Published in final edited form as:
Anthropol Philos. 2011 January 1; 10(1-2): 61–81.

The Hidden Strengths of Weak Theories

Frank Keil
Yale University

Abstract

There has been a strong tradition of assuming that concepts, and their patterns of formation might be best understood in terms of how they are embedded in theory-like sets of beliefs. Although such views of concepts as embedded in theories have been criticized on five distinct grounds, there are reasonable responses to each of these usual objections. There is, however, a newly emerging concern that is much more challenging to address – people's intuitive theories seem to be remarkably impoverished. In fact, they are so impoverished it is difficult to see how they could provide the necessary structure to explain differences between concepts and how they might form in development. One response to this recent challenge is to abandon all views of concept structure as being related to people's intuitive theories and see concepts as essentially structure-free atoms. The alternative proposed here argues that our very weak theories might in fact do a great deal of work in explaining how we form concepts and are able to use them to successfully refer.

For many years it has been assumed that concepts are embedded within larger systems of beliefs that help articulate their structure. These systems of beliefs are often thought of as intuitive or naïve theories and are thought to be a key way of explaining concept formation and conceptual change. In particular, the emergence of new theories out of old ones in which inconsistencies become apparent are thought to be a primary vehicle for the formation of new concepts (Carey, 2009). In the philosophy of science, it has long been held that theories provide critical frameworks within which concepts are articulated, frameworks that give sense of ontological kinds and of relations between concepts (Kuhn, 1977). In developmental psychology, intuitive theories have even been attributed to infants and have been argued to be the best ways to understand early concepts (Gopnik & Meltzoff; 1997). Intuitive theories have also been seen as causing a potential conflict with more associative views of concepts in which a concept is little more than tabulations of how often features occur and co-occur for certain entities (Johnson & Keil, 2000; Keil, 1989).

Yet, this view of concept formation faces a major challenge. Are the intuitive theories of lay people and of the folk sciences, namely the ways lay people make sense of various phenomena in the world, adequate as means for understanding concept formation, growth and use? The answer is unclear and has led some to propose either minimalist views of concepts (Fodor, 1998) or even to do away with concepts altogether (Machery, 2009). Here I want to suggest that there may be ways to maintain an account in which concepts are

Department of Psychology, Yale University, frank.keil@yale.edu.

associated with rich structures but which also acknowledge the many limitations of intuitive theories.

Let us therefore consider in some detail the view that concepts are embedded in theories, and that they derive their structures from theories. More precisely, concepts are embedded in theory-like structures and are distinguished from each other by the particular ways that each concept is embedded in a web of relations that make up a theory. That web might be characterized as a “web of belief”, (Quine & Ullian, 1978), or perhaps more primitively as less belief-like cognitively implicit links to other concepts and properties. Thus, one might argue that more brute force associations between networks of concepts define new concepts (Rogers & Mclelland, 2004). The same idea can be advanced for constituents of concepts, namely that a concept such as DOG is made up of various “perceptual” and “conceptual” features that are presumed to make up the meaning of dog, such as having four legs, barking, being a living creature and having an essence. (There is much to worry about in such accounts, including whether any appeals to features of concepts are in fact simply making more links to other concepts (Fodor,1998), but since much bigger questions will emerge about the larger enterprise, those worries do not need to be dwelt on here.)

One reason for thinking that concepts are made up of theories in this way is because of apparently powerful links between conceptual change and theory change. Whether it be in the history of ideas or a particular child, concepts seemed to travel in groups (Carey, 1985; Keil, 1989; Thagard, 1992). When a child has a particular kinship term that shifts in meaning over the course of development, many others concepts seem to shift at the same time in terms of what they mean to the child (Keil, 1989). When a child’s concepts of weight seem to change, concepts of density may change as well, and similarly for heat and temperature (Smith, Carey & Wiser, 1985; Wiser & Carey, 1983). When the concept of evolution by natural selection emerged in the 19th century, it seemed to be related to a change in the concept of a species. Many other examples exist in the history of science and in cognitive development (Thagard, 1992).

In addition to cases of conceptual change, many concepts seem to be interdependent. It is not clear that it is even coherent for someone to claim to have a concept of a mechanical NUT without having the concept of a BOLT that accepts that nut. Similarly. BUY can’t seem to stand on its own without the concept SELL, MOTHER without CHILD, and so on. These cases appear to illustrate the idea that concepts are part of larger relational complexes that both give meaning to them and make up their meaning. Come to understand one concept and, in at least some cases, you will automatically understand the other. If concepts are embedded in theories and the same theory applies to two different concepts, it offers an explanation of why they should be linked in understanding and in conceptual change.

The concepts-in-theories idea also seems to gain support from the ways in which notions of “how” and “why” seem to influence all aspects of concept acquisition and use. Often, one of the most striking aspects to concepts is not how often features occur with instances – that is what makes up prototypes (or syndromes, whatever you want to call them), but rather the causal explanatory roles filled by the features associated with concepts. Even very young children and infants do not judge category membership by merely weighing features on the basic of their typicality and then doing some aggregation of such weights over features to determine category membership. The degree to which they think a property is causally important will often trump typicality. Thus, even if all known tires have been black and only 95% have been fully round (as opposed to flat), shape is considered much more important to being a tire than color (Keil, 1994; Keil et al., 1998; Keil, 2010). Similarly, as adults at least, we discount highly salient and reliable features of hair length and clothing as central to a concept of male and female and emphasize other features that are much less frequently observed. In the same way, when we make inductions about other things that are likely to be true of thing, we make those inductions not merely on the basis of past frequencies of features, but also on the basis of guesses about their causal roles in category. So, if told that a sampling of ten cats revealed that all had a particular enzyme for digesting meat and that all had two syllable names, we are much more likely to infer that the enzyme is a critical feature of all cats not the two syllables (Heit, 2000; Proffitt, Coley and Medin, 2000; Wisniewski and Medin, 1994). In both adults and young children causal explanatory knowledge seems to influence how features are used to structure categories and their associated concepts (Hayes & Thompson, 2007; Rehder & Kim, 2006).

In short, concepts as theories (or at least as embedded in theories) seem like a compelling way to characterize their nature in adults. At the same time, there are some concerns in the adult psychological literature as well. For example, frequency based information can have a strong influence in at least some contexts (Hampton, 2000). Moreover, the importance of causal information may vary as a function of domain with it possibly being more influential for living kinds than for artifacts (Hampton, Storms, Simmons & Heussen, 2009). Thus, there are theory-like effects, but they can vary in strength, raising potential questions about how central they are to concepts in general.

Theories and Development

The alleged centrality of theories to our concepts also seemed to be further reinforced by their role in development. Indeed, research in cognitive development was a primary impetus towards this view of concepts. Concepts appeared to change in the course of development in ways that reflected growing webs of belief. Concepts weren’t nodes in this network, they were clusters of nodes and links. There was a tacit, but not well articulated assumption, that these clusters were somehow bounded, like one of circles in Figure 1.

As the network of beliefs grew, the clusters of concepts changed, and new concepts emerged. Thus, web growth could spawn new concepts (new clusters) elaborate on old ones in quantitative ways (more links with the same sort of structure) or become more dramatically restructure in qualitative ways (new kinds of structures). There were nagging details about how to determine clusters, how to decide when a new concept emerged, and so on, but these seemed like details that could be worked out.

Five classes of problems

Other problems, however, were deeper and were more problematic; although here too, many researchers in the field felt that reasonable solutions to these would emerge with time. Most of these were also documented by Fodor (1998 (2005): Fodor & Lepore, 2002). Five classes of problems are particularly relevant.

First, there is “the lost in thought” problem. If concepts were really the same as theories of how all their components worked together, wouldn’t theories be too slow? Wouldn’t we get lost working through all these theoretical implications? Yet we use concepts quickly and effortlessly. How could our apparent speed of use of concepts be reconciled with the richly textured theories that had to be considered to use them appropriately?

A second problem was that there seemed to be too much change in the theories surrounding concepts, while the concepts seemed to stay the change or only change more modestly. When William Blake wrote “Tiger, Tiger” near the beginning of the 19th century, the state of biology was radically different from what it is today. Evolution was not yet on the horizon, molecular biology didn’t exist, and there were many popular misconceptions about the dispositions of tigers. Yet, it isn’t unreasonable to say that William Blake and Siegfried and Roy really meant pretty much the same thing when they referred to tigers. How could their tiger concepts be so similar if biological theories had changed so much?

A third problem was that of meaning holism. Where does one theory stop and another begin? As one pursues full and exhaustive explanations of just about any phenomenon one runs the risk of traversing the full extent of the web of beliefs to track down an additional explanatory insight. An exhaustive theory explaining everything about cars might pull on physical mechanics and even quantum phenomena, on chemistry and thermodynamics, on electricity and magnetism, on human physical and cognitive ergonomics (and from there to all of cognitive science and biology), on the economics and geopolitics of fuels, and so on. It would seem we might be at the mercy of a “cognitive butterfly effect.” Change a belief at some far off point in the web of belief and why couldn’t it traverse back and cause a shift in our core concepts of cars?

There is also the problem of conceptual combinations. If concepts can be understood in terms of the structure of the theories in which they are embedded, shouldn’t it be possible to use those structures to predict what happens when they are merged in conceptual combinations, such as pet fish, junkyard dog, and the like? At first glance, it appears that we are not well equipped to explain these combinations. Properties seem to emerge in ways that are not derivable from their alleged internal structures, whether those structures are understood as prototypes or as theories (Fodor and Lepore, 1996; 2002; see also Jönsson & Hampton, 2008). Moreover, it was not clear what it means to combine concepts when those concepts themselves are understood as parts of larger networks of beliefs. Are those beliefs modified or is the circle of interest simply expanded to include both of the constituents with the larger circle of beliefs equal to the new combination? (Neither of these seems very feasible). Moreover, to the extent that meaning holism is a problem, it is an even larger problem in knowing how to combine such massively extended networks.

Finally, there are difficult questions concerning how new nodes in the network of explanatory beliefs might emerge. What does it mean for a network to grow and how could one distinguish between the addition of new beliefs linking up familiar nodes and the actual emergence of new nodes? More concretely, when a child learns the concept of an Apple I-pad, is that child simply welding together a new constellation of beliefs about familiar concepts of electronic devices and music, or does a new node emerge which then can be used as another connector for beliefs? There appeared to be little consensus on how to implement such ideas.

Some Possible Responses to the Five Problems

These problems are hardly new (e.g., Fodor, 1998). For many researchers in cognitive science, however, these did not seem to be insurmountable obstacles to a view of concepts as based on theories. It is far beyond the scope of this paper to go through in any detail the possible ways to address these problems, but a brief mention of some potential ways out is useful because it helps to keep them in mind in considering what may a be a much more profound problem with the concepts-in-theories view.

The “lost in thought” problem might be solved by a mechanism of “pre-compiling”. Thus, people might not have to work out all the theoretical implications of a concept in real time and instead could build up a set of useful expectations based on their theories that could then be stored and used quickly in various situations. For example, I do not need to go through the full set of causal explanatory beliefs relating to birds every time I encounter a bird. Instead, over a longer period of time preceding any encounter with a bird, I may have for theory-based reasons, learned to weigh some features such as kind of wings, more importantly than others, such as whether a bird is facing to the right or the left when I first encounter it. Theories could guide one’s attention towards certain features over others and in doing so shift even what features are most typically noticed in conjunction with a category. In fact, through such methods as pre-compiling, adults at least can sometimes use theory-based information about features faster than mere frequency based information (Luhman, Ahn, & Palmeri, 2006).

The problem of theory change without concept change might be addressed by discounting the magnitude of theory change for most people. To be sure, the biological sciences have advanced massively in the past 100 years, but perhaps the layperson’s concepts of the biological world have changed to a far smaller extent and thus do not pose as much of a problem. If intuitive theories are ones in which concepts are normally embedded, there may be more continuity than change in their nature. There may also be less theory change in the folk sciences where, as we will see, the adult end states are often quite sparse. If the theories are sufficiently sparse, not as much has to change.

Meaning holism might be addressed by a kind of pragmatic pruning that gives locality in real time. Thus, while it is true that explanatory inferences can eventually branch out from a concept to almost anywhere in network of beliefs, perhaps we use some simple heuristics to keep holism in check, such as attaching an exponentially higher cost to traversing each extra link away from the concept we are trying to understand. Moreover, there has been a surge of interest in the philosophy of science to document somewhat similar heuristics used to localize phenomena and rule out some sets of causal relations through simplifying assumptions and idealizations (Elga, 2007; Godfrey-Smith, 2009; Strevens, 2008; Weisberg, 2007).

Conceptual combinations remain a tricky problem, but there are attempts to explain how coarser patterns in domains might actually be being brought into alignment in ways that do enable predictions (Johnson and Keil, 2000; Hampton, 1994). Also, people may have underestimated the extent to which prototypes can foster comprehension of many conceptual combinations (Jönsson & Hampton, 2008). Finally, there are neuropsychological suggestions of how constituents might be superimposed in systematic and predictable ways (Baron, Thompson-Schill, Weber & Osherson, 2010) as well as suggestions of ways in which perceptual simulations might be involved for at least some conceptual combinations (Wu & Barsalou, 2009).

Finally, we might be able to distinguish between core concepts that exist as primary nodes in a network and which exist from early infancy and other concepts that only exist in a derived form based on networks of beliefs to those core concepts. This might provide us with a principled way of deciding how new nodes emerge in the network of beliefs. Thus, there are arguments that core concepts have a special status and character in early development that enable researchers to distinguish them from later concepts that are formed out of these primitives (Carey, 2009; Spelke, 2000; Spelke & Kinzler, 2007).

There remain many major problems to be solved with each of these responses, but they do allow one to believe that the theory-based way of studying concepts could still survive. It seemed reasonable to develop these responses further given the many appeals of the concepts-in-theories view. A much bigger, problem, however emerged when attention turned to the actual complexity of naïve theories.

A Bigger Problem-Weak Theories

From my own perspective, the most difficult problem with the theory-based view of concepts emerged gradually through a series of studies that more frontally tried to ask about what intuitive theories really were like in the minds of others. We began to uncover truly devastating gaps in people’s knowledge - gaps that much of the time existed without people have any awareness of them. We called this lack of awareness an “illusion of explanatory depth” (Rozenblit & Keil, 2002). People tend to overestimate how well we can explain things, and children do so to an even more extreme degree (Mills & Keil, 2004). To show this, an experimenter simply asks participants how well they think they know how something works, and then subsequently ask them to explain it (with appropriate training on scales and other experimental design particulars). People are often shocked at how much worse their explanations are than from what they thought they knew. Interestingly, people tend to be much better at assessing their knowledge of how well they know facts, procedures (such as how to make international phone calls), and narratives (how well they know the particular plots of books or movies). In contrast, they are very poor at estimating their knowledge of how and why. The effect has also been found for estimates of understanding of political candidate’s explanations (Alter, Oppenheimer & Zemla, 2009).

The failure to recognize the shallowness of our explanatory understandings creates a problem. If the theories are much weaker than we think, how much work can they do? We can retreat to talk of framework theories, or core theories, and kindred kinds of notions (e.g., Gopnik & Wellman, 1992l; Wellman & Gelman; 1992;), but these retreats have their own serious problems. If you look at some of the theories that ascribed to young children – they are at best sometimes 3 nodes and 3 links. The very young child's theory of mind has been characterized as roughly: "I have desires that cause me to engage in actions." That is the entire "theory". A little older child might have the following: "I have beliefs. Beliefs cause desires. Desires cause me to engage in actions." Similarly, very early folk biology might be: "I believe in a vital force. That vital force helps me to move; and if there is some force left over, it helps me grow." (Inagaki & Hatano, 2002). If those simple components are all there is to framework theories, they are not going to do a lot for us in terms of articulating the structure of concepts. If there is more to framework theories, it is not clear what those additional details look like.

To make matters worse, we also have a high tolerance of contradictions. As has been shown repeatedly, people can believe rather large chunks of information that are completely contradictory to each other and not realize it until its explicitly pointed out to them (Chin and Brewer, 1993). To flesh out two examples a bit more, many adults will state that they believe that animal kinds have fixed essences yet also state that they believe in gradual evolution through natural selection (Shtulman & Schulz, 2008). Yet, natural selection can only operate on a species that is defined as a distribution of traits rather than having some set of necessary defining features. In the realm of describing human behavior, the same person can state that that human behavior is s result of strict causal determinism and not free will while also later stating that people are morally responsible for their actions (Nahmias, Coates, & Kvaran, 2007).

Perhaps the happy existence of such contradictions in one person's mind can be seen optimistically as a sign that holism cannot really be at work in real minds (if you automatically traversed the full net of beliefs you would be aware of all the contradictions), but it makes all the more difficult any idea that concepts emerge out of a richly articulated and coherent set of theory-like beliefs. It has been repeatedly suggested that people strive towards coherence and use it to structure their beliefs and certainly some preferences for coherence occur (Thagard, 2000), but at the same time, there are clearly other factors that limit the reach of coherence.

One extreme reaction to contradictory beliefs is to say that there is no overall linking structure to our beliefs, that knowledge falls apart into little tiny pieces. DiSessa (1993; Di Sessa, Gillespie, & Esterly, 2004), for example, argues that our knowledge may be nothing more than a collection of phenomenal primitives, or "p-prims". It may, however, not be necessary to abandon all structure. There may be ways we can talk about a more relational structure, but it just can't be like a traditional "theory".

The core problem may be the following: There is, or most of us, no theoretical difference between lions and tigers. I know they are different, I think I know they mean different things, but I cannot provide a theoretical reason that distinguishes them (Fodor, 1998). How then, can concepts be differentiated in terms of the theoretical frameworks within which they are embedded? I may believe that lions and tigers differ for interesting reasons related to theoretical notions in biology, but do not actually know any of those differences. I do have a weak sense of what matters for the difference, namely DNA and the ways a genetic code leads to proteins and other products that in turn give an animal its properties; but I have no idea of what it is about lion DNA that makes it a lion and not a tiger. So, at best, I have some hunches about the kinds of things that make a difference in distinguishing between these two animal kinds; but I cannot provide any details about them whatever.

In other cases, an understanding of what would make a difference may be even weaker. I may know that something tiny inside gold makes it different from silver, but know nothing at all about the nature of that tiny micro-structural component beyond the idea that it somehow causes gold to behave like gold, and silver like silver. This is what we might call “blind faith essentialism” in its purest form.

Do weak theories force us to accept notions of concepts as atoms, with no constitutive structure (see for example, Fodor 1998)? One strong reason to resist that move is the issue mentioned earlier that concepts travel in groups. If my concept of MOTHER changes, so does my concept of CHILD. Concepts can be mutually parasitic off each other for their meanings in ways that seem to defy the idea that they have no internal structure. In addition, there is the critical centrality of how and why. Why it is that features that co-occur equally for instances of concepts are either ignored or attended to because they fit with some notion of how and why? Similarly why are some correlations between features ignored or embraced in ways that reflect intuitions about their causal centrality to a domain? One way out may be to propose that we do track causal and relational structures in the world in a way that is less theory-like than we used to think. This way of tracking causal structure may often not even be belief-like, but might still work in a manner that supports concept acquisition and use and is part of the concepts themselves.

Tracking Causal Structure

What are some of the alternative ways that we do track causal structure? One very simple way involves knowing what kinds of property types are likely to do important causal work in a domain. Very young children, infants, and even some other primates seem to know, for example, that when one is thinking about tools, shape is going to usually matter more than color. In contrast, when one is talking about foodstuff, color usually matters more than shape (Keil, 1994; Keil et. al., 1998; Keil, 2010; Santos et al., 2002). These intuitions about the relative importance of property can be used to construct “ causal relevancy profiles “ that help constrain though about members of categories.

A different level could be understood as that of causal powers, knowing that certain classes of things have certain dispositions to produce certain effects. For example, pre-verbal infants (12 month olds) know that intentional agents have the power to create order out of disorder while non-intentional ones do not (Newman, Keil, Kulhmeir & Wynn, 2010). In that set of studies, infants see a pile of disordered blocks, a barrier comes up, comes down again, and reveals the blocks in a neatly ordered array. Infants think that only an intentional agent can bring about such a change, not something non-intentional, like a rolling ball. If the event is reversed and blocks go from an ordered array to a disordered one, the infants understand that both intentional and non-intentional agents have the power to bring that second kind of change about (Newman et. al., 2010).

At yet another level of causal analysis, young children can think about what kinds of functional interpretations make sense with different sorts of kinds. Thus, if preschoolers are invited to ask questions about novel artifacts and novel animals that they have never seen before, they approach them very differently in terms of the kinds of causal regularities they think are at work (Greif, Kemler-Nelson, Keil, & Guiterrez, 2006). For a novel artifact, they are likely to ask what the artifact as a whole is for: “What’s that for?” There are far fewer spontaneous questions of this sort about novel animals – they are unlikely to ask what the animal as a whole is for. The children will, however, ask about what parts of an animal are for. “What are their claws for” or “What is this long beak for?” Even if they have no idea what an animal or machine is called and have never seen it before, they seem to have quite sophisticated expectations about the kinds of relational and causal patterns that go with different domains.

Children use these notions of causal patterns to guide their intuitions about the division of cognitive labor. Thus, even if they don’t know who knows what, they know there are different kinds of experts out there that they can defer to (Keil, Stein, Webb, Billings & Rozenblit, 2008). This knowledge may be critical to how they set up concepts when they have almost none of the details themselves. It may provide a sense of the kinds of properties and relations that are important and which experts are likely to know about such matters.

Concepts as Chimeras

Perhaps concepts are best understood as chimeras. They are not simply prototypes, they are certainly not definitions, and they are not theories, yet elements of each of these seem to be at work. There may also be a rich causal relational structure that is part of the story. In addition, it may be necessary to incorporate into concepts and their formation the idea of “locking” (Fodor, 1998). People to lock onto objects in ways that often do not have a rich underlying propositional structure serving as support. They seem to use a grab-bag of components to stably refer, ranging from probabilistic tabulations of features associated with categories to evaluations of trustworthiness in a social network of deference in order to ground their use of words. For example, many people may freely use the word “wombat” but have no idea of wombat perceptual or behavioral features. Yet, they arguably lock onto wombats by being reliably plugged into a network of deference and expertise. Even young children are surprisingly sophisticated at linking abstract causal-relational patterns to broad domains, such as social interactions, artifacts, intentional beings, mechanical agents and the like, and they use those to guide categorization, deference and learning. Locking in this way, with all its variations of methods, may be in place from the earliest moments of word learning.

Consider a concrete example of how locking might use such components. One of the animals in Figure 2 is a weasel and the other is a ferret.

Even though one may have no knowledge of any specific features that distinguish these two kinds of animals, one might nonetheless firmly believe that one has both of those concepts. What does it mean to say that a person has two distinct concepts of a weasel and of a ferret yet has absolutely no idea what the difference is between them? One answer suggests that such a person thinks he knows who knows. He thinks he knows who the appropriate experts are and how to access them. He may be mistaken, but his beliefs in such experts are enough to convince him that he "has" two separate meanings (see Putnam, 1975).

One part of this knowing who knows may involve the notion of "sustaining mechanism" (Laurence and Margolis, 2000; Margolis & Laurence, 2003). Sustaining mechanisms are mental operations that enable our concepts to lock onto the appropriate classes of entities. That they exist is self-evident – their relation to concepts is more controversial. Laurence and Margolis discuss three kinds of sustaining mechanisms: 1. those that are theoretical and allow you to lock onto objects, 2. those that are based on deference to experts, and 3. those that are based on a syndrome, or something like a prototype. In most cases, the issue may not be which one of these at work. Instead, all three may usually be involved at the same time in instances of locking. Here is why. Our theories are too weak to work on their own. But often, when we decide whether something is a ferret or a weasel, what we are doing is having a crude notion of who the right expert is and using those weak theories to find the right expert realm. We then use that idea of appropriate expert realm to help us defer to others, and we also then use that deference to determine which features of the syndrome to attend to. No one sustaining mechanism may be enough in most real-world cases.

A critical question is whether the sustaining mechanisms are part of the concept itself, as opposed to just being tools that helps us lock. Are sustaining mechanisms like a microscope that helps us lock but which should not be confused with the locked thing? There are reasons to resist that conclusion. It may be that, even for microscopes, part of what it means to have the concept BACTERIA is to know what kind of tool a microscope is. It is not just an abstract tool in the most general sense. One has to know, for example, that it is a way to get information about invisible microscopic structures, that the microscope has some causal efficacy. Moreover, that understanding may be critical to my concept of BACTERIA. The same holds for experts and thinking of experts as sentient "tools" like the microscope. You cannot point any expert at any object and expect to get the right answer. You have to know what kind of expert you are talking about. You have to know that there are different kinds of experts, who have different specializations in different causal regularities.

It turns out that very young children are sensitive to some of the ways expertise works. Even by the age of 3 or so, they start to know that there are different kinds of experts in the world (Lutz & Keil, 2002; Danovitch & Keil, 2004; Keil et. al., 2008). They know that adults are not omniscient and have different zones of cognitive competence. These children must have some mastery of the causal structure of the world to even be able to engage in the practice of deference and the use of expertise. Thus, it can be shown that they are relying on abstract causal schemas to solve the division of cognitive labor problem (Keil et. al., 2008).

Consider how this all might come together in the acquisition of the concept of a CARBURETOR. This truly a speculative account, but it will serve to demonstrate how the idea might work. (It may be an especially interesting case because soon carburetors will no longer exist. They are vanishing due to their replacement by fuel injection systems). We might hear the word “carburetor”, and we might then hypothesis-test whether it is an artifact or a natural kind. Here, we might grant that the notion of artifact is innate as well as perhaps the simplest sense of natural kind (not the more complex sense inherent in the philosophy of science). We might then quickly map the word onto the artifact domain; there are lots of perceptual heuristics to tell whether something is an artifact or not (e.g., Levin, Takarae, Miner & Keil, 2001). This locking onto a broad category such as artifact raises the question of whether there is whether there is a notion of differentiating sustaining mechanisms that allow us to go beyond those broad categories. Thus, initially, our locking is so crude that we really can’t have different kinds of concepts below a certain level, but we then come to have them as our locking mechanisms become more refined. This is one way in which concepts become formed out of earlier substrates.

In these cases, to have differentiating concepts is to have differentiating sets of sustaining mechanisms. Those increasingly fine-grained sustaining mechanisms may be what allow us to be more and more successful in picking out appropriate categories. Moreover, the mechanisms may not be sharpened just by hypothesis-testing. They may proceed instead by becoming more and more sensitive to the kinds of causal patterns that are associated with different kinds of experts, that is with increasing ability to pick out different kinds of regularities in the world.

In such accounts, concepts might still be considered as autonomous atoms; however, the sustaining mechanisms are so linked to them that they may not be able to be separated from the concepts proper. What then is the role of theories in all of this, especially if theories are so weak? How do weak theories strongly constrain? It may be that weak theories do set up boundary conditions on concepts, namely that if you don’t have the abstract causal patterns that tell you the proper domain of a concept, you simply don’t have the concept. Someone who thinks that carburetors do have micro-structural essences, that they have no overall function, or are non-physical, simply may not have the concept. Plenty of more detailed beliefs could be wrong, but someone cannot go so far as to violate these overarching patterns. Someone could have mistaken beliefs about the shape, the local function, or the material substrate of carburetors, but they are not licensed to have any mistaken beliefs at all (see also Keil, 1979). A different set of beliefs will be at work for living kinds, such as that they do have micro-structural essences and that they do not have functions as wholes even as their parts can have functions People will therefore have different expectations for living kinds that also cannot be violated. Weak theories do not directly tell lions apart from tigers. But they may provide guidance to deference and ways of access to information. They may guide the construction and the differentiation, of domain-specific sustaining mechanisms and, in this way, are involved in accounts of concept formation.

It is difficult to distinguish this sort of account from Fodor’s atomism in which sustaining mechanisms are exterior to the concepts themselves. There are no easy algorithms for telling what is in the concept proper versus what might be in a distinct enabling cognitive structure. But there seem to be powerful constraints in terms of causal-relational patterns that we all pick up on from an early age and which we see as fitting with high-level domains. These are domains like living kinds, artifacts, and intentional agents. These domains are not like traditional theories. Indeed we may sense many of the patterns associated with those domains at a highly implicit level that is only revealed when we look at what information children and adults must be aware of to solve certain tasks. Somehow, we have to learn how to look at equally typical features and weigh them differentially because of beliefs about their causal centrality, and then use that information to guide locking. As a first pass, all of that might still be best thought of as part of the concept proper.

A final issue concerns whether compositionality clearly argues against having sustaining mechanisms being parts of the concepts themselves. It is not at clear, for example, that even the most exhaustive analysis of sustaining mechanisms for two concepts would allow us to explain how they compose. What, for example, is the relation between the sustaining mechanisms for red things and for shirts, that allows us to pick out red shirts? If the mechanisms are parts of the concepts, should not the structure as revealed by such mechanisms enable us to predict the ways in which concepts compose? Given how difficult it is for any approach to provide full accounts of composition, it is not clear that sustaining mechanisms are especially vulnerable. After all, we think that hydrogen and oxygen gas molecules are usefully described in terms of their constituent atoms and bonding relations even though it remains a mystery to fully explain the properties of water from their combination. The inadequacy of some alleged constituents of concepts to fully explain conceptual combination may therefore not be sufficient grounds for dismissing their roles as constituents of concepts. If other phenomena, such as conceptual change and concept formation, can be usefully understood in that matter, then perhaps that is enough.

Conclusion

I have tried to sketch out how the concepts-in-theories view, while very appealing, also has serious limits when one considers the minimalist nature of many intuitive theories. In the end, the theory-like effects associated with so many aspects of concept acquisition and use, argue for still trying to have a way in which causal explanatory structure is part of concepts. It may be that such a route exists through the ways in which weak theory-like structures guide notions of expertise, deference, and feature centrality. It is far too early, however, to know whether this sort of program will provide a fully satisfactory answer to the problem of what sorts of structures make up concepts and can explain their formation in development and learning.

Acknowledgments

Note: Some of the research reported on in this article was supported by NIH grant # R37HD023922 to F. Keil

References

Ahn W. Why are different features central for natural kinds and artifacts? The role of causal status in determining feature centrality. Cognition. 1998; 69:135–78. [PubMed: 9894403]

Alter, AL.; Oppenheimer, DM.; Zemla, JC. A construal-based mechanism for the Illusion of Explanatory Depth. Paper presented at the Society of Judgment and Decision Making annual conference; Boston, MA. 2009.

Baron SG, Thompson-Schill SL, Weber M, Osherson D. An early stage of conceptual combination: Superimposition of constituent concepts in left anterolateral temporal lobe. Cognitive Neuroscience. 2010; 1:44–51. [PubMed: 24168244]

Boyd, R. Homeostasis, species, and higher taxa. In: Wilson, R., editor. Species: New Interdisciplinary Studies. Cambridge, MA: MIT Press; 1999. p. 141-85.

Carey, S. Conceptual Change in Childhood. Cambridge, MA: Bradford Books, MIT; 1985.

Chinn CA, Brewer WF. The role of anomalous data in knowledge acquisition: A theoretical framework and implications for science instruction. Review of Educational Research. 1993; 63:1–49.

Danovitch J, Keil FC. Should you ask a fisherman or a biologist? Developmental Shifts in Ways of Clustering Knowledge. Child Development. 2004; 75:918–931. [PubMed: 15144494]

di Sessa AA. Toward an epistemology of physics. Cognition and Instruction. 1993; 10:165–255.

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

Elga, A. Isolation and folk physics Causation, Physics, and the Constitution of Reality: Russell's Republic Revisited. Price, Huw; Corry, Richard, editors. Oxford University Press; 2007.

Fodor JA. Some reflections on L. S. Vygotsky's thought and language. Cognition. 1972; 1:83–95.

Fodor, JA. The Language of Thought. NewYork: Thomas Crowell; 1975.

Fodor, JA. Concepts: Where cognitive science went wrong. Oxford: Clarendon Press; 1998.

Fodor. meaning and world order. 2005

Fodor JA, Lepore E. The Red Herring and the Pet Fish: Why Concepts Still Can't Be Prototypes. Cognition. 1996; 58:253–270. [PubMed: 8820389]

Fodor, JA.; Lepore, E. The compositionality papers. Oxford: Clarendon Press; 2002.

Gelman, SA. The essential child: Origins of essentialism in everyday thought. Oxford: Oxford University Press; 2003.

Goldman, A. Pathways to knowledge. Oxford: Oxford University Press; 2002.

Godfrey-Smith P. Models and Fictions in Science. Philosophical Studies. 2009; 143:101–116.

Gopnik A, Wellman HM. Why the Child's Theory of Mind Really is a Theory. Mind and Language. 1992; 7:145–71.

Greif M, Kemler-Nelson D, Keil FC, Guiterrez F. What do children want to know about animals and artifacts?: Domain-specific requests for information. Psychological Science. 2006; 17(6):455–459. [PubMed: 16771792]

Gopnik, A.; Meltzoff, A. Words, thoughts and theories. Cambridge: MIT Press; 1997.

Greif ML, Kemler Nelson DG, Keil FC, Gutierrez F. What do children want to know about animals and artifacts? Domain-specific requests for information. Psychological Science. 2006; 17:455–459. [PubMed: 16771792]

Hampton JA. Conceptual combination: Conjunction and negation of natural concepts. Memory & Cognition. 1997; 25:888–909. [PubMed: 9421575]

Hampton, JA. Concepts in Human Adults. In: Mareschal, D.; Quinn, P.; Lea, SEG., editors. The Making of Human Concepts. Oxford: Oxford University Press; 2010. p. 293-311.

Hampton JA, Storms G, Simmons CL, Heussen D. Feature Integration in Natural Language Concepts. Memory & Cognition. 2009; 37:1721–30.

Harris, P. What do children learn from testimony?. In: Carruthers, P.; Stich, S.; Siegal, M., editors. The Cognitive Basis of Science. Cambridge: Cambridge University Press; 2002. p. 316-334.

Hayes BK, Thompson SP. Causal relations and feature similarity in children's inductive reasoning. Journal of Experimental Psychology: General. 2007; 136:470–484. [PubMed: 17696694]

Heit E. Properties of inductive reasoning. Psychonomic Bulletin & Review. 2000; 7:569–592. [PubMed: 11206199]

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

Johnson, C.; Keil, FC. Theoretical Centrality vs Typicality in Conceptual Combinations. In: Keil, FC.; Wilson, RA., editors. Explanation and Cognition. Cambridge: MIT Press; 2000. p. 327-360.

Jönsson MC, Hampton JA. On prototypes as defaults. (Comment on Connolly, Fodor, Gleitman and Gleitman, 2007). Cognition. 2008; 106:913–923. [PubMed: 17433281]

Keil, FC. Semantic and conceptual development: An ontological perspective. Cambridge, MA: Harvard University Press; 1979.

Keil, FC. Concepts, Kinds and Cognitive Development. Cambridge, MA: MIT Press; 1989.

Keil FC. Explanation Based Constraints on the Acquisition of Word Meaning. Lingua. 1994; 92:169–196.

Keil FC. The Feasibility of Folk Science. Cognitive Science. 2010; 34:826–862. [PubMed: 20625446]

Keil FC, Smith CS, Simons D, Levin D. Two dogmas of conceptual empiricism. Cognition. 1998; 65:103–135. [PubMed: 9557380]

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]

Kuhn, TS. The Essential Tension Selected Studies in Scientific Tradition and Change. Chicago: University of Chicago Press; 1977.

Laurence S, Margolis E. Radical Concept Nativism. Cognition. 2002; 86:22–55.

Levin DT, Takarae Y, Miner A, Keil FC. Efficient visual search by category: Specifying the features that mark the difference between artifacts and animals in preattentive vision. Perception and Psychophysics. 2001; 63:676–697. [PubMed: 11436737]

Luhmann CC, Ahn W, Palmeri TJ. Theory-based categorization under speeded conditions. Memory & Cognition. 2006; 34:1102–1111. [PubMed: 17128608]

Lutz DJ, Keil FC. Early understandings of the division of cognitive labor. Child Development. 2003; 73:1073–1084. [PubMed: 12146734]

Machery, E. Doing without concepts. Oxford, England: Oxford University Press; 2009.

Mandler, JM. The foundations of mind. Oxford, England: Oxford University Press; 2004.

Margolis, E.; Laurence, S. The Blackwell Guide to Philosophy of Mind. Stich, S.; Warfield, T., editors. Blackwell Publishers; 2003. p. 190-213.

Mills CM, 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]

Mills CM, Keil FC. The development of cynicism. Psychological Science. 2005; 16:385–390. [PubMed: 15869698]

Murphy, GL. The big book of concepts. Cambridge, MA: MIT Press; 2002.

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

Newman GE, Keil FC, Kuhlmeier V, Wynn K. Early understandings of the link between agents and order. Proceedings of the National Academy of Sciences. 2010; 107:17140–17145.

Proffitt JB, Coley JD, Medin DL. Expertise and category-based induction. Journal of Experimental Psychology: Learning, Memory, and Cognition. 2000; 26(4):811–828.

Putnam, H. The meaning of “meaning”. In: Gunderson, K., editor. Language, mind, and knowledge. Vol. 2. Minneapolis: University of Minnesota Press; 1975. p. 131-193.

Quine, WVO.; Ullian, JS. The web of belief. New York, NY: Random House; 1978.

Rehder B, Kim SW. How causal knowledge affects classification: A generative theory of categorization. Journal of Experimental Psychology: Learning, Memory, & Cognition. 2006; 32:659–683.

Rogers, TT.; McClelland, JL. Semantic Cognition: A Parallel Distributed Processing Approach. Cambridge, MA: MIT Press; 2004.

Rozenblit LR, Keil FC. The missunderstood limits of folk science: An illusion of explanatory depth. Cognitive Science. 2002; 26:521–562. [PubMed: 21442007]

Santos, LR.; Hauser, MD.; Spelke, ES. Domain-specific knowledge in human children and non-human primates: Artifact and food kinds. In: Bekoff, M.; Allen, C.; Burghardt, G., editors. The Cognitive Animal. Cambridge: MIT Press; 2002. p. 205-216.

Shtulman A, Schulz L. The relation between essentialist beliefs and evolutionary reasoning. Cognitive Science. 2008; 32:1049–1062. [PubMed: 21585442]

Smith C, Carey S, Wiser M. On differentiation: A case study of the development of the concepts of size, weight, and density. Cognition. 1985; 21:177–237. [PubMed: 3830547]

Spelke ES, Kinzler KD. Core knowledge. Developmental Science. 2007; 10:89–96. [PubMed: 17181705]

Spelke ES. Core knowledge. American Psychologist. 2000; 55:1233–1243. [PubMed: 11280937]

Strevens, M. Depth: An Account of Scientific Explanation. Harvard: Harvard University Press; 2008.

Thagard, P. Conceptual revolutions. Princeton, NJ: Princeton University Press; 1992.

Thagard, P. Coherence in Thought and Action. Cambridge, MA: MIT Press; 2000.

Vosniadou S, Brewer William F. Theories of knowledge restructuring in development. Review of Educational Research. 1987; 57:51–67.

Weisberg M. Three Kinds of Idealization. The Journal of Philosophy. 2007; 104(12):639–59.

Wellman HM, Gelman SA. Cognitive development: foundational theories of core domains. Annual review of psychology. 1992; 43:337–375.

Wiser, M.; Carey, S. When heat and temperature were one. In: Gentner, D.; Stevens, A., editors. Mental models. New York: Academic Press; 1983. p. 75-98.

Wisniewski EJ, Medin DL. On the interaction of theory and data in concept learning. Cognitive Science. 1994; 18:221–281.

Wu LL, Barsalou LW. Perceptual simulation in conceptual combination: Evidence from property generation. Acta Psychologica. 2009; 132:173–189. [PubMed: 19298949]

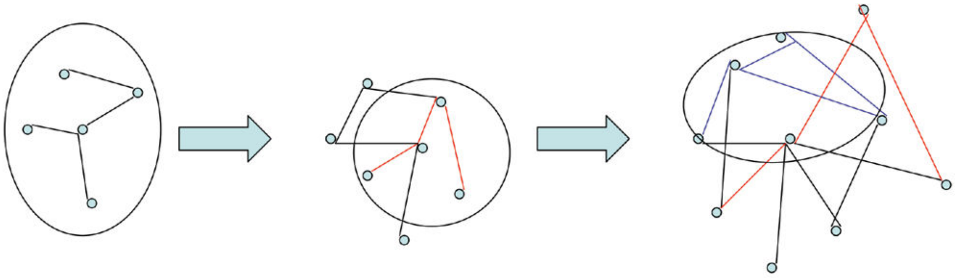

Figure 1.

Are concepts regions in networks of beliefs (shown here by ovals), with conceptual change occurring as those networks expand?

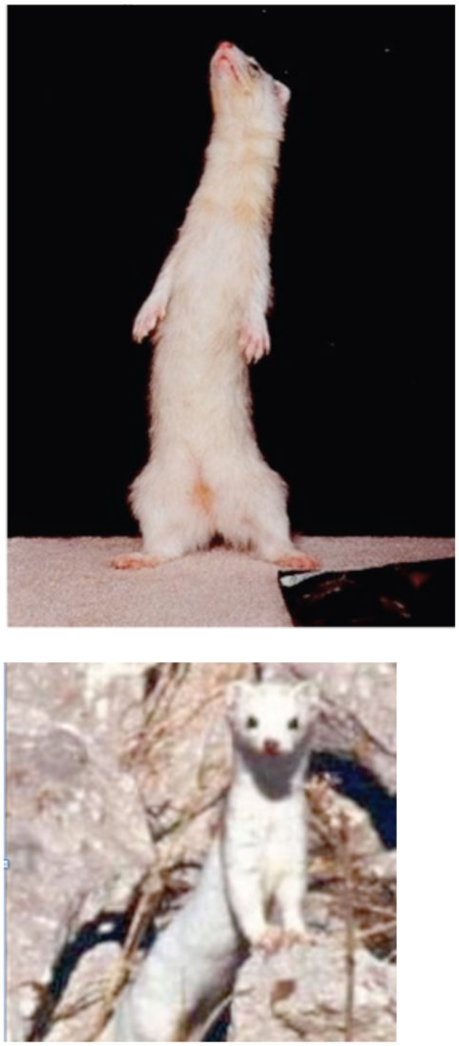

Figure 2.

Weasels versus ferrets. One of these creatures is a ferret and one is a weasel. We may believe these to be importantly different kinds, but may have no idea of any particular differences between them.