenomena were again equated for difficulty in adults. In Study 3, we explored the basis for these results in intuitions about common knowledge and firsthand experience. Study 4 showed that the intuitions about the differences between the disciplines were based on intuitions about difficulty of understanding and not on the basis of more general intuitions about the feasibility or truth of the phenomena in question. Taken together, in the studies, the authors find an early emerging basis for judgments that some sciences are intrinsically more difficult than others, a bias that may persevere in adults in subtler forms in such settings as the courtroom.

Keywords

cognition; intuitive theories; knowledge evaluation; cognitive development

> Physics investigates the essential nature of the world, and biology describes a local bump. Psychology, human psychology, describes a bump on the bump. (Quine, 1981)

Are the sciences unequal in their "scientificness"? Is there a hierarchy with physics at the highest levels and the social sciences at the lowest (e.g., Jefferson, 2000)? These questions have been related to claims that all sciences are ultimately reducible to physics. Nonphysicists are sometimes said to suffer from "physics envy" (Cohen, 1971), in which they envy physics as the most pure, or "real," science. One dimension promoting this envy may be the perceived difficulty of understanding phenomena in the various disciplines. If a phenomenon is seen as relatively trivial to explain, then its science may also seem unimpressive, as might its practitioners. We explore here how views of the relative difficulty of the sciences develop.

Correspondence concerning this article should be addressed to Frank C. Keil, Department of Psychology, Yale University, 2 Hillhouse Avenue, New Haven, CT 06520. frank.keil@yale.edu.

It is difficult to document, in adults at least, whether views of the relative difficulty of the sciences are rational. Perhaps problems in the natural sciences are in general more difficult to fully explain than problems in the social sciences, and this view is appropriate. Alternatively, people may use heuristics to estimate the difficulty and complexity of problems, heuristics that bias judgments more for some sciences than others. Deciding between these two possibilities is difficult in adults because there are no absolute standards of difficulty or complexity (Sober, 1975). How is one to judge that a full explanation of human sentence processing will be more or less difficult than a full explanation of super-conductivity? We therefore do not see any objective way to set up complexity scales for phenomena across the natural and social sciences and to use those as a basis for evaluating the accuracy of difficulty judgments. Instead, we ask here whether developmental changes in intuitions might be used as a basis for revealing biases.

Consider how a developmental approach might reveal biases that cannot be uncovered by methods involving looking solely at adults. If children and adults alike use heuristics to evaluate how difficult it would be to understand various phenomena, the influence of those heuristics might well be stronger in children because they are lacking experiences that lead them to moderate the effects of the heuristics. They simply have not had enough time to see real complexities revealed and to use those complexities to infer the level of difficulty involved in understanding those phenomena. For example, one heuristic for evaluating the difficulty of understanding a phenomenon may refer to the degree to which the phenomenon seems to be immediately experienced and seems to be controllable. Thus, many psychological phenomena are experienced directly and may seem to be strongly influenced by willful action. One can, for example, influence one's ability to remember phone numbers by rehearsing them, and one can immediately experience the effect. Phenomena in other domains, such as the healing of one's skin after a laceration, may not be as likely to be either immediately experienced or changeable by will. Even when one can easily initiate an action in physical mechanics, such as dropping a ball, one may then recognize that one has little influence over the path and velocity of the ball once set in motion. Nonpsychological phenomena may seem to have a life of their own in ways that psychological ones do not. Perceptions of controllability may also be related to intuitions about predictability. For example, if whining reliably produces a certain effect in one's parent, one may feel that one has strong predictive skills in that area as well.

Another heuristic might be based on the greater availability of simple (even if inaccurate) theories in some domains than in others. Thus, if a young child's folk psychology is based on a simple belief– desire model (Wellman & Liu, 2004), but that child does not have readily available a comparably easy or apparently coherent theory about physical mechanics or biology, the child might assume that all phenomena in psychology could be explained in terms of that simple theory and therefore see them as intrinsically simpler. This contrast is supported by arguments that intuitive theories in areas such as physics may lack coherence and may instead be a complex array of fragments (di Sessa, 1993). These two heuristics may also be related. Easily available and simplified theories of mind may exaggerate the extent to which a child views beliefs and desires as being able to influence psychological phenomena. This possibility is further supported by suggestions that children might be especially prone to interpret the behaviors of others as acts of free will as opposed to being deterministic (Bloom, 2006; Nichols, 2004). If most psychological phenomena are seen as controllable and immediate, they might fit nicely into a simple belief– desire theory that makes them seem easy to understand.

Over time, however, as children gain more experience trying to understand phenomena in domains and as their domain theories become further differentiated, they may start to see layers of complexity that will moderate the effects of these simple heuristics. If the heuristic is based on a sense of immediacy and control, the child might start to see how immediacy and control unpack in much more complex ways and how a sense of control may often be illusory. Similarly, as children's theories of mind seem to gradually unfold in ever more complex forms in ways that hold across cultures (Wellman, Fang, Liu, Zhu, & Liu, 2006), they may come to see that ever more complex models are needed to account for psychological behaviors. In other domains, children's early theories may not be so salient and well organized and therefore may not appear to be as simple early on or show such a regular progression of increasing complexity.

An additional heuristic that might influence judgments concerns the level of mathematics that is seen as required for proficiency in a field (Blinder, 1999; Cohen, 1971). If a child sees a discipline as more intrinsically linked to numbers and calculations, that discipline may seem more complex. We suspect important developmental changes here as children gain more insight into other areas in which mathematics is involved and also start to appreciate more clearly the different levels of complexity in mathematics itself.

We assume here that heuristics along these lines are early emerging and exert strong domain effects on difficulty ratings in young children. Regardless of the details of the heuristics involved, in the studies in this article, we ask whether differences in difficulty ratings might be revealed in their strongest forms in children who have not yet been exposed to the actual complexity of explanations in the various disciplines and, therefore, are more purely at the mercy of early biases and heuristics. We further suggest that vestiges of these biases may persevere in adults and reveal themselves in certain contexts such as judgments about the qualifications of expert witnesses.

Evaluating Knowledge and Ability

If people use heuristics and biases to judge the relative difficulty of fields of intellectual endeavor, young children might also do so, given that many biases and heuristics have been shown to be present in children, often to a stronger degree than in adults (Bernstein, Atance, Loftus, & Meltzoff, 2004; Jacobs & Klaczynski, 2002). Because young children can evaluate knowledge in other minds and can refer to the causal structure of the world to guide their judgments (Keil, Stein, Webb, Billings, & Rozenblit, 2008), they are likely to have intuitions about the relative complexity of the disciplines. In addition, because young children are able to evaluate the relative difficulty of performing various tasks in ways that seem to take into account their intrinsic complexity (Kinlaw & Kurtz-Costes, 2003; Nicholls, 1980), it is reasonable to assume that they might also be able to rate the difficulty of understanding phenomena in ways that reflect intuitions about their complexity.

Although there are no prior studies in which domain differences in complexity or difficulty ratings by children were examined, there are related lines of work on both adults and children that suggest why systematic differences might be found. In adults, for example, the correlations between estimates of competence and actual performance tend to be relatively high for abilities that are public and relatively easy to measure objectively. In contrast, for abilities for which actual performance is more private and difficult to measure objectively, adults tend to overestimate their ability (Dunning, Heath, & Suls, 2004; Mabe & West, 1982). Thus, people tend to be more accurate about their athletic abilities than about their emotional intelligence. People might therefore also estimate that phenomena related to their own immediate and internal mental experiences, namely those of psychology, are easier to understand than are those related to more objectively observable events, such as those of physics. They may be confronted less often with clear evidence that they do not understand a social science phenomenon than with clear evidence that they do not understand a natural science one.

In a related vein, adults tend to give higher estimates of values of traits that they think are controllable (Alicke, 1985). Although controllability has been largely looked at in terms of personality traits, it may extend to a much broader range of attributes (Dunning et al., 2004). With respect to physics versus psychology, it is certainly reasonable that adults and children alike might think they have more control over psychological phenomena. That sense of control in turn may lead them to feel that they understand controlled domains more fully.

Finally, when adults and adolescents study passages in various domains, they tend to think they have more easily grasped complex information in some domains than others, even when later tests show that they have not. In particular, junior-high, high-school, and college students all show a tendency to think they have understood psychology textbooklike passages better than they really have, while being less confident in their comprehension of biology, chemistry, and physics passages (Zimmerman, Gerson, Monroe, & Kearney, 2007). There is also a tendency to become less and less confident about the understanding of chemistry and physics passages with increasing age. Thus, there is some reason to believe that a bias to see psychology as relatively easy compared with other disciplines might be early emerging. In the studies that follow, we ask whether such biases might be particularly robust in young children who have had little or no exposure to the idea of formal academic disciplines as such.

Developmentally, younger children certainly do make more errors in estimating their competencies. They fail, for example, to discriminate between memory problems that are relatively hard versus relatively easy (Dufresne & Kobasigawa, 1989). More generally, they tend to overestimate their performance on cognitive tasks (Schneider & Pressley, 1997), as well as overestimating their own explanatory understanding (Mills & Keil, 2004). This overconfidence is also found in their beliefs about their future abilities (Lockhart, Chang, & Story, 2002). Moreover, their overconfidence is particularly strong for psychological traits (Lockhart, Nakashima, Inagaki, & Keil, 2008). Younger children also believe that they have greater control over psychological traits (Lockhart et al., 2008; Kalish, 2002), a finding that resonates with suggestions mentioned earlier that they emphasize free will more than determinism in behavioral explanations. One reason for developmental improvements in the quality of estimates is thought to be an increased ability to engage in metacognitive assessments of their own mental states (Schneider & Pressley, 1997).

Vestiges of the Biases in Adults

There seem to be certain circumstances in which a discounting of the difficulty of psychology relative to the natural sciences continues to occur in adults. In the legal profession, it appears that there are domain differences in the extent to which expert witnesses are challenged on the grounds that their alleged area of expertise is largely common sense. For example, in an influential 1975 case in England, a court ruled that experts in psychology were generally inadmissible because their supposed expertise was in fact common knowledge (Colman & Mackay, 1993, 1995; Mackay, Colman, & Thornton, 1999). Similarly, in the United States legal system, there have been decades of opposition, often by judges, to the idea that psychology experts know anything beyond that which is already obvious and beyond that which random jurors can learn from everyday experience (Loftus, 1986).

A recent U.S. case involving perjury charges against I. Lewis “Scooter” Libby illustrates the pattern vividly. Judge Reginald Walton disallowed the use of psychology experts on memory as relevant, arguing that jurors had no need of memory experts because they could rely on their common sense understanding of how memory works. By contrast, in the natural sciences, such common knowledge challenges are almost never made (Mackay et al., 1999).

Even for cases involving experts on complex psychological findings that are clearly unknown or counterintuitive to laypeople, judges have repeatedly ruled them to be not really legitimate areas of expertise. This pattern, which continues in many legal systems today, suggests powerful biases to assume that some areas of science are intrinsically easier than are others and are able to be understood without access to experts.

A different example of the bias in adults may be seen in how people regard the appeal of various forms of explanations. In particular, although adults find it quite easy to tell a vacuous, wholly psychological explanation from an informative one, they have much more difficulty telling the two kinds of explanation apart when the same explanations contain additional irrelevant biological information, such as descriptions of uninformative functional magnetic resonance imaging scans (Weisberg, Keil, Goodstein, Rawson, & Gray, 2008). Thus, the addition of new information at a lower level of explanation in the reductionist hierarchy can result in illusions of insight, even when that information has no real explanatory value. It may be that the addition of biological information to a psychological explanation causes a rise in perceived complexity that in turn makes the explanation seem more profound or deep.

Taken together, the adult findings of domain differences and the developmental data showing greater errors in self-assessment point toward the following developmental hypotheses. Given that younger children may have a greater sense of immediacy and control over psychological and social phenomena, and given that they have less insight into mental states and their complexity, they might be expected to especially underestimate the difficulty of understanding in social science areas relative to natural science ones. In addition, given that children's theories of mind follow a lawful progression of increasing complexity that may not be so orderly in other domains, children might especially underestimate the complexity of psychological phenomena early on because they have highly available, simple, and relatively coherent theories. More precisely, one can ask whether some phenomena further down in the reductionist chain are considered more complicated than are those higher up, roughly creating the continuum physics > chemistry > biology > psychology > economics. The reductionist ordering is based on the idea that a higher science in the chain, such as biology, can be explained in terms of the next lower one, chemistry, which in turn can be explained in terms of physics (Anderson, 1972). This hierarchical ordering of the disciplines has been repeatedly proposed on a wide range of other grounds as well, such as citation patterns and rated disciplinary “hardness” and “softness” (Simonton, 2009). The bias need not always favor physics, it could also favor biology over psychology, or it could be largely an effect of discounting the difficulty of psychology. Moreover, it may be that some domains along this continuum all have roughly the same level of perceived difficulty, such as all the natural sciences. Our most conservative prediction is therefore that difficulty biases should follow the reductionist continuum but allow for adjacent domains to be seen as the same in difficulty.

We also posited that economics might be a special case in which young children believe that economics involves counting and arithmetic, and because that skill is both more objective and difficult for them, they might think that the area in general is difficult. Older children, however, who can easily count large numbers and do simple arithmetic and who do not yet have any sense of the complex mathematics associated with academic economics, may start to discount its difficulty in a manner similar to other social sciences such as psychology.

To explore the developmental hypotheses, we created problems in each of the disciplines that adults rated as equal in difficulty. We then asked whether the tendency to discount the difficulty of some sciences relative to others is stronger in young children. We predicted such an effect and moreover predicted that economics, because of its perceived use of numbers, might also be seen as especially difficult by young children, despite its remoteness in the reductionist chain from physics. In Study 1, we explore the origins of a bias across all the disciplines. In Study 2, we narrow the focus to a contrast that explores subfields within psychology, the area that is seen as easiest by children. In Study 3, we examine in more detail the separate contributions of intuitions about ease of learning and common knowledge to domain differences. Finally, in Study 4, we ask whether the effects might arise from much broader cognitive biases about domains and not really from difficulty judgments per se.

Study 1

Method

Participants—Twenty-two kindergartners (mean age = 5 years, 9 months; range = 5 years, 4 months to 6 years, 1 month; 13 females, 9 males), 26 second graders (mean age = 7 years, 8 months; range = 7 years, 2 months to 8 years, 4 months; 18 females, 8 males), and 26 fourth graders (mean age = 9 years, 10 months; range = 9 years, 5 months to 10 years, 4 months; 13 females, 13 males) participated. In addition, 33 sixth graders (mean age = 11 years, 6 months; range = 11 years, 0 months to 12 years, 0 months; 22 females, 11 males), 23 eighth graders (mean age = 13 years, 4 months; range = 13 years, 0 months to 14 years, 0 months; 14 females, 9 males), and 28 adults (mean age = 20 years, 7 months; range = 18 years, 0 months to 23 years, 0 months; 16 females, 12 males) participated. All the children were from the greater New Haven, Connecticut, area and reflected the general demographics of that population (approximately 75% White, 13% African American, 6% Asian, and 6% other, with most children being of middle-class backgrounds). The adults were all full-time students at a highly selective university and were either paid for their participation or served as participants as part of a course requirement.

Materials—There were six items in each of five categories: physics, chemistry, biology, psychology, and economics (see Table 1). A larger set of 13 stimuli for each category were created and then used in a series of pilot studies with adults to narrow down the set to a final list of six for each category that college-level adults saw as all comparable in difficulty. Adults in the pilot studies were given the instructions shown in Appendix A.

The stimuli were randomized in two different orders. Participants evaluated each item on a 5-point scale concerning the difficulty of being able to fully understand the phenomenon described. All stimuli described phenomena in simple terms that would be easily accessible to young children, even though they might have no idea about the mechanisms underlying those phenomena. As shown in the stimuli in Table 1 and in the training and instructions used in Appendices A and B, the stimuli were equated in the extent to which surface descriptions of the phenomena were equally comprehensible (i.e., ease in identifying phenomena such as that iron is magnetic but gold is not), not in the extent to which one could explain the phenomena.

Procedure—Prior to testing, participants were given training to use a scale of 1 to 5 to rate how difficult each item would be to understand, with 1 meaning *very easy to understand* and 5 meaning *very hard to understand.* For the sixth-grade, eighth-grade, and college participants, this training was given through written instructions (see Appendix B). Training for the kindergartners, second graders, and fourth graders was done verbally with the experimenter. In addition, the kindergartners, second graders, and fourth graders were given several examples of items from other categories and asked to use the scale to determine how hard or how easy each would be to understand (see Appendix C). The training emphasized that children were to judge how difficult it would be “to learn everything there is to know about each thing … if you were taught all about them.” This part of the instructions ensured that children were not judging how hard it would be to access information through a source such as the Internet. All children were able to complete the training successfully, demonstrating an understanding of the different levels of ratings.

After training, participants were asked to rate the questions on how easy or hard they would be to understand. Ratings were collected with pencil and paper with the older participants. The kindergartners, second graders, and fourth graders were read each question aloud by the experimenter and asked to point to a number on a large scale or to say the number aloud. Pictures representing each question were also displayed on the table to help keep the children's attention. Each picture represented the topic in the question in a brief symbolic manner but did not help the participants rate the item's difficulty. For example, “Why is it that magnets pull on iron but not on glass?” was accompanied by a picture of a magnet picking up metal filings; “Why do the very best of friends still have fights?” was accompanied by a picture showing two people engaged in a heated verbal argument; and “Why are gas prices different in different parts of the country?” was accompanied by a picture of gasoline prices in the foreground with an unfamiliar city in the background. At random points in the procedure, the experimenter asked the participant to either repeat the question or tell why they chose that number, in order to ensure the participant was paying attention and using the scale correctly.

Results

Overall average scores for each of the five categories (physics, chemistry, biology, psychology, and economics) were computed and used as the basis for data analysis (Cronbach's alpha was calculated to assess the internal consistency of the rating scale, $\alpha = .874$). An initial analysis of gender of participants showed no overall effects on perceived difficulty of areas, $F(1, 156) = 1.04$, *ns*. There was also no difference due to stimulus order, $F(1, 156) = 0.76$, *ns*. The data were accordingly collapsed across these categories and then analyzed in a repeated measures analysis of variance (ANOVA) with mean scores for each category type as a repeated measure and grade as a between-subjects variable.

There was a general tendency for difficulty ratings to drop with increasing age, $F(5, 152) = 16.26$, $p < .001$, $\eta^2 = .349$. In addition, psychology items were generally rated as less difficult than both natural science items and economics, $F(4, 152) = 13.30$, $p < .001$, $\eta^2 = .080$. Finally, although economics items were seen as among the most difficult items by the younger children, they were seen as among the easier items by the oldest children and adults (as shown by an Age × Category interaction), $F(20, 152) = 3.75$, $p < .001$, $\eta^2 = .110$.

When differences among item types are examined for each age group through separate repeated measures ANOVAs, a difference was found among item types for kindergarteners, $F(4, 21) = 2.48$, $p < .05$, larger differences were found for Grades 2, 4, and 6 (*p*s < .001), intermediate differences were found for Grade 8, $F(4, 22) = 3.32$, $p < .02$, and no difference was found among item types for adults, $F(4, 27) = 1.36$, *ns*. The lack of differences among adults verifies that the pilot studies had indeed identified items that are judged as equal in difficulty by adults. In the kindergarteners, there was no significant difference among the psychology, biology, chemistry, and physics items, $F(3, 21) = 1.22$, *ns*, indicating that economics was primarily responsible for the overall category difference in that age group. In Grades 2, 4, and 6 there were strongly significant differences among these four domains (*p*s < .001), moderate differences for eighth graders ($p < .03$), and no significant differences for adults. (All significant differences represent main effects for domain type for the four domains in separate repeated measures ANOVAs for each grade).

Further analysis revealed that there were no significant differences among the judgments about the three natural sciences, $F(2, 152) = 0.23$, *ns*, $\eta^2 = .001$, physics, chemistry, and biology. A mean overall score for the natural sciences was therefore computed and plotted against economics and psychology, as shown in Figure 1. When specific comparisons are made among the natural sciences, psychology, and economics, economics was rated significantly more difficult than the natural sciences in kindergarten and significantly more difficult than psychology in Grades 2, 4, and 6 ($ps < .05$, Bonferroni corrected t tests). Economics was not significantly different from either psychology or the natural sciences at any other age groups, although there were trends to see economics as more difficult than the natural sciences in Grade 2 and less difficult than the natural sciences in Grade 8 ($ps < .10$, Bonferroni corrected t tests). In addition, psychology was seen as easier than the natural sciences in Grades 2, 4, 6, and 8 ($ps < .02$, Bonferroni corrected t tests).

Discussion

By the second grade, children think that the psychological phenomena are easiest and continue to do so, but with diminished contrast, until some point past the eighth grade. The adults saw no differences in difficulty across the domains. The kindergarteners only saw economics as an outlier of higher difficulty relative to the natural sciences, a difference that disappeared by the fourth grade. The general decline of rated difficulty with increasing age is assumed to reflect increased levels of actual understanding by participants. The lower scores in adults do raise the possibility of a floor effect that suppresses possible differences in difficulty ratings at that age. As the absolute floor would have been ratings of 1, there would still seem to be some possibility for ratings to occur that were lower than the adult mean of approximately 2.2. This view is further supported by the finding that every adult but one (27/28) gave at least one rating of 1 to a phenomenon. Thus, virtually all adult participants were willing and able to use the bottom end of the scale.

Study 2

The strongest discipline effect in Study 1 is the judgment, starting in the second grade, that the psychology phenomena are distinctly easier than those in the natural sciences and economics, a judgment that gradually diminishes until all phenomena are seen as being of equal difficulty by adults. Why should the psychological phenomena seem so much easier to younger participants? Perhaps any phenomena having to do with human behavior, broadly construed, are seen as easy. Alternatively, it may be that phenomena closer to immediate experience and feelings of self-involvement and self-efficacy seem easier. It may also be that phenomena that are more readily interpreted in terms of oversimplified folk psychological theories are judged as especially simple. Perhaps within the broad sense of what is considered psychology, some general kinds of phenomena are seen as easier to understand than are others. In particular, perhaps those more closely associated with desires, emotions, and “hot cognition” lend themselves to facile interpretations in terms of oversimplified folk psychological theories. To explore these alternatives, in Study 2, we created a continuum within psychology running from neuroscience to emotions. Would an early bias downplaying the difficulty of the “softer” areas again be created, or would all items involving human behavior be judged as equal in conceptual difficulty?

Method

Participants—Twenty-two kindergartners (mean age = 6 years, 0 months; range = 5 years, 7 months to 6 years, 8 months; 12 females, 10 males), 24 second graders (mean age = 7 years, 11 months; range = 7 years, 6 months to 8 years, 8 months; 14 females, 10 males), and 17 fourth graders (mean age = 9 years, 11 months; range = 9 years, 5 months to 10 years, 6 months; 10 females, 7 males) participated. In addition, 19 sixth graders (mean age = 11 years, 8 months; range 11 years, 0 months to 12 years, 0 months; 9 females, 10 males), 15 eighth graders (mean age = 13 years, 6 months; range = 13 years, 0 months to 14 years, 0 months; 6 females, 9 males), and 23 adults (mean age = 19 years, 10 months; range 18 years, 0 months to 24 years, 0 months; 11 females, 12 males) participated. The participants were drawn from the same population groups as those described in Study 1.

Materials—A continuum from neuroscience to emotions was set up, based on a familiar progression seen in many introductory psychology courses. There were four items in each of seven categories: neuroscience, sensation and perception, cognition, attention and memory, social psychology, personality, and emotions (see Table 2). Larger sets of eight items for each category were piloted with adults until a final set of four each was reached for which pilot studies showed roughly equal levels of rated difficulty.

Procedure—Study 2 was conducted in the same manner as Study 1, with only the test questions and stimuli pictures differing. As in Study 1, a 5-point scale was used.

Results

Overall average scores for each of the seven categories (neuroscience, sensation and perception, cognition, attention and memory, social psychology, personality, and emotions) were computed and used as the basis for data analysis ($\alpha = .791$). After initial analyses showed no gender differences, $F(1, 108) = 2.97$, *ns*, or effects due to stimulus order, $F(1, 108) = 1.02$, *ns*, the data were accordingly collapsed across these categories and were then analyzed in a repeated measures ANOVA with category type as the repeated-subjects measure and grade as a between-subjects factor. Difficulty ratings across the domains clearly differed, $F(6, 114) = 24.54$, $p < .001$ $\eta^2 = .177$. In addition, difficulty ratings dropped with increasing age, $F(5, 114) = 6.72$, $p < .001$, $\eta^2 = .228$. The differences in ratings between the domains also narrowed with increasing age (as shown by an interaction of category type with grade), $F(30, 113) = 2.29$, $p < .001$, $\eta^2 = .09$.

Overall, means for the seven domains revealed the following ordering of difficulty from most difficult to easiest to understand: neuroscience > memory and attention > sensation and perception > cognition > personality > social > emotion. Neuroscience items were significantly different from memory and attention, sensation and perception, and cognition, which were not significantly different from each other. Memory and attention, sensation and perception, and cognition scores were significantly different from personality, social, and emotional psychology scores, which were not significantly different from each other (all significant differences at $p < .05$ with Bonferroni corrections for multiple *t* tests). Figure 2 captures this pattern with the plotting of one curve for neuroscience; one for memory and attention, sensation and perception, and cognition scores averaged together; and one for personality, social, and emotional psychology scores averaged together.

When differences among item types are examined for each age group through separate repeated measures ANOVAs, there were significant differences among item types at the $p < .003$ level for the kindergarten grade group, $F(6, 21) = 3.47$, and at the $p < .001$ level for Grades 2, 4, 6, and 8. There were no significant differences among item types for the adult group, $F(6, 22) = 0.56$, *ns*. The lack of differences among adults verifies that the pilot studies had indeed identified items that are judged as equal in difficulty by adults. At a more focused level of *t* test comparisons with Bonferroni corrections, neuroscience was seen as more difficult than perception, attention, and cognition at Grades 2, 4, and 6 (ps < .02); neuroscience was seen as more difficult than social psychology, emotions, and personality at Grades 2, 4, 6, and 8 (ps < .002); and perception, attention, and cognition was seen as more difficult than social psychology, emotions, and personality at Grades 2, 4, and 8 (ps < .005).

Discussion

There was a relatively constant order of judged difficulty from kindergarten to the eighth grade, even as overall rated difficulty declined with age. The order reflected a continuum from biology to emotions as described earlier. The differences between the domains narrowed with age. Thus, within the domain of psychology, those phenomena that are closest to phenomenal experience and feelings of self-involvement or explanations in terms of simple folk psychological theories are seen as easier than others are. In particular, phenomena related to feelings, motivation, and interpersonal behavior are seen as distinctly easier to understand than are those having to do with the brain bases of behavior, with phenomena of cognition, attention, and perception being seen as having an intermediate level of difficulty. These differences seem especially dramatic to children, as opposed to young adolescents and adults. It is also notable that ratings of phenomena related to social psychology, emotions, and personality were the one case that adults started to rate phenomena as harder to understand than younger aged people did, $t(36) = 1.82$, $p < .08$ (two tailed). Study 2 also tends to support argument against the idea that a floor effect is driving the equated ratings shown by adults in Study 1. If the scale had an intrinsic floor of 2.2, as found in that study, it is more difficult to explain why the adult ratings all leveled off at 2.6 in this study.

The especially high difficulty ratings for all the neuroscience items may be related to the explicit use of the word *brain* in all descriptions. Prior research has shown that young children have trouble understanding what *brain* refers to, and thus it may be triggering a sense of difficulty in ways that are particularly related to the use of that word (Johnson & Wellman, 1982). Specific lexical items, however, cannot account for the other differences found in this study, as there were no constant lexical items distinguishing the cognition, attention, and perception phenomena from those concerning social psychology, emotions, and personality.

Study 3

Studies 1 and 2 suggest that the tendency to underrate the difficulty of areas such as psychology relative to physics may arise from a feeling that many psychological phenomena are intrinsically easy to grasp through direct experience and do not really require help from experts. In Study 3, we tested this suggestion more directly by asking children whether they could figure out various phenomena on their own or whether they would need to consult others for a full explanation. Prior studies have shown that adults are to able to judge whether experts are needed for making certain category-related decisions and that such judgments can vary considerably across such categories as artifacts and natural kinds (Kalish, 2002; Malt, 1990). Thus, it is plausible that adults and children might also be able to evaluate the need for experts to help understand various phenomena and that such intuitions might vary across broad domains. We approached this question in two ways that mirror criteria used by the courts to make decisions about what counts as an expert witness: being self-evident and being common knowledge. Thus, we asked children and adults about whether one could figure out a phenomenon all on one's own or whether one would have to ask others. We also asked them what proportion of adults would be able to explain a phenomenon. Because the largest differences between the domains used in Study 1 occurred in the second and fourth grade groups, Study 3 focused on children in kindergarten and in Grades 2 and 4, as well as on adults.

Method

Participants—Twenty-five kindergartners (mean age = 5 years, 8 months; range = 5 years, 1 month to 6 years, 4 months; 13 females, 12 males), 25 second graders (mean age = 7 years, 9 months; range = 7 years, 3 months to 8 years, 7 months; 16 females, 9 males), and 25 fourth graders (mean age = 9 years, 8 months; range = 9 years, 1 month to 10 years, 3 months; 16 females, 9 male) participated. In addition, 25 adults (mean age = 20 years, 5 months; range 18 years, 0 months to 24 years, 0 months; 16 females, 9 males) participated. The participants were drawn from the same population groups as those described in Study 1.

Materials—The same stimuli were used as in Study 1, except that now children were asked to judge, on separate 5-point scales, the ease of learning on one's own and how many adults would know the topic well (see Appendix D).

Procedure—All children were trained on two examples that either clearly required expert assistance to understand or that clearly could be understood without any help. Both examples involved artifacts, so as to use a different domain from all the natural phenomena used as stimuli. The case for expert assistance involved being asked how the insides of a computer worked, whereas the case requiring no help involved understanding how a crayon worked. Children were given both examples and then asked their opinions. They were then trained on a second set of items (a television and a tricycle) and asked how many grown-ups would know how they worked. The proportion of grown-ups with the queried knowledge was depicted by a crowd of adult silhouettes in which a subset of shaded ones were described as having the knowledge.

All child participants passed the training by the end of the second set of items. The full training protocol is shown in Appendix D.

Results

Overall average scores for “on-one's-own” judgments for each of the five categories (physics, chemistry, biology, psychology, and economics) were computed and used as the basis for data analysis ($\alpha = .792$). After initial analyses showed no gender differences, the data for “on one's own” judgments were subjected to a repeated measures ANOVA with category type as the repeated-subjects measure and grade as a between-subjects factor. Psychology items were seen as more likely to be learned on one's own at all ages than either any of the natural sciences or economics, $F(4, 96) = 38.61$, $p < .001$, $\eta^2 = .287$. In addition, the youngest children tended to see all items as harder to learn on one's own than did the older age groups, $F(3, 96) = 5.48$, $p < .002$, $\eta^2 = .146$. Finally, although the difference between all the natural sciences and psychology ratings stayed relatively constant across ages, economics was initially seen as the hardest domain to learn on one's own but shifted with age until it was judged by adults to be almost as easy to learn on one's own as psychology (as seen by a category type by grade interaction), $F(12, 96) = 4.62$, $p < .001$, $\eta^2 = .126$. Specific comparisons between domains revealed that all the natural science domain scores were not significantly different from each other at any age, but were significantly different from the psychology scores at all grades (all at $p < .001$, Bonferroni corrected *t* tests). Because they were not significantly different from each other at any age, the natural science scores were collapsed into one general score and plotted with the economics and psychology scores in Figure 3.

Overall average scores for “how many adults” judgments for each of the five categories (physics, chemistry, biology, psychology, and economics) were computed and used as the basis for data analysis ($\alpha = .885$). After initial analyses showed no gender differences, the data for “how many adults” judgments were subjected to a repeated measures ANOVA with category type as the repeated measure and grade as a between-subjects factor. Overall, psychology items were judged to be known by a larger proportion of adults than were all of the natural science items and the economics items, $F(4, 96) = 18.28$, $p < .001$, $\eta^2 = .160$.

Ratings of how many adults would know a domain well were different across age groups, $F(3, 96) = 17.87$, $p < .001$, $\eta^2 = .113$, the largest difference being a strong drop in adult participants in judgments of the number of adults who knew any phenomena well. Finally, the differences among domains changed with ages (as shown by a Grade × Item Type interaction), $F(12, 96) = 4.09$, $p < .001$, $\eta^2 = .137$, changing from no differences in the kindergarten group to a perception of psychology as known by more adults than both the natural sciences and economics in Grade 2, and as known by more adults than the natural sciences in Grade 4 and adults. Economics was also seen as known by more adults than the natural sciences among the adult grade group (ps < .01, Bonferroni corrected t tests). Because they were not significantly different from each other at any age, the natural science scores were collapsed into one general score and plotted with the economics and psychology scores in Figure 4.

As seen in Figures 3 and 4, the developmental profile of judgments of ability to learn on one's own was quite different from those for common knowledge. To explore this difference quantitatively, the scores for each discipline for each of the two judgment types were subjected to a principal components factor analysis with varimax rotation in which the initial data matrix contained all average judgment scores for each of the five disciplines and for each of the two judgment types, or 10 scores for each participant. The two judgments types clearly loaded strongly on two different factors, with both factors accounting for 62.78% of the total variance. Component 1, on which all "common knowledge" judgments had much higher varimax rotation loadings (mean of .823 vs. .053 for Component 2), accounted for 34.96% of the variance, whereas Component 2, on which all "learn on your own" judgments had much higher loadings (mean of .740 vs. .050 for component 1), accounted for 27.82% of the variance.

Discussion

"Common knowledge" and "on-one's-own" judgments may contribute to the difficulty effect in Studies 1 and 2 in relatively independent ways. In addition, the "on-one's-own" judgments distinguished social and natural science areas for the youngest children, whereas "how many adults" judgments did not distinguish between these areas until the second grade. This pattern suggests that judgments of the self-evident nature of phenomena may be first linked with judgments of difficulty and that this complex of intuitions then feeds into inferences about common knowledge that emerge somewhat later in development. Thus, even though preschoolers and young school children do have some sense that there are divisions of cognitive labor among adults (Keil et al., 2008; Lutz & Keil, 2002), it may take time to link up notions of difficulty with judgments of which domains are more likely to be common knowledge as opposed to specialized areas of expertise. It is not possible to tell whether judgments of being self-evident are the basis for judgments of difficulty or vice versa. It may also be that each intuition tends to suggest the other in a reciprocal manner.

Study 4

Studies 1 and 2 document an early emerging bias to consider phenomena in the natural sciences as intrinsically more difficult than those in the social sciences. Study 3 suggests that two factors may drive these biases, with intuitions about being self-evident preceding those about common knowledge. Taken together, these findings support the idea that heuristics for evaluating difficulty result in systematic distortions across broad domains. The possibility remains, however, that despite the training on the notion of difficulty in understanding, children might be using a broader metric to evaluate the phenomena. In particular, they could be simply evaluating how plausible or likely a phenomenon was to actually be true. Thus, if a child is so unfamiliar with a phenomenon as to not feel capable of evaluating its difficulty at all, she might default to an assessment of its plausibility, assuming that if something is implausible or impossible, it should be very difficult to understand as well. With respect to Study 1, children might have thought that social science phenomena were more likely to be true and were therefore using such judgments to stand in for what seemed to be judgments of difficulty. It is well known that children do sometimes make judgments on different grounds than the question or the training intended. For example, children can confuse the highly implausible with the strictly impossible (Shtulman & Carey, 2007). Thus, it would be helpful to show that for the exact same stimuli items, children can give quite different ratings from those given when they were evaluating difficulty of understanding the phenomena. Such a study would also rule out the possibility that children were responding on the basis of a global bias (for example, that they simply liked psychology more).

To explore this issue, in Study 4, we used the same stimuli phenomena as in Study 1, but instead of training children on difficulty and asking how difficult things were to understand, we asked how sure they were that the phenomena were true. If some very general notion of plausibility was driving children's intuitions in Studies 1 and 2, then judgments in Study 4 should show the same pattern. If, however, they clearly distinguished plausibility from difficulty, their judgments in Study 4 should pattern differently from those in Study 1. On the basis of occasional justifications suggesting attention to both difficulty and complexity, we indeed predicted that intuitions about plausibility would not replicate the systematic domain differences in judgments of difficulty.

Method

Participants—Twenty kindergartners (mean age = 6 years, 4 months; range = 5 years, 8 months to 6 years, 10 months; 9 females, 11 males), 20 second graders (mean age = 8 years, 3 months; range = 7 years, 10 months to 8 years, 8 months; 11 females, 9 males), and 20 fourth graders (mean age = 10 years, 4 months; range = 9 years, 10 months to 10 years, 10 months; 13 females, 7 males) participated. In addition, 20 adults (mean age = 20 years, 9 months; range 18 years, 0 months to 50 years, 0 months; 7 females, 13 males) participated. Children were drawn from the same population groups as those in Study 1. Eighteen of the adults were drawn from the same population as those in Study 1. Two were professionals working at the same university who were paid for their participation. Because the primary question was whether the judgments would be the same as in Study 1 and because the most dramatic patterns in that study involved kindergarten, Grades 2 and 4, and adults, this study focused on those ages.

Materials—The same stimuli were used as in Study 1, except that now children were asked about how likely it was that the phenomena were true. As in Study 1, all stimuli were accompanied by pictures that served to remind children of the phenomenon.

Procedure—All children were trained in a protocol that gave them an example of something “you are really sure about even if you don't know exactly why that is.” The example was that the “moon looks different on some nights than on other nights.” They were then told that they would be asked about how sure they were that various things were true, whether or not they knew why. They were then given examples of two other clearly true things that most people do not understand: that refrigerators keep things inside cool and that iPods store music.

Both examples involved artifacts, so as to use a different domain from all the natural phenomena used as stimuli. All children were shown how to give their responses on a 5-point scale that ranged from 1 (“*I think this is never the case*”) to 5 (“*I think this is always the case*”). All children seemed to easily understand the scale. The training protocols for children and adults are shown in Appendix E.

Results

Overall average scores for likelihood judgments for each of the five categories (physics, chemistry, biology, psychology, and economics) were computed and used as the basis for data analysis (α= .672). After initial analyses showed no gender differences, likelihood judgments were subjected to a repeated measures ANOVA with category type (economics, psychology, biology, chemistry, and physics) as the repeated-subjects measure and grade (kindergarten, 2, 4, and adult) as a between-subjects factor. Truth ratings varied across domains, with biology showing a general tendency to be seen as more true across all ages and economics as the least likely to be true, with the other domains in between, $F(4, 76) = 312.61$, $p < .001$, $\eta^2 = .142$. There was no general change in truth judgments with grade, $F(3, 76) = .28$, *ns*. The differences between domains did vary by grades, with the largest difference occurring in the two youngest grades, $F(12, 76) = 2.09$, $p < .018$, $\eta^2 = .076$. The ratings for each of the five domains are shown in Figure 5.

Although there are differences in ratings among the five domains, they show a very different pattern from that found in Study 1. There was no significant age difference, in strong contrast to the large effects due to grade in Study 1. In addition, the natural sciences all pattern in the same ways in Study 1 and can be collapsed together (as they could be in Studies 2 and 3 as well). In contrast, in Study 4, biology was seen as markedly different from the two other natural sciences (both comparisons at $p < .001$ level, Bonferroni corrected *t* tests). Overall, ratings were also higher in Study 4, but the results do not seem to be limited by a ceiling effect, given that there was still a wide range of responses and given that significant differences between the fields were found. Finally, economics judgments did not show the marked crossover effect found in Study 1 and instead were at or below the overall mean rating level in all age groups.

General Discussion

By equating judged difficulties between phenomena in adults, these studies asked whether biases construing the natural sciences as more difficult are present in young children. Biases about the difficulty of the sciences might seem to be late emerging and reflecting sophisticated insights into the practices of the disciplines, but the opposite seems to be the case. Studies 1 and 2 both point to a powerful bias by at least the second grade. The high rating for economics early on may reflect inferences about the need for manipulation of numbers. There is also a distinct pattern in which matters having to do with psychological phenomena are seen as especially easy to understand. This seems to be especially true for psychology items related to beliefs and emotionally laden phenomena. Thus, in Study 2, neuroscience accounts of behavior were judged as much more difficult than more purely psychological ones. The discipline bias is unlikely to be driven by mere familiarity of the items, as several items in physics, chemistry, and biology (e.g., how soap cleans dishes, how skin heals) were more familiar than several items in psychology and economics. Similarly, some neuroscience items (e.g., right and left-handedness) were more familiar than some social psychology ones (e.g., blame shifting). Children can look beyond familiarity to judge the quality of arguments (Ackerman, Spiker, & Glickman, 1990); here, they use heuristics that discount the difficulty of the social sciences relative to the natural sciences. Moreover, when average word frequencies are computed for all the words used in descriptions of the phenomena in Studies 1 and 2, there are no relationships between average frequencies and difficulty judgments, as well as no systematic differences across domains.

Study 3 suggests that judgments of difficulty may be mediated by judgments of the seemingly self-evident nature of knowledge in some domains, which appears to precede judgments of the extent to which a full understanding of a phenomenon is common knowledge. Study 3 also suggests that when adults are asked to make “on-one's-own” judgments or "how many adults" judgments, their latent biases about the relative simplicity of the social sciences leak through, even though when asked to directly rate the difficulty of the phenomena used in these studies, adults rated all of them of equal difficulty across all the disciplines. The legal system's use of intuitions about common knowledge and about a phenomenon being self-evident, instead of about difficulty or complexity, may help explain why social science experts have been excluded so often in the courtroom.

Finally, Study 4 indicates that children were not basing their difficulty judgments simply on how likely it was for various phenomena to be true. The pattern of judgments in Study 4 looked strikingly different from those found in Studies 1–3. In particular, the natural sciences did not cluster as a coherent group, and biology received the highest rating instead of psychology, which received more intermediate ratings of likelihood. Moreover, there was no developmental shift in how economics was perceived with respect to the likelihood of phenomena being true. Taken together, these results support the view that at all ages studied, difficulty judgments pattern very differently from truth judgments.

The findings do not support a fine-grained hierarchy in which physics phenomena are seen as more difficult to understand than those in chemistry, and chemistry phenomena are seen as more difficult than those in biology. Instead, all the natural sciences were seen as being equal in difficulty. Strictly speaking then, these studies do not support a physics envy, in which phenomena in physics are seen as the most difficult to understand of all domains. If such views do exist, they may emerge in some adults rather than be present in children. Psychological phenomena, however, do seem to be seen as especially easy to understand in comparison with all other domains of phenomena. Thus, the primary cause of domain differences may lie in a discounting of the difficulty of understanding psychology. This discounting may arise from a sense that psychological phenomena fit nicely into simple and easily available folk-psychological theories of behavior, theories that may well be relatively coherent and comprehensive in comparison with more fragmentary theories in other domains such as physics and biology. In addition, as seen in Study 2, the softer areas of psychology and those least like the natural sciences seem to be those in which the bias is the strongest.

We have also suggested that a sense of immediacy and control may make psychological phenomena seem easier to understand. When one has privileged access to certain phenomena (a form of immediacy), such as those relating to internal psychological states, one has more confidence about one's abilities related to those states (Dunning, Heath, & Suls, 2004). People also tend to also see psychological abilities as more controllable (Lockhart et al., 2008), and feelings of controllability of abilities are also related to higher levels of confidence about one's level of performance on those abilities (Alicke, 1985). This relatively high level of overconfidence about one's psychological abilities (especially for ones that are difficult to objectively measure) may spread to an overconfidence in the ability to understand those phenomena. If people think that they are very good at something, they may also think that they will find it easy to understand. This is a plausible extrapolation from the ways that people learn many skills, such as how to play various games or how to use a simple mechanical tool such as an adjustable wrench. High skill levels often make something more immediate and controllable, as well as easier to understand. Perceived control and skill levels in an area may therefore serve as false proxies for having explanatory understanding as well.

Simplified folk psychological theories may be related to feelings of immediacy and control, and they may mutually reinforce each other. If one thinks that most psychological phenomena can be experienced more directly and can be easily controlled, one might well think that they are easy to explain as well. Conversely, if one thinks that one can provide simple explanations for psychological phenomena, one might also think that they are easy to experience and to control. Despite these interrelationships, it does not seem that control and immediacy are simply tautologically part of psychology. After all, within psychology itself, those phenomena that seem to be especially immediate and controllable are those that also seem the easiest. If there are phenomena outside of psychology that seem equally immediate and controllable, they too might seem easy to explain.

Intuitions about the difficulty of understanding economics phenomena have a distinct developmental trajectory. Although only one of the items used explicitly mentioned numbers (the reference to $100 for currency), it is possible that the mention of prices and currency leads children to infer that those concepts require some facility with numbers. For young children, the inferred presence of mathematics may be stronger than in other disciplines in which the role of mathematics may only be gradually understood with increasing exposure to expertise in that domain. However, by the eighth grade, phenomena in economics were seen as almost as easy to understand as are those in psychology. This shift may be a result of three related developmental changes. First, by the eighth grade, the kind of mathematics that is apparent in descriptions of economic phenomena, such as price changes and calculations of costs, may seem quite easy to understand, whereas the complex mathematics used in economic theory would surely still be completely unknown to an eighth grader. Second, 13-year-olds are much more in control of their economic activities than are younger children and have more of a sense of mastery and control over economic transactions. Finally, there may be an emerging sense of some of the psychological dimensions of economic phenomena. Because the presence of numbers could only have been inferred by the younger children for four of the five items used, the primary reason economics was seen as so difficult by younger children might have more to do with a feeling of having no mastery or control over economic transactions or with no ability to predict economic phenomena such as prices and costs, which may seem utterly mysterious to children until they become more active participants in economic systems and transactions.

To what extent are these biases in difficulty ratings that are so apparent in children still at work in adults? In the studies presented here, the ratings were deliberately equated for difficulty in adults, but a bias may still persist. If there were an objective way to measure complexity, perhaps the psychological phenomena would be found to be more complex, but then discounted by cognitive biases. Such biases may be observed when tasks do not so directly ask about difficulty of understanding and instead use measures such as ease of learning on one's own. Even the adults in Study 3 felt that psychological phenomena would be easier to learn on one's own than phenomena in other domains. The biases might also surface in cases in which adults are distracted or forced to quickly assess the difficulty of understanding various phenomena. They might also appear when one has to judge the need to use experts in a domain. For example, people might be more prone to see a designated expert for a biological disorder, such as a skin rash, than for a psychological disorder, such as anxiety, even if the efficacy of treatment is the same for the two disorders. Instead, for the psychological disorder, people might be more prone to think they can figure out a solution on their own or through the advice of nonexpert friends.

Systematic biases in how people evaluate the difficulty of attaining a good understanding in various domains have broad implications for how people decide when they need to turn to others for insight, when they should defer to claims made by others, when they should value an area of instruction, and when they should doubt their own understanding. As described in the introduction, a particularly vivid effect of such biases might be on the admission of expert witnesses into the courtroom. If psychological phenomena are seen as usually quite easy to understand and largely self-evident and if such judgments are inaccurate and underestimate the need for experts, cases might well be decided in ways that unfairly exclude valuable expert insights. Even salaries for professionals in various disciplines may partially reflect conceptions of how difficult it is to attain expertise in those areas. The single most effective way to reduce such biases at any age may be to acquaint lay people as much as possible with the real complexity of phenomena in all domains and with the extraordinary level of intellectual investment that is required to gain a deep understanding. Even if one cannot fully know the complexity oneself, education might well enable one to more realistically appreciate the abilities of those who do have a deep understanding in a domain.

Acknowledgments

This research was supported by National Institutes of Health Grant R-37-HD023922 to Frank C. Keil. We thank the Fairfield Board of Education and Osborn Hill Elementary School, Fairfield, Connecticut. We thank John Bargh and Paul Bloom for comments on drafts of this article, Eric Smith for assistance with Study 3, Kara Gaughen for assistance with Study 4, and Jonathan Kominsky for help in preparation of this article.

Appendix A

Instructions Given to Adults in Pilot Studies That Narrowed Set of Candidate Stimuli From 65 Items (13 per Category) Down to 30 Items (6 per Category)

In this study, you will be reading about different problems and asked to rate how difficult the problem is. You will be asked to use a 5-point scale that represents the difficulty level of the problems. Think of it as a continuum, in which 1 represents the least difficult problems and 5 represents the most difficult problem.

Some of these problems will be about very complicated things that are very, very hard to understand. It would be very difficult to know everything about the problem and would take you a very long time to learn everything there is to know about the problem. If you feel that a problem is like this, very complicated and very, very difficult to understand, then you would give it a rating of 5.

Some of the problems will be very, very easy to understand and not very complicated. It would only take to a couple minutes to learn all there is to know about the problem. These problems would be rated as a 1.

And finally, some problems will be in the middle, they will be somewhat difficult to understand, and somewhat complicated. You might be able to learn everything there is to know about these problems in a couple hours or days.

Appendix B

Instructions Given to Adults, Eighth Graders, and Sixth Graders in Study 1

Instructions: You know how some things are very easy to understand, while other things are very hard to understand? And for some things, even if you don't know about it, it would be easy to learn all about it. Other things it would take a really long time to learn all about, and you might not even be able to because it's so hard to understand.

In this survey you will be presented with lots of different questions, and for each one, I want you to rate on a scale of 1–5 how easy or hard it would be to understand.

For some of the questions, you might not know much about them at all, and that's ok! I DON'T want you to rate how much you know about the topic; rather, rate how easy or hard the topic would be to understand if you were taught all about it.

If you think it would be very easy to understand … you'd say 1.

If you think it would be very hard to understand … you'd say 5.

The numbers 1–5 represent a scale going from very easy to understand to very hard to understand.

For example, you can think of it like this:

1. very easy to understand
2. pretty easy to understand
3. sort of hard/sort of easy to understand
4. pretty hard to understand
5. very hard to understand

Read each question carefully and please rate honestly.

Appendix C

Instructions Given to Fourth Graders, Second Graders, and Kindergarteners in Study 1

Today we are going to be playing a type of game. Now, do you know how some things are very easy to understand while others are hard to understand? And for some things, even if you don't know about it, it would be very easy to learn all about it? Other things it would take a really long time to learn all about, and you might not even be able to because it's so hard to understand.

Well in this game, I'm going to be telling you about a lot of different things and for each one, I want you to tell me how easy or how hard it is to understand. Does that make sense?

For a lot of these you probably don't know much about them; I don't even know how they work. But what I want you to tell me is how hard you think it would be to learn everything there is to know about each thing. So, tell me how easy or hard it would be to understand these things really well if you were taught all about them.

Again, don't tell me how much you know about them, tell me how easy or hard it would be to learn and understand them if someone tried to teach you about them.

If you think it would be really, really easy to understand … you'd say 1.

If you think it would be really, really hard to understand … you'd say 5.

If you think it would be sort of hard, you'd say 3; pretty hard you'd say 4; and pretty easy you'd say 2.

For example, if I asked you how hard you thought it would be to understand how a button works, you probably would say it is really very easy to understand, even if you have never seen a button before. How a button goes sideways in a button hole and then holds the two sides together is easy to understand, right? So you would probably say 1 for that, right?

But if I asked you how hard it would be to understand how a computer works, you might think that it is really hard because there is so much to know about all the parts of the computer and how they work together. Because this is really, really hard to understand, you'd probably say what number??

… right, a 5!

So again, some things are pretty easy to learn about and understand and some really, really hard. Let's try a couple more examples …

How hard would it be to understand what 1 + 1 equals?

How hard would it be to understand what 1,437,932 + 245,346 equals?

How hard would it be to understand why socks go on your feet and not your hands? Right, this is really easy … even if you didn't know at first, it would be easy to learn, so you'd probably say 1 or 2.

How hard would it be to understand or learn how a car works? Right … that is really hard … so you'd say 5.

OK, let's get started. Remember, tell me how easy or hard it would be to understand everything there is to know about each thing.

Now remember, some will be easy and some will be hard, so make sure you use all the numbers, ok?

Appendix D

Training Protocol Used in Study 3

Today we are going to be playing a type of game. Now, do you know how some things are very easy to understand and you can figure out all about how they work just by looking at them and/or playing with them yourself? For some things though, no matter how long you look at them or play with them they are so hard to understand that you have to ask others with special knowledge to help explain how they work.

You probably don't know much about a lot of these things we are going to be talking about and that's ok. What I would like you to tell me is how well you think you would be able to figure out everything there is to know about each thing on your own just by living and watching things. Let's try an example …

If I asked you to figure out everything about how a computer works, do you think you could figure that out all on your own or would you need to learn some things from a person who has special knowledge about how computers work?

That's right! You might be able to figure out a little bit abut how a computer works by looking at the parts and playing with it, but in order to figure out everything about how it works you would have to learn many things from a person who has special knowledge about computers. Computers have so many small parts that help them work, it would be almost impossible for you to learn and understand all of them just on your own.

If I asked you to figure out everything about how a crayon works, do you think you could figure that out all on your own or would you need to learn some things from a person who has special knowledge about how crayons work?

That's right! You would be able to figure how a crayon works all by yourself just by looking at it and playing with it yourself. After all, a crayon just has a place for you to hold it and a pointed tip to help you color things. It is not very complicated at all.

I am also going to ask you to tell me how many grown-ups you think would know these things.

Let's try an example …

What if I asked you how many grown-ups would know how a TV works?

That's right…probably only a few grown-ups would know how a TV works because it has many parts that work together in ways that are hard to understand.

Let's try another example …

What if I asked you how many grown-ups would know how a tricycle works?

That's right, probably every adult would know how a tricycle works. After all, there aren't that many parts to a tricycle besides the wheels, pedals, and steering bar. It would be easy for an adult to look at something with so few parts and explain how it works.

Appendix E

Training Protocols for Study 4

Child Participants: Today we are going to be playing a type of game. Now, do you know how some things you are really sure about even if you don't know exactly why that is? So I know for sure that moon looks different on some nights than on other nights. I would say I'm very sure that is true, even though I don't think I could tell you why that is.

Some of the things we are going to talk about today you may know about or you may not. We are going to be talking about how sure you are about whether different things are true.

For a lot of these things you probably don't know much about them; I don't even know how they all work. But what I want you to tell me is how sure you are about whether these different things are true.

Let's try an example …

If I told you that an IPod stores music inside, how sure are you about that? Do you think that is really the case?

Yes that's right, it is true!

Even though you may not know how an IPod does that or all about the ways an IPod works, you are really sure that an IPod DOES store music inside.

And if I told you that a refrigerator keeps everything inside cool, how sure are you about that? Do you think that is really the case?

Yes, you're right! That is true! We might not know exactly how it does that, but we know that a refrigerator REALLY does keep the things inside it cool.

So when I tell you about these different things, I am going to ask you how sure you are that each thing is true.

Adult Participants: Today we are going to asking you about different things and how sure you are that they are true. Now, do you know how some things you are really sure about, even if you don't know exactly why that is? You may know for sure that the moon looks different on some nights than on other nights. You would say that you are very sure that is true, even though you might not be able to give a detailed explanation for why that is.

Some of the things we are going to ask about today you may know about or you may not. What we are asking is how sure you are about whether different things are true.

Please read through the following survey and pay close attention to the scale. Mark your answers accordingly.

References

Ackerman BP, Spiker K, Glickman I. Children's sensitivity to topical discontinuity in judging story adequacy. Developmental Psychology 1990;26:837–844.

Alicke MD. Global self-evaluation as determined by the desirability and controllability of trait adjectives. Journal of Personality and Social Psychology 1985;49:1621–1630.

Anderson PW. More is different. Science August;1972 177:393–396. [PubMed: 17796623]

Bernstein DM, Atance C, Loftus G, Meltzoff AN. We saw it all along: Visual hindsight bias in children and adults. Psychological Science 2004;15:264–267. [PubMed: 15043645]

Blinder, AS. Economics becomes a science—Or does it?. In: Beam, A., editor. Useful knowledge: The American Philosophical Society Millennium Program. American Philosophical Society; Philadelphia, PA: 1999. p. 141-154.

Bloom P. My brain made me do it. Journal of Culture and Cognition 2006;6:209–214.

Cohen JE. Mathematics as metaphor. Science May;1971 172:674–675.

Colman AM, Mackay RD. Legal issues surrounding the admissibility of expert psychological and psychiatric testimony. Issues in Criminological and Legal Psychology 1993;20:46–50.

Colman AM, Mackay RD. Psychological evidence in court: Legal developments in England and the United States. Psychology, Crime, and Law 1995;1:261–268.

di Sessa A. Towards an epistemology of physics. Cognition and Instruction 1993;10:105–225.

Dufresne A, Kobasigawa A. Children's spontaneous allocation of study time: Differential and sufficient aspects. Journal of Experimental Child Psychology 1989;47:274–296.

Dunning D, Heath C, Suls JM. Flawed self-assessment: Implications for health, education, and the workplace. Psychological Science in the Public Interest 2004;5:69–106.

Jacobs JE, Klaczynski PA. The development of judgment and decision making during childhood and adolescence. Current Directions in Psychological Science 2002;11:145–149.

Jefferson M. What anthropology can do for psychology: Facing physics envy, ethnocentrism, and a belief in “race.”. American Anthropologist 2000;102:552–563.

Johnson CN, Wellman HM. Children's developing conceptions of the mind and brain. Child Development 1982;53:222–234. [PubMed: 7060424]

Kalish CW. Children's predictions of consistency in people's actions. Cognition 2002;84:237–265. [PubMed: 12044735]

Keil FC, Stein C, Webb L, Billings VD, Rozenblit L. Discerning the division of cognitive labor: An emerging understanding of how knowledge is clustered in other minds. Cognitive Science 2008;32:259–300. [PubMed: 19759842]

Kinlaw CR, Kurtz-Costes B. The development of children's beliefs about intelligence. Developmental Review 2003;23:125–161.

Lockhart KL, Chang B, Story T. Young children's beliefs about the stability of traits: Protective optimism? Child Development 2002;73:1408–1430. [PubMed: 12361309]

Lockhart KL, Nakashima N, Inagaki K, Keil FC. From ugly duckling to swan? Japanese and American beliefs about the stability and origins of traits. Cognitive Development 2008;23:155–179. [PubMed: 19337584]

Loftus EF. Ten years in the life of an expert witness. Law and Human Behavior 1986;10:241–263.

Lutz DR, Keil FC. Early understanding of the division of cognitive labor. Child Development 2002;73:1073–1084. [PubMed: 12146734]

Mabe PA III, West SG. Validity of self-evaluation of ability: A review and meta-analysis. Journal of Applied Psychology 1982;67:280–286.

Mackay, RD.; Colman, AM.; Thornton, P. The admissibility of expert psychological and psychiatric testimony. In: Heaton-Armstrong, A.; Shepherd, E.; Wolchover, D., editors. Analysing witness testimony: A guide for legal practitioners and other professionals. Blackstone; London, England: 1999. p. 321-334.

Malt BC. Features and beliefs in the mental representations of categories. Journal of Memory and Language 1990;29:289–315.

Mills C, Keil FC. Knowing the limits of one's understanding: The development of an awareness of an illusion of explanatory depth. Journal of Experimental Child Psychology 2004;87:1–32. [PubMed: 14698687]

Nicholls JG. The development of the concept of difficulty. Merrill-Palmer Quarterly 1980;26:271–281.

Nichols S. The folk psychology of free will: Fits and starts. Mind & Language 2004;19:473–502.

Quine, WVO. Theories and things. Harvard University Press; Cambridge, MA: 1981.

Schneider, W.; Pressley, M. Memory development between two and twenty. Springer-Verlag; New York, NY: 1989.

Shtulman A, Carey S. Improbable or impossible? How children reason about the possibility of extraordinary events. Child Development 2007;78:1015–1032. [PubMed: 17517019]

Simonton DK. Varieties of (scientific) creativity: A hierarchical model of domain-specific disposition, development, and achievement. Perspectives on Psychological Science 2009;4:441–452.

Sober, E. Simplicity. Clarendon Press; Oxford, England: 1975.

Weisberg DS, Keil FC, Goodstein J, Rawson E, Gray J. The seductive allure of neuroscience explanations. Journal of Cognitive Neuroscience 2008;20:470–477. [PubMed: 18004955]

Wellman HM, Liu D. Scaling of theory-of-mind tasks. Child Development 2004;75:523–541. [PubMed: 15056204]

Wellman HM, Fang F, Liu D, Zhu L, Liu G. Scaling of theory-of-mind understandings in Chinese children. Psychological Science 2006;17:1075–1081. [PubMed: 17201790]

Zimmerman, C.; Gerson, S.; Monroe, A.; Kearney, AM. Physics is harder than psychology (or is it?): Developmental differences in calibration of domain-specific texts. In: McNamara, DS.; Trafton, JG., editors. Proceedings of the Twenty-Ninth Annual Cognitive Science Society. Cognitive Science Society; Austin, TX: 2007. p. 1683-1688.

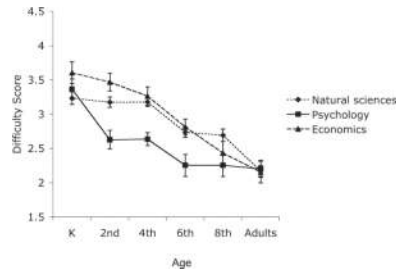

Figure 1.
Difficulty ratings across the natural and social sciences. By the second grade, children see natural science phenomena as much more difficult to understand than psychological ones. Although they initially see economics as very difficult, older children come to see it as more akin to psychology. Error bars represent standard errors. K = kindergarten.

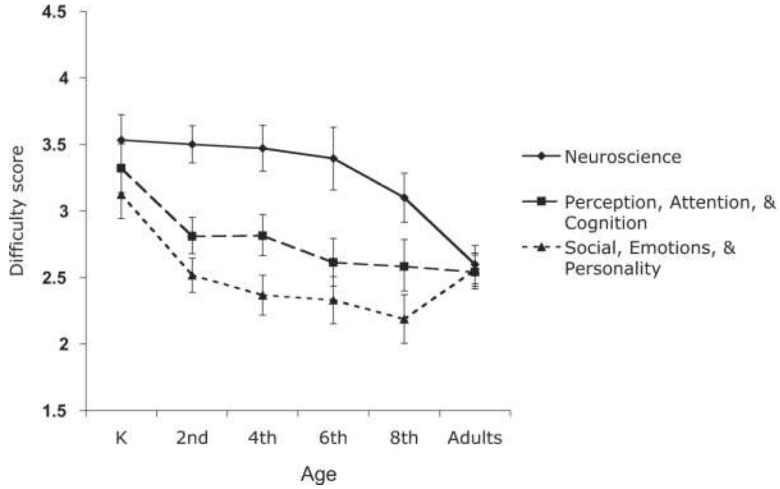

Figure 2.
Difficulty ratings across areas of psychology. From kindergarten on, children see neuroscience phenomena as much more difficult to understand than cognitive, attention, and perception ones, which in turn are seen as more difficult than those relating to social psychology, emotions, and personality. Error bars represent standard errors. K = kindergarten.

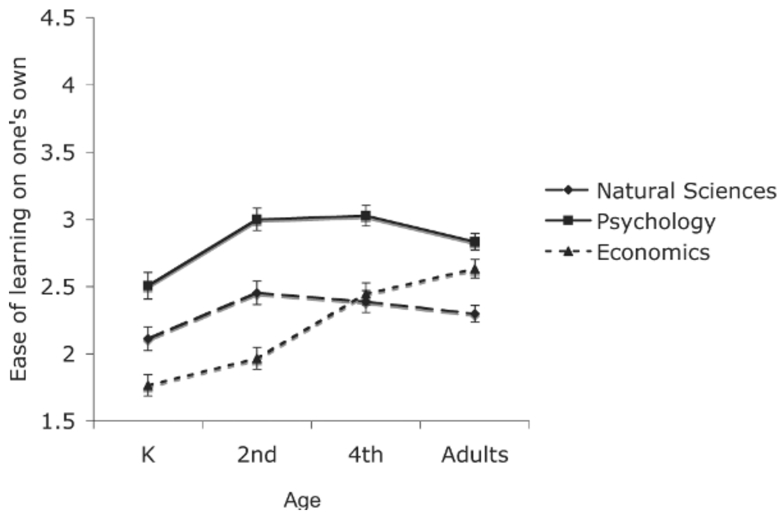

Figure 3.
Judgments, across the natural and social sciences, of "How much of this could you figure out on your own just by living and watching things?" Error bars represent standard errors. K = kindergarten.

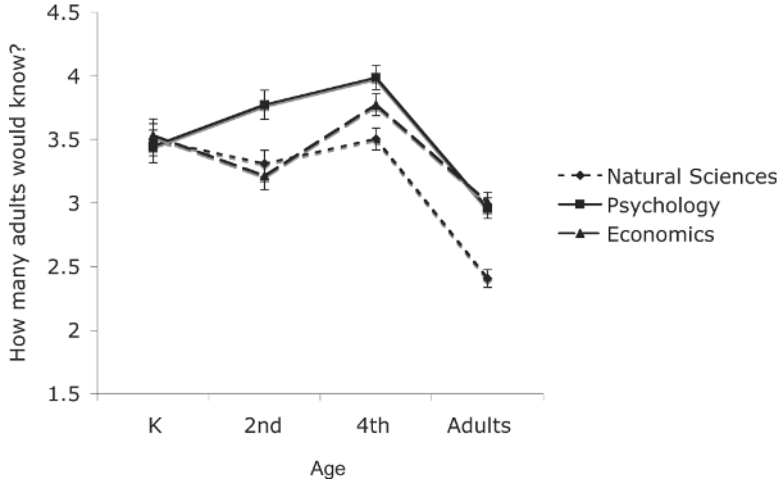

Figure 4.
Judgments, across the natural and social sciences, of “About how many grown-ups would know this well?” Error bars represent standard errors. K = kindergarten

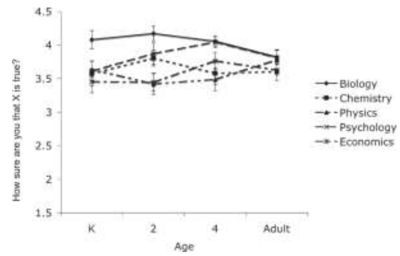

Figure 5.
Judgments, across the natural and social sciences, of “How sure are you that *X* is true?” *X* represents the item being queried. Error bars represent standard errors. K = kindergarten.

Table 1

Stimuli Used in Study 1

Stimuli
Physics
How does a top stay spinning upright?
Why do rainbows always have the blue/purple color on the bottom stripe and red on top?
How can some boats sail faster than the wind?
Why are sunsets orange and red?
Why does light travel faster than sound?
Why is it that magnets pull on iron but not on glass?
Chemistry
Why do eggs go from a liquid to a solid when you cook them?
How does water mixed with soft sand and soft powder make hard concrete?
Why does paper burn but not aluminum foil?
Why can you bend copper tubes but not steel ones of the same size?
Why is diamond much harder than glass?
Why does soap clean grease off of dishes?
Biology
How does your skin get a tan after it has been in the sun for a while?
How do seals stay underwater for such a long time?
How do birds with really skinny legs keep from freezing their feet in the winter?
Why are we allergic to some things and not others?
Why do some people have freckles and others don't?
How does your skin heal after it has been cut?
Psychology
Why is it hard to understand two people talking at once?
Why do you have to repeat phone numbers in your head in order to remember them?
Why can't you remember things from when you were a baby?
Why do the very best of friends still have fights?
Why is it harder to recognize a photo of someone if it is upside-down?
Why can children learn new languages more easily than adults?
Economics
Why are gas prices different in different parts of the country?
How do different countries know how much of their money to give you in trade for $100 dollars?
Why can bigger companies make cars cheaper than smaller companies?
Why do silver prices go up and down more than gold prices?
Why do house prices go up and down over the years?
Why is it so hard to figure out what something will cost a year from now?

Table 2

Stimuli Used in Study 2

Stimuli
Neuroscience and biological psychology
Why does the brain make it easier for some people to use their left hand and not their right hand?
Why is it that when you get tired, your brain doesn't work as well?
How does your brain know when to have you wake-up?
Why does more blood flow to your brain when you are thinking really hard?
Sensation and perception
Why is it harder to recognize a photo of someone if it is upside-down?
Why can some people only see black and white and not colors?
Why does yellow look brighter next to purple than next to orange?
Why can dogs can hear really high sounds that humans can't hear?
Cognitive psychology
Why can children learn new languages more easily than adults?
Why do some people have trouble following directions?
How can you be really good at reading and really bad at math?
How do you recognize yourself in the minor?
Attention and memory
Why do you have to repeat phone numbers in your head in order to remember them?
Why is it hard to understand two people talking at once?
Why do people forget things?
Why is it hard to make yourself forget something?
Social psychology
Why do the very best of friends still have fights?
Why are people are more likely to get angry when they are tired?
Why do people sometimes blame their mistakes on other people?
Why do people sometimes feels like they have to do what everyone else is doing?
Personality
Why do people sometimes lie about something bad they did?
Why is it hard to break a bad habit?
Why do some people give up more easily than others?
Why is it hard to wait to do something you REALLY want to do?
Emotions
Why do people cry at a sad movie?
Why do you get nervous when you're doing something new?
Why does laughing make you happy?
Why are some people afraid of the dark?