Working paper: do not distribute or quote without permission!

This is a relatively breezy version of an exploration of some issues about how to provide a theory of concepts and conceptual semantics. I have also written more conventional versions of some of this material (without the Three Bears motif), though those are set in a broader context. "Logical and Psychological Views of Semantics" addresses these issues in the context of arguing that semantics can be approached from a psychological side that ought not to be "logicized" much as Frege argued that logic ought not be psychologized. "Concepts and Biology: How (not) to give a theory of concepts" devotes more time to a sketch of what the history of protoconcepts across phyla might look like, and deals with material cognate with this article in an introductory section which identifies several problems with previous approaches to concepts.

Goldilocks Searches for a Conceptual Semantics

Steven Horst
Wesleyan University
Shorst@wesleyan.edu

I shall begin with a brief fairy tale: Goldilocks, searching in the Enchanted Forest for a theory of conceptual semantics, is confronted with a problem. The Mama Bear's theory requires too much of such a theory. The Papa Bear's theory does too little. But is there a Baby Bear theory that is just right?

Philosophers interested in a theory of concepts and conceptual semantics tend to fall into two camps. Those in the first camp—the Mama Bears in the fairy tale—are interested in a theory that is, in a broad sense, *reductive*. That is, they want an account of concepts and conceptual semantics in terms of something else, most likely in the form of a set of necessary and sufficient conditions for what it is to be a concept *simpliciter*, and/or what it is to be the concept-of-X. On the current scene, Mama Bear has a taste for reductive theories that are also naturalistic—that is, theories that account for mentalistic notions like "concept" and "meaning" in entirely non-mentalistic and non-intentional terms taken from the natural sciences. Philosophers in the second camp—the Papa

Bears—insist that a theory of concepts and of conceptual semantics cannot be framed in isolation from other mental phenomena, such as consciousness or inferential role. Some Papa Bears hold additionally that concepts cannot be properly understood without reference to *social* phenomena, particularly participation in a linguistic community, as well.

It seems to me that it is reasonable to feel the pull of each of these positions. However, the positions are in important ways incompatible with one another, and so the very things that make one position attractive are missing in the other, and vice-versa. Unless Goldilocks can find a third approach that incorporates the best of both of these, she is faced with a dilemma of choosing between two approaches, each of which is promising in some respects but not in others.

Reductive Views (Mama Bear)

Mama Bear is a reductionist. She believes that science is in the business of reduction, and that philosophical analysis is in the business of providing necessary and sufficient conditions. For her, an account of concepts would be something that could be put into the form, "X is a concept iff …" And an account of conceptual semantics would be something that could put into the form "X is the concept-of-Y iff…" Moreover, the lacunae must not be filled in some question-begging way that presumes some prior stratum of conceptuality or semantics, nor in a way that presumes some other mental phenomenon (like consciousness or the norms of inference) that is itself left unreduced.

Reductionist theories might be largely philosophical: for example, the view that concepts are something like syntactically-individuated representation-types in a language

of thought, whose semantic properties are determined by causal covariation in perception. (Fodor, 1987) Alternatively, they might be scientific: say, a theory that views concepts as a construction out of successive stages of evolutionary adaptations over phylogenetic history (Millikan 1984, 2000) and reconstructable today through comparative ethology and neuroscience.

Theories of this sort—when you can actually lay hands on one—have a number of virtues. Philosophically, they are "neat". That is, they say what they say fairly clearly, and so it is comparatively easy to assess their coherence, explanatory power and truth. Likewise, it is comparatively easy to frame alternatives and revisions to a theory cast in terms of necessary and sufficient conditions. They also agree with (what is presently) a deeply-seated atomistic intuition: that one understands more complex phenomena by understanding the relations of simpler phenomena. Thus, as Hobbes suggested [], a train of reasoning ought to be understood in terms of the constituent thoughts that make it up and the process through which they are ordered, and not vice-versa.

This sort of theory also has a distinct appeal for the cognitive scientist. For one thing, it can be made to highlight continuities between human and animal cognition. No one thinks (*pace* the antics of philosophical bears in fairy tales) that other terrestrial species possess *all* of the same cognitive abilities that we possess, of course. But it is clear that there are strong continuities between nearby species: it makes perfect sense to say that my cat "sees a bird", "wants his breakfast" or "likes to lie in the sunlight." If he cannot also see a bird *as* a member of an endangered species, want expensive catfood (*qua* expensive), or appreciate the play of light on cut crystal, these differences pose interesting research questions about what additional accretions there are to human

conceptuality that are lacking in cats. And if there are levels of the animal kingdom where we begin to doubt whether it is right to speak of "seeing", "wanting" or "liking" altogether, this too bespeaks of earlier stages of accretion of cognitive abilities. Somewhere along this line it is appropriate to start speaking of concepts and semantics. At other points along this line—presumably later—it is appropriate to start speaking of language or reasoning. And somewhere along this line—perhaps earlier, perhaps later—it is appropriate to begin speaking of consciousness.

The cognitive scientist also likes the fact that reductive accounts are able to exploit empirical knowledge about the parts of the brain that are implicated in various forms of cognition. One can, for instance, take a fact about cognition—such as the fuzziness of concepts, which seemed anomalous when one approached semantics from the standpoint of the logician—and give them what seem to be straightforward explanations. One might say, for example, that the fuzziness of conceptual semantics is a consequence of the fact that concepts are implemented through a neural network architecture that is optimized (and selected) for efficient learning of distinctions, and that the fuzziness of the resulting concepts can be traced to features of the network architecture. (One might then go on to ponder how so many 20th century analytic philosophers had theorized about semantics as though concepts were designed with syllogistic reasoning in mind.)

The Goldilocks in me is quite attracted to the ability to give explanations of these sorts. There are surely things about cognition that are subject to explanation through cognitive science. And if analytic philosophers were right to follow Frege in rejecting

psychologistic approaches to logic, I should urge that they also resist the urge to pursue logicist approaches to psychology.

Yet there are philosophical problems with this approach as well. If we separate an account of concepts from "later" developments like reasoning, language and consciousness in this way, we begin to be committed to things that seem counter-intuitive. If we are just thinking in a commonsensical way about feline cognition, it may seem natural to speak of the cat "having the concept of a bird." But if we consider the conditions of concept-identity philosophically, this begins to seem more problematic. Does the cat have the *same* concept that I do? Here the intuitive answer would seem to be *no*. My concept BIRD would seem to be partially constituted by the various inferential commitments that I associate with it. Someone who lacked or denied enough of these would at some point cross the line into having a different concept, even if she used the same English word and used it to track the same objects in the world. And my cat is almost certainly more of a logical and inferential alien to me than just about any human being. Indeed, he is innocent of certain important forms of inference altogether. So whether you make this point with two humans or humans and non-humans, or even the same human at different times, there are issues of how we individuate concepts that are at odds with the reductive approach.

One can make a similar point in the case of concepts that involve linguistic deference. [] In such cases, my ability to use, or indeed to possess, a concept is derivative from the abilities of an expert. I can use words like 'quark' meaningfully, for example, even though I do not have a firm grasp of subatomic physics. And likewise I have a concept QUARK. I can possess such a concept because the word 'quark' has

entered the language by way of the usage of English-speakers who *are* experts in subatomic physics and who have a more robust concept. I can then have a minimal concept of quarks which means something like "the things that the physicists call 'quarks'." Here, my possession of some concepts turns out to be dependent upon a larger web of language in which my concept-acquisition and –use are embedded. Non-linguistic beings are not candidates for possession of concepts acquired in this way.

One might even wonder whether it makes sense to speak of *concepts* at all in cases where there is *too* little in the way of connection with language, inference or consciousness. Any notion we might have that a creature had a concept of some sort might very well be undercut entirely if we were to additionally suppose that it could *infer nothing at all* from it, or that it *could not be made part of a conscious thought*. Here the very strategy of pursuing cognitive continuities down the phylogenetic tree suggests that we are faced with a situation where we cannot simply provide a philosophical analysis of the existing notion of "concepts", but must go beyond our present analytical tools and make some additional distinctions.

Concepts, Language, Reasoning and Consciousness (Papa Bear)

Other theorists have taken this relation between concepts and other aspects of our mental lives as a starting point, holding that conceptual semantics is either derivative from, or at least cannot be divorced from, something else: the network of public language [Wittgenstein, Blackburn], inferential role [Brandom], or consciousness [Searle, Horst]. Indeed, one might even take all of these factors together as a seamless whole, and hold

that one cannot break into the intentional-linguistic-conscious-inferential circle from the outside, as the reductionist wishes.

This view also has a strong appeal. Insofar as we have a notion of unconscious, inarticulate or inexplicit concepts, we have to cash this out as a notion of things that might potentially be made conscious, articulated in a language, and whose inferential commitments might be made explicit. We can only view them by "looking backwards" through the lens of our own conscious and articulate adult conceptuality. When we speak of the "concepts" of frogs, cats or even human infants, we know at some level that we are applying terms that are *aptly* applied to us to things to which their application is less apt, and at some point along this continuum we are bound to find ourselves either anthropomorphizing or using the intentional and semantic vocabulary in a metaphorical or extended sense. Additionally, all of the concerns raised in the previous section as problems for the reductive view can also count as intuitive support for this more holistic view. ("Holistic" in the sense of the totality of our mental lives, not in the sense of holistic theories of meaning.)

But for the cognitive scientist, or the philosopher who wishes to be able to use insights from cognitive science to provide any kind of explanations of mental phenomena, this view also has its problems. Indeed, these are almost exactly the mirror-image of the benefits of the reductive view. The holistic view completely obscures any continuities between human and animal cognition. And while one might reasonably say that it is part of our very notions of "thought", "concept" and "meaning" that they are bound up with the possibility of consciousness, articulation and explicitness, this does not vitiate the comparative ethologist's point. Even if you reserve *these* terms for our own

case—or at least the case of beings that possess the same general features of language, consciousness and inference that we possess—there are nevertheless continuities of *other* sorts between us and creatures that lack these features. And the continuities are not *simply* "of other sorts". Rather, they are closely connected with features that are deeply a part of our own cognition. My cat cannot make anything explicit, and may or may not experience qualia. But there is surely a great deal of continuity between what I call "seeing a bird" and "wanting breakfast" in him and what I mean when I apply those words in my own case. And if Papa Bear's approach to concepts and conceptual semantics doesn't have room to talk about this continuity, so much the worse for Papa Bear's approach to semantics, says Goldilocks.

Likewise, the Papa Bear approach does not seem to be able to accommodate ways of explaining features of cognition, like the fuzziness of concepts, in terms of features of human cognitive architecture, such as formal properties of neural networks, or as features that are a process of natural selection. If one is not entitled to speak of *concepts* at all until one reaches human beings, it is hard to see how one can then explain features about concepts *per se* by appealing to features shared with, say, all vertebrates, and selected at a much earlier stage of evolutionary history.

Goldilocks's Wish List

In the end, Goldilocks is not satisfied with either sort of approach to concepts. One approach allows her to give some of the explanations that she wants to be able to give; but that approach ends up offering these as reductive theories that are unable to really explain everything about conceptual semantics: something could fulfill the terms

set out in these cognitive theories and yet fail to be a *bona fide concept* because it was not a conscious being, or incapable of inference, or not able to be part of a linguistic community. The other approach seems to do better justice both to our ordinary *usage* of words like 'concept' and 'meaning', and to our *notions* of what concepts and conceptual semantics must be like; but it does so in a way that seems to afford us no room to say anything *about* concepts *per se* unless it applies precisely to conscious, linguistic, reasoning beings.

What Goldilocks would *like* would be an approach to conceptual semantics that would do justice to our usage and intuitions, and yet also permits us to explain things *about* conceptuality and semantics, in a *non*-reductive way, using insights from the empirical sciences of cognition. And part of the problem seems to rest in something that the two approaches seem to agree on: *The problem with these old bears*, opines Goldilocks, *is that they assume that there is a forced choice between explanations of concepts that provide sufficient conditions for their intentional and semantic properties and no explanations at all that appeal to non-intentional and non-semantic properties at all.*

A Baby Bear Approach

Some elements of the previous sections suggest a third approach which Goldilocks might consider "just right." Both the "Mama Bear" and "Papa Bear" sections ended with a suggestion that *part* of the impasse we face stems from the attempt to use *ordinary* terms to do *theoretical* work, and that we need to refine and even coin additional terms. Terms like 'concept' and 'meaning' are themselves semantically-rich.

Their paradigm instances are taken from our own case, in which concepts and meaning are bound up with things like inference, language, linguistic deference and conscious thought. The holistic theorist looks at this and infers that one cannot properly speak of "concepts" and "meaning" outside of this network. The reductionist posits that we can, and that these broader connections are not essential. Here we are in danger of arguing over an equivocation. The holist is pushing the pre-existing use of words like 'concept' and 'meaning' in one direction; the reductionist in another.

There is an obvious way out here. Indeed there are several alternative ways out, which are more or less notational variants of one another. I shall explore one of these, the first part of which is cognate with some suggestions made by Hilary Putnam []. Let us reserve the words in the intentional and semantic vocabulary—words like 'intentionality', 'thought', 'concept' and 'meaning' for cases like our own, where these are bound up with things like language, inference and consciousness. This will forestall any concern that attention to the continuities between human and animal cognition will commit any category errors or turn upon any equivocations in these words. It will also keep us grounded in the fact that we look at any *other* kind of cognition through the lens of our own case. For the cases of other animals, let us speak instead of "protoconcepts", where that term is initially used to pick out the continuities with human conceptuality that are to be found in other species. The term is thus initially characterized very loosely, with the intent that its meaning will be filled in more concretely in the course of further investigation. (Or, indeed, it might in turn need to be abandoned in favor of a yet more fine-grained set of distinctions.)

As these types and stages of protoconceptuality are filled out by the sciences of cognition, they would provide us of saying things that are literally true of the species to which they apply. But they would also give us a way of saying things that are literally true of ourselves as well, because our concepts *are*, among other things, protoconcepts. The human conceptual system *is* (for purposes of argument here) a distinction engine. This is not *all* it is; and something could be a similar sort of distinction engine without possessing by-golly concepts at all. This is not a reductive analysis. But the human conceptual system is, *among other things*, a distinction engine, and individual concepts are the products of a learning process employing this engine.

Mama and Papa's Ideal Little Bear

This Baby Bear approach is, I think, one of which both its fore-bears can approve. Like its reductionist mother, it seeks to understand human cognition by examining simpler elements of cognition that also appear in other species. Indeed, it asks the question of what we can learn about human cognition by seeing how it is "built" out of a succession of phylogenetic stages, and assumes that these answers can be illuminating in ways that nothing else can. (That is, they will provide answers to particular questions that cannot be answered in any other way.)

But, like its holist papa, this approach is not reductionist in spirit. It does not assume that our mental life is "nothing but" the sum of its animal parts, and follows Papa Bear's admonition not to "hang out on streetcorners with nothing-butheads". In our own case, protoconcepts are *concepts*, only seen through a lens or filter that idealizes away

some features that are essential to full-fledged conceptuality in order to see others more clearly. Our concepts "are" protoconcepts in the way that dollar bills "are" sheets of paper; and protoconcepts in us *are* concepts in the way that certain sheets of paper are dollar bills. The Baby Bear approach is thus anti-reductionist not only in the shallow sense of denying that conceptuality is reducible to something non-mental and non-intention, but in the deep sense of rejecting reductive *strategies* generally in favor of an alternative strategy of *idealization*.

Idealization is a common feature in the sciences, and is a special form of abstraction. In abstraction, one brackets features of the world that are then treated as irrelevant for the theoretical purpose at hand. Sometimes these features are truly irrelevant. But sometimes they are not really irrelevant—for example, a theory of electromagnetism abstracts away from gravitational influence, but gravitation plays a role in real-world dynamics, and so real objects never behave *exactly* as your electromagnetic theory says. [Cartwright 1983; Horst MWN] I use the word 'idealization' for abstractions of this sort: i.e., the sort where the abstraction matters *in vivo*. Some idealizations even *distort*, such as when gravitational models treat bodies as point-masses. (Consider the very different dynamics of a paper airplane and the same piece of paper crumpled into a ball.)

All scientific accounts screen out some features of the world in order to emphasize others, and the same is true of philosophical analyses. None of them tell the *whole* truth about things like dynamics, organic life, or cognition. Many of them simplify or distort their subject-matter in ways that are innocent in a core set of contexts but would be disastrous in others. (E.g., you cannot treat bodies as point-masses when

aerodynamics matters.) Knowing when a particular idealizing move is innocent and when it matters is part of the practical knowledge of the scientist, exercised in deploying the various models at her disposal. [] Scientific theories only tell "nothing but the truth" when you know when to keep silent. But they do tell truths. You just need to know the background assumptions of the theory, such as the idealizing moves and the pragmatic context, to understand *what* truths they are telling, or to assess whether what they are telling you is true.

The assumption of the Baby Bear approach is that, in the case of cognition, we can perform some idealizing moves that allow us to screen out things like consciousness, language and inference in order to see "older" features of cognition, but that what we find by doing this will also illuminate our understanding of full-fledged human cognition as well, and not merely animal cognition or the cognition of brain-damaged humans. In fact, what is revealed will be literal truths about human cognition; they just will not be the entire truth. The justification of such an assumption cannot be provided in advance. It is the working hypothesis of a research strategy in cognitive science, and whether it pans out in the end will depend upon an assessment of the ideas that arise in the course of pursuing it.

An Example

Let us consider a brief example of how the Baby Bear strategy might work. Let us take the explanation of the fuzziness of concepts mentioned earlier. Such an explanation might claim that concepts are fuzzy because (1) the conceptual system is a variation upon a *proto*conceptual system which is a "discrimination engine" optimized to

learn salient features of the environment, (2) this system is implemented through a particular sort of neural network architecture that has properties that make it good enough at performing this task to be adaptive, (3) this architecture was selected because of these adaptive features, (4) this particular network architecture, in learning, creates a partition of state space with fuzzy boundaries, and (5) individual concepts are realized through areas in the partition of state-space thus created. The theorist might go on to try to identify where in phylogenetic history such a discrimination engine appeared on the basis of evidence from comparative ethology: for example, if the ability to learn distinctions is found in all mammals and in no non-mammals, one might postulate that it appeared in a common ancestor of all present-day mammals after that lineage had separated from those of other phyla.

Like its reductionist Mama, the Baby Bear theory attempts to explain a feature of intentional states—in this case, the fuzziness of the concepts employed therein—by appeal to features of the brain and the selection history of those features. But, unlike its reductionist forbear, the explanations thus offered are not considered to be *reductions*. There is no claim that this, or any other account of concepts (especially ones cast in non-semantic, non-intentional, non-conscious, non-inferential terms) could provide sufficient conditions for being a *particular* concept (say, the concept of a cat) or indeed for being a concept at all. What it claims is merely that this story is one of many true stories about concepts, and that in saying something about (proto)concepts-as-implemented-in-neural-nets, we are saying something about concepts in us as well. When Baby Bear says, "the conceptual system is a distinction engine," he does not mean that that is its *essence*. The "is" here is *not* the "is" of analysis, but the "is" of predication. (And hence there is no

implication that a neural network we program on a computer will also have bona fide concepts, or for that matter that cats or prelinguistic children do so, just because they have neural networks very much like our own.) However, it is not just *any* sort of predication, but one that aims at pinning down some of the invariants found in the subject matter itself. There may be many such invariants, just as someone who wants to solve problems in dymamics must sometimes employ gravitational models, sometimes electromagnetic models, sometimes aerodynamic models, and sometimes must find a way to factor them together tolerably well.

Likewise, viewing such an explanation as an idealized explanation rather than as an analysis or a reduction leaves us free to allow that other things, things left out of the story entirely, may be essential to concepts. When we say that concepts "are" protoconcepts, and that protoconcepts are in turn determinate states of a discrimination engine, this is perfectly compatible with holding that, in order to count as full-fledged concepts, they must also be (potentially) conscious, linguistic or deployable in inferences and reason-giving. Indeed, this was the whole *point* of making a distinction between protoconceptuality and conceptuality-full-stop in the first place. However, it is important to bear in mind here that this was done in such a fashion that the notion of "protoconceptuality" is also applicable in the case of full-fledged concepts, in much the way that the notion "airplane" still applies to Air Force One, even though in some contexts it might be important to view it under other aspects, such as "Presidential symbol".

Conclusion

This paper has been an attempt to develop, in a light and breezy way, several important philosophical points. The broadest of these is that philosophers need to disabuse themselves of the view that the only kind of explanation that is any good is a reductive analysis in terms of necessary and sufficient conditions. That kind of explanation is in fact rather rare outside of mathematics. An important alternative can be found in the style of explanations employed in the sciences, which characteristically give idealized accounts of particular features of real-world objects. If we view cognitive-scientific explanations of features of the mind such as intentionality and semantics this way, the porridge is just right. It preserves the explanatory power of the explanations without turning them into attempts at reduction that are doomed to failure because of the situation of conceptuality and semantics within a broader web of human mental life.

Bibliography

Blackburn, Simon. 1984. *Spreading the Word.* Oxford/Clarendon Press.

Brandom, Robert. 1994. *Making It Explicit*. Cambridge, Mass.: Harvard University Press.

Fodor, Jerrold. 1987. *Psychosemantics*. Cambridge, Mass.: MIT Press/Bradford Books.

Horst, Steven. 1996. *Symbols, Computation and Intentionality: A Critique of the Computational Theory of Mind*. Berkeley and Los Angeles: University of California Press.

Horst, Steven. Forthcoming. *Mind and the World of Nature*.

Millikan, Ruth. 1984. *Language, Thought and Other Biological Categories*. Cambridge, Mass.: MIT Press.

Millikan, Ruth. 2001. *Clear and Confused Ideas*.

Putnam, Hilary. 1992. *Renewing Philosophy*. Cambridge, Mass.: Harvard University Press.

Searle, John. 1993. *The Rediscovery of the Mind*. Cambridge, Mass.: MIT Press.

Wittgenstein, Ludwig. 1958. *Philosophical Investigations*. Third Edition, translated by G.E.M. Anscombe. New York: Macmillan.