The origins of concepts

Daniel A. Weiskopf

Abstract: Certain of our concepts are innate, but many others are learned. Despite the plausibility of this claim, some have argued that the very idea of concept learning is incoherent. I present a conception of learning that sidesteps the arguments against the possibility of concept learning, and sketch several mechanisms that result in the generation of new primitive concepts. Given the rational considerations that motivate their deployment, I argue that these deserve to be called learning mechanisms. I conclude by replying to the objections that these mechanisms cannot produce genuinely new content and cannot be part of genuinely cognitive explanations.

1. Introduction

Our concepts enable us to entertain thoughts about an enormous variety of objects and states of affairs, both actual and possible. The richness of our conceptual repertoire determines the limits of these thoughts. The question I will take up here is whether this repertoire is fixed or open. While it is generally agreed that at least some of our concepts must be primitive, and indefinitely others may be formed from that primitive basis by combinatorial mechanisms, much debate has circled around the issue of whether we can learn to entertain new concepts that aren't simply part of the closure of this primitive basis under these combinatorial operations.

I argue that there are several mechanisms available for learning new primitive concepts—that is, new concepts that aren't simply combinations of previously given ones. Many have argued that learning new concepts must in *some* sense be possible, but few have either spelled out what conception of learning is involved when we talk about 'learning concepts' or laid out possible mechanisms that might implement such learning.[1] Here I aim to do both.

[1] The most prominent recent discussions of concept learning and the arguments against it are those of Cowie (1999), Margolis (Laurence & Margolis, 2002; Margolis, 1998), and Rupert (2001).

First, I lay out the argument against the possibility of learning new primitive concepts (Sec. 2). This argument rests on a rather narrow conception of what learning might be. I offer an alternative account of learning and distinguish it from this narrow conception (Sec. 3). On this account, learning a concept is understood as learning to token a new representation of a property. This new representation—a new vehicle paired with a new content—is precisely a new concept. I then show how the ability to token new representations can be acquired by psychological processes that are sensitive to a range of epistemic factors. It is because the acquisition of these new representations is under the control of such epistemically sensitive processes that this deserves to be called a learning story.

Given that the account offered here is a learning story, it is couched in terms of specifically psychological mechanisms. I first describe in relatively abstract terms the sort of psychological mechanism that is at work in concept learning (Sec. 4), then present a specific example of how this mechanism might work in practice, involving how concepts are learned from linguistic inputs (Sec. 5). Finally, I address two major objections to this proposal (Sec. 6). One is that it involves bootstrapping in a specially problematic way, and the other is that no cognitive mechanism can produce changes to the basic architecture of the cognitive system. While I conclude that neither objection is well-founded, addressing them will sharpen the content of the positive account itself.

2. Framing the Acquisition Problem

Debates about concept acquisition are bound up with debates about conceptual structure. The initial question all theorists of concepts have to answer is: do concepts have or lack internal structure?[2] *Compositionalists* adopt the former view. They hold that concepts are decomposable into features that contribute to their semantic content and psychological role. This has arguably been the default view in psychology. Prototype theory, on which concepts are composed of features that capture the central statistical tendency of a category, provides the most familiar example of a compositionalist view (Hampton, 1995). Many linguists, particularly those in the lexical semantic tradition, have also been compositionalists (Jackendoff, 1983; Katz, 1992; Pustejovsky, 1995).

On the other side are *atomistic* theories, on which most lexical concepts are unstructured or primitive representations (Fodor, 1994, 1998; Roelofs, 1997). According to atomism, concepts, like monomorphemic terms of natural language, lack semantically significant constituents. They are simple mental 'labels' for properties, kinds, events, and individuals. They may be associated with an arbitrarily large database of information about the categories that they represent, but none of these associative links are part of concepts' possession or individuation conditions.

Where, then, do new concepts come from? One way of getting a new concept is by recombining concepts already possessed. For phrasal concepts like FANCY DUCK, this means combining lexical concepts like FANCY and DUCK. For lexical concepts, answers differ depending on the view one takes of their structure. Compositionalists hold that lexical concepts are acquired by recombining the ultimate conceptual primitives in new ways. All possible lexical concepts must be constructible from these primitives, whatever they might be. Atomists, on the other

[2] By 'concepts' here I mean to restrict the focus only to *lexical* concepts (e.g., CAT, GLASS, TABLE, etc.). In accordance with most who adopt the Representational Theory of Mind, I treat concepts as a kind of mental representation (in opposition to Fregeans, who treat concepts as extramental objects such as senses).

hand, hold that the lexical concepts are themselves the primitives. So their acquisition story cannot be combinatorial, but must have some different form.

Both of these views, however, run aground on a common problem. For the compositionalist, the problem arises from the fact that it is unclear whether any *fixed* set of primitive concepts is, in fact, sufficient to generate all of the possible lexical concepts that humans can entertain. The search for such primitives has not been encouraging.[3] If no fixed set of primitives is sufficient, perhaps the notion that the set of primitives is fixed is the culprit. Whatever the initial primitive concepts that are given to us happen to be, it seems that it must be possible to extend this endowment somehow. But this is impossible if recombination is the only available method of arriving at new concepts.

Adopting atomism, of course, does not solve this problem, since for atomists most or all concepts are primitive. On this view, the initially given resources of the conceptual system are not a smallish set of primitives, but rather an extensive set of primitives corresponding to every possible lexical concept. This, too, has seemed implausible to many people (Levinson, 2003). For one thing, languages differ widely in their basic lexical repertoire; for another, within each language there are several open lexical classes that allow new words to be introduced at will. So the problem that compositionalists arguably face at a later stage is one that atomists face from the beginning. If this is correct, everybody needs a story about where new primitives come from.

[3] The most recent attempts at solving the problem of the primitives have been empiricist in character (Barsalou, 1999; Prinz, 2002). These views attempt to resuscitate the classical notion that all concepts are ultimately composed of copies of perceptual representations. While I won't take up the empiricist case in detail here, it is not at all clear that all concepts can be adequately reduced to percepts (Weiskopf, forthcoming-b). Alternative, non-empiricist views tend just to supplement the perceptual basis with a set of relatively abstract concepts such as CAUSE, EVENT, and AGENT (Jackendoff, 1992a; Miller & Johnson-Laird, 1976). Again, it is not at all clear that this simple form of supplementation will do the required trick of capturing the content of all possible lexical concepts.

The *Acquisition Problem*, as I will call it, is the problem of explaining how it is that new primitive concepts can be introduced into our repertoire. It is, intuitively, an extremely plausible idea that getting new primitive concepts must involve *learning* them. But both compositionalists and atomists have argued that the Problem cannot be solved by appeal to learning. Ray Jackendoff, for instance, says:

> ... one can't simultaneously claim both that something is a primitive and that it is learned, for learning entails construction from primitives. If I were to claim that some conceptual unit is a primitive, and it turned out to be learned, I would have no choice but to change my claim and to look for a deeper or more primitive set of basic units from which my unit could be constructed. This is a simple point of the logic of combinatorial systems. (Jackendoff, 1992b, p. 59)

For Jackendoff, then, it seems to be virtually analytic that primitive concepts cannot be learned.

The *locus classicus* of the argument against the learnability of primitives is found in the work of Fodor (1975; 1981). His argument can be reconstructed as follows:

1. 'Concept learning' necessarily involves a hypothesis-formation and confirmation (HFC) process.

2. At the stage of hypothesis-formation, an HFC process presupposes the possession of the to-be-learned concept itself.

∴ 3. 'Concept learning' necessarily presupposes the possession of the to-be-learned concept.

So, Fodor concludes, '[i]f the mechanism of concept learning is the projection and confirmation of hypotheses (and what else *could* it be), then there is a sense in which there is no such thing as learning a new concept' (1975, p. 95).

How does concept learning involve hypotheses? Consider one typical experimental setup: The subject sees, represents, and stores representations of instances I_1, I_2, ..., I_n. These instances might be pictures of animals, dot patterns, verbal feature lists, or whatever. The subject then makes judgments about whether the instances are members of the to-be-learned category, and (perhaps) receives feedback on her judgments. The stage of generating a response-guiding judgment requires representing a generalization about how the instances should be classified. An example hypothesis would be:

(H) Instances that are Φ should be categorized together.

Here, Φ is a representation of the conditions under which a positive categorization response should be made. The subject uses (H) in generating new behavioral responses, and continues until the experimenter says it's all over. But if Φ just is the concept that the experimenter intends the subject to learn, then learning presupposes the ability to entertain the concept already. So the concept can't have been learned at all, if learning is (minimally) a process in which the initial state is being *unable* to entertain the concept, and the final state is being *able* to entertain it.

Fundamentally, Fodor thinks, the mistake made by proponents of concept learning is to confuse concepts with hypotheses (Fodor, 2001, p. 133). Hypotheses are thoughts; that is, they are states with complete propositional contents. You can reason about them, wonder whether they are true or false, and decide whether the evidence, on balance, confirms or disconfirms them. Concepts, though, are constituents of thoughts. They correspond to sub-propositional contents. But learning—*real* learning—is only a matter of confirming or disconfirming full propositions. Hence concepts aren't even the right sorts of *things* to be confirmed or disconfirmed. You don't 'learn' concepts, you need concepts to learn!

This raises, rather pointedly, the question of how else concepts might be acquired if *not* by learning. The most commonly proposed alternative is that concepts are activated, or caused to become available to the conceptual system, by a process called 'triggering' (Fodor, 1981). Triggering, which is often described as a 'brute-causal' way of making a concept (or other representation) available to thought, has two central characteristics. The first is that the relationship between a trigger and what it releases 'may often be *extremely* arbitrary' (Fodor, 1981, p. 280); what concept is acquired may have nothing to do with what causes it to be acquired. This is a principal advantage of triggering mechanisms over learning. The second is that triggering is *epistemically insensitive*: it produces new psychologically contentful 'raw materials' (concepts and other representations) in a way that is indifferent to the kind of psychological process leading up to the triggering. All that matters for triggering is that the key stimulus occurs. It doesn't matter whether the process that leads to the triggering is, for example, sensitive to relevant sorts of evidence, or characterized by the appropriate sorts of inference.

While it may be plausible that some concepts are simply triggered by experience, it is difficult to believe that this is how all primitives are acquired (Sterelny, 1989). Many concepts are seemingly acquired only in the right sort of epistemic setting. The concepts X-RAY or CELL are ones whose attainment requires a significant amount of theoretical stage-setting and evidence-gathering. The potential arbitrariness of the triggering relation is even harder to swallow. It is difficult to believe that it is simply an accident that TIGER is acquired from experiences with tigers, pictures of tigers, and hearing tiger-related information from one's teachers. Given the dubiousness of extending the triggering model to cover all primitives, it seems incumbent on both compositionalists and atomists to seek an alternative solution to the Acquisition Problem.

3. Two conceptions of learning

As Fodor comments, '[w]e badly need—and have not got—an empirically defensible taxonomy of kinds of learning' (Fodor, 1975, pp. 34, fn. 34). While this remains true, I won't present such a taxonomy here. Instead, I will focus on just two conceptions of learning.

According to the Fodorian view, what happens in concept learning experiments is just inductive hypothesis confirmation. This requires the cognitive resources to entertain the proposition in question, to represent the evidence base, and to compute the degree of support that the proposition receives from the evidence. The model for this process is the logic of confirmation and the formalization of the norms of statistical reasoning. Call this an 'inductivist' conception of learning. Most of the work in the automated learning tradition also falls under this rubric. It is typical to provide programmed learning systems with a vocabulary of representations adequate to formulate all of the observations and hypotheses they will need in performing their tasks (Schank, Collins, & Hunter, 1986).

This restrictive notion of learning is close to the heart of Fodor's insistence that concept learning is incoherent. Unfortunately, though, it has an air of stipulation about it. The term 'learn' as it is used both by cognitive psychologists and in everyday discourse applies more widely than just to propositional thoughts. The restricted sense undeniably captures one thing that we might usefully mean by 'learning', but there is another sense in which concepts might be learned.

Among researchers on perception, it is common to talk about perceptual learning as a process in which a person (or animal) undergoes a change in her perceptual faculties as a result of an interaction with the environment that results in her being able to carry out some perceptual task better or perform some new perceptual function. As Goldstone (1998, p. 586) puts it, ‘Perceptual learning involves relatively long-lasting changes to an organism’s perceptual system that improve its ability to respond to its environment and are caused by this environment’. We should add to this list of properties the important proviso that these changes are mediated by psychological processes.

These conditions can be generalized to create a conception of learning applicable to all cognitive faculties and functions, not just perception. This conception has four components:

1. learning is a psychological process;
2. it is driven by and appropriately sensitive to environmental causes;[4]
3. it involves a relatively lasting change in the creature’s cognitive systems; and
4. it is adaptive, that is, it either (a) improves an aspect of cognitive functioning or task performance or (b) bestows the ability to carry out a new cognitive function or task.

Learning, in this more general sense, is *relatively long-lasting environmentally driven adaptive change in a creature’s cognitive systems brought about by a psychological process*. That is, it is a change in those systems that results in improvements to one’s ability to perform various cognitive tasks. Since these changes function to improve a creature’s cognitive functioning in a

[4] The ‘environment’ should be broadly interpreted to include other creatures—teachers—who guide the learner by providing specially chosen examples, delivering comprehensible error signals, and drawing attention to contrasts in classes that are not antecedently obvious to the individual.

way that reflects the structure of the task and the input, call this an ‘adaptationist’ conception of learning.[5]

Inductivism is sometimes presented as simply the only thing learning *could* be. This claim seems to be advanced by Louise Antony: ‘Just bear in mind that, to learn a word meaning, is *inter alia*, to learn a *proposition*. As, indeed, to learn *anything* is to learn a proposition’ (Antony, 2001, p. 205). Induction is one way to learn, but that doesn’t show that being an inductive process is necessary for learning. Here I am claiming only that satisfying (i)-(iv) is sufficient for something’s counting as learning.

The adaptationist conception of learning is generic by design. It is meant to capture what happens in many cases in which a creature learns something, independent of what it is that is learned. So, strictly speaking, adaptationist and inductivist learning are not exclusive. Rather, they define partially overlapping domains of processes. Some inductivist learning processes may be ways of *implementing* adaptationist learning processes, since one way to bring about a creature’s improvement in performance on a certain task might be precisely by a process of hypothesis formation and confirmation. Getting better at playing chess, for instance, might involve trying out hypotheses about what your opponent is trying to do (he’s setting up a knight fork), and improving at tennis might involve forming hypotheses about how to adjust one’s grip during a backhand shot.

However, the two sorts of learning do not overlap completely, since not every hypothesis confirmation process needs to be connected with improvement on some cognitive task, or

[5] It’s important to note that this sense of ‘adaptation’ has nothing to do with the term’s use in evolutionary biology. Not every thing learned in the adaptationist sense needs to improve a creature’s fitness or help it live longer or make it happier. It’s perfectly coherent that what is learned might have a negative effect on a creature’s fitness, lifespan, or happiness. You could, for instance, coherently hone your ability to prevent pregnancy (fitness-reducing), or your ability to commit suicide (lifespan-reducing).

acquisition of the ability to carry out a new cognitive function. One might acquire confirmation for a proposition that is perfectly useless, and that bestows no new advantage in carrying out cognitive tasks (e.g., that the cat has precisely *n* hairs at time *t*). So while some inductively learned propositions contribute to acts of adaptive learning, it may be that not all of them do.

Further, and more importantly for present purposes, not every adaptive learning process is an inductivist learning process. In the case of inductivist learning, what is learned is the proposition that has been confirmed (if the change in its confirmation value pushes it above some threshold). Given the generic character of adaptationist learning, what is learned need not be a proposition. Three brief examples should suffice to make this point.

First, consider learning associations. A learned association can take a number of forms. At a cellular level, learned associations among neurons exist when one neuron's activation preferentially causes another's activation. Hebbian mechanisms typically produce such links: cells that have fired together in the past tend to have their connections strengthened so that they fire together in the future. Psychologically, paired-associate learning involves learning to recall a particular arbitrary item when given a distinct cue, such as producing the three-letter string XOL when given the string WEV. At both the neurobiological and psychological levels laws of association exist that govern how these connections will be established. These associative connections are not propositions. They may be described propositionally (e.g., 'that XOL is associated with WEV'), but that does not show that what is learned is a proposition. What is learned is an association: a certain tendency for one representation to cause another to be tokened.

Second, consider learning skills or abilities. For instance, take a psychological ability such as the ability to recognize certain distinctive visual shapes. These shapes may be arbitrary, as with letters and experimental stimuli, or have some significance, such as the look of a key or a human face. Initially, one might not find a certain complex shape especially salient, and be unable to distinguish it from other, similar shapes. With repeated exposure, including perhaps supervised training, one can develop one's perceptual sensitivity to such complex shapes (Gibson & Gibson, 1955). The end result of this training process is a heightened perceptual sensitivity to them; they tend to 'pop out' from their surroundings (LaBerge, 1973), and are spontaneously seen as distinct perceptual units (rather than hard-to-distinguish parts of the perceptual background).

What is learned here is the ability to rapidly and spontaneously distinguish certain perceptual stimuli. This process may involve re-tuning some aspects of one's visual system, heightening attention to certain dimensions that were previously neglected, storing images of previously seen shapes (Nosofsky, 1986), and perhaps even producing new perceptual templates to aid in recognition (Schyns, Goldstone, & Thibaut, 1998). But neither the skill of identification itself, nor the representations that underlie it, need to be seen as learned propositions. Copying stimuli into memory and adjusting attentional weights are not processes that involve confirming any propositions about the input.

Third, consider concept learning itself. Learning a concept is learning a new kind of representation. This links up with learning abilities because possessing a representation seems to be a matter of having a certain ability. In particular, it minimally entails having the ability to token a contentful mental representation that plays the functional role distinctive of concepts. This isn't to say that concepts simply *are* abilities, only that having a concept necessarily involves the ability to token and employ a type of mental representation. The ability to token a concept entails indefinitely many other abilities, namely abilities to entertain thoughts having that concept as a constituent. When this ability comes about as the result of a psychological process that is sensitive to a range of epistemic factors, we can say that it is learned.

In all three of these cases we have epistemic sensitivity of a sort that is not present in triggering. Nature sets up triggering mechanisms in cases where what is to be acquired is important for the organism, too important to risk the chance that the organism will either not encounter the crucial evidence that would lead it to acquire the right end state, or that the organism is simply too dumb or too unlucky to hit on the right conclusion to draw from the evidence that it has. Learning involves the end state being dependent on the particular course of experience the creature has, and on how the creature reshapes its cognitive structures in response to that course of experience.

How this epistemic sensitivity is exhibited differs from case to case. Associations, for instance, are learned in virtue of the fact that they are established by experience of associated properties in the environment. Association strength is determined in part by frequency of association, which is evidence for the degree of correlation among properties in the environment. Negative evidence, meanwhile, can weaken associations. Creatures have associative learning mechanisms precisely because they allow them to organize their representations in ways that will be useful given the worldly distribution of properties. This means setting creatures up in such a way that experiencing A with B is treated as a piece of evidence that generally A's go with B's.

Abilities are learned in virtue of specific sorts of training (practice at discriminating objects or executing behaviors), and variations in the training regimen result in different end states. For example, punishment after a poor performance is evidence for the learner that this way of carrying out the task isn't to be repeated. This can induce self-generated variations in how one does it, so that the next performance constitutes an improvement. Successful performance, on the other hand, tends to stabilize the complex of representations that underlies skilled performance (although it doesn't preclude further improvements such as getting faster or more fluent in exercising one's abilities). The fact that skill learning involves treating successes or failures as evidence that one is performing properly or improperly indicates that it, too, is an epistemically sensitive process.

Finally, learning concepts, as we will see in Section 4, involves sensitivity to evidence about the structure of properties and kinds in the environment. In concept learning, one's evidence has to do with the existence of a category worth singling out in thought. Without evidence that a certain category of interest exists in the environment, one cannot learn a concept of that category. So in each case we have epistemic sensitivity of a sort not found in the classical cases of triggering.

One worry is that adaptationist learning entails that too many systems learn. Many systems alter their propensities under the influence of the environment, but not all of these changes count as learning. A system for maintaining homeostasis, such as the ones that regulate endocrine levels or core temperature, is not a learning system, since it is not part of the cognitive system proper; neither is the immune system. Self-regulating changes in natural and (non-cognitive) biological systems, then, fail to count as learning.

Fictional examples like the 'language pills' discussed in the literature are not ways of learning, since they do not involve psychological processes. Taking a Latin pill simply re-wires one's brain into the configuration had by the normal Latin speaker. This produces the end state without forcing the brain through the delicate process of environmentally guided adjustment that occurs over years in the normal case of second language learning. No psychological mechanisms of memory, attention, perception, emotion and motivation, and so on, are implicated. So while Latin pills produce knowledge of Latin, they aren't devices for *learning* Latin.[6]

Consider, finally, two canonical examples of 'unlearned' cognitive changes: filial imprinting and syntactic parameter-setting. Imprinting and parameter-setting are similar in that both involve a complex set of representations and behaviors becoming available as a result of a fairly specific kind of experience.[7] Imprinting involves a young precocial bird rapidly forming an attachment to an object that it thinks of as its mother and directing a stereotyped range of behaviors towards it. Parameter-setting involves a child arriving at a mentally represented grammar having certain determinate characteristics (e.g., placing heads initially rather than finally). In both cases, there are characteristics of this complex end-state that are plausibly *not* learned. Chicks appear to come equipped with a general template for what their mother ought to look like, as well as what sorts of behavior should be directed towards her. Similarly, human children appear to have information that, at a minimum, constrains the kind of grammars that they will settle on and enables them to arrive at these grammars rapidly and with little explicit instruction. In neither case does this information itself appear to be learned.

However, that doesn't mean that imprinting and parameter-setting don't involve learning. In both cases, the activation of some body of innately specified information is *preceded* by a

[6] A remarkable real-life example along these lines is song acquisition in the juvenile canaries studied by Gardner, Naef, & Nottebohm (2005). If these young birds are isolated from their conspecifics and played samples of artificially generated irregular songs, they can learn to imitate these songs reasonably well. But if they are given a testosterone injection to stimulate the onset of adulthood, their song shifts rapidly over to the normal adult song pattern, although they have not been exposed to this previously. The presence of the hormone, then, appears to induce rearrangement of the song patterns the birds can sing. But they don't thereby *learn* their song. (I heard of this study from Ariew (in press).)

[7] The specificity of the occasioning event, however, differs in each case. Imprinting requires a stimulus having certain perceivable characteristics, but may allow for significant variation within broad constraints. Parameter-setting, on the other hand, is highly constrained. It is only by hearing a Head-Initial language, for instance, that one comes to speak such a language. It is interesting, given the relatively constrained nature of the inputs in each case, that both have been held up as good examples of 'triggering' processes. Neither truly seems to possess the required degree of arbitrariness that is allowed in genuine cases of triggering. This has sometimes been interpreted as casting doubt on triggering explanations of these phenomena (Cowie, 1999; Fodor, 1998; Sterelny, 1989).

learning process. In the case of imprinting, chicks come with a perceptual template for what mother should look like. The chick needs to learn neither the mother template, nor how to behave towards mother. What is learned is precisely which object in the environment is mother. In the case of parameter-setting, children come with a Universal Grammar (UG) device with a small number of 'switches' on it.[8] In the classic principles and parameters approach, children do not need to learn the range of possible human grammars, nor any language-universal principles governing tree structure and movement. What is learned is what these linguistic switches should be set to, given the linguistic information in the environment. In both cases, then, what is learned is precisely how these distinctive bodies of innate information should be deployed given the circumstances the creature finds itself in.

Learning which object is mother requires a chick to search its environment for a best fit to the pre-existing template it possesses. Chicks do engage in this sort of active search. Moreover, imprinting is more successful with stimuli that more closely resemble the template. Some objects can be rejected as being poor fits: e.g., better fits have elements resembling a head and neck. Finally, imprinting can be reversed within a certain period of time, suggesting that chicks are willing to revise previous 'judgments' about what the best fitting object happens to be (for a review of these findings, see Bolhuis, 1991). These facts suggest that, while imprinting is rapid, it also involves actively collecting information, weighting it differentially, and entertaining rudimentary hypotheses, such as thinking '*this one* is mother'. While there is disagreement about whether this process is best seen as perceptual or associative learning, most contemporary

[8] While parameter-setting approaches are commonly adopted by syntacticians working in the Chomskyan tradition, linguists working in different theoretical frameworks may reject the assumption that there exists such a thing as Universal Grammar. I want to leave aside the question of whether this is in fact the right sort of approach to take to syntax acquisition. The point of evaluating the case is just to see whether parameter-setting is a form of learning on the present criterion, not to settle the issue of what linguistic information is innate, or what form that information takes.

ethologists characterize the process of settling on an object for imprinting as being a learning process (Bateson, 1990, 2000; Staddon, 2003, Ch. 14; ten Cate, 1994; van Kempen, 1996).

The case of language acquisition involves deciding how linguistic input should be analyzed, and detecting the units that correspond to the pre-existing parameters. What needs to be learned by the child is, for instance, whether *this* language is Head-Initial or Head-Final. Figuring this out requires information about what a head is, where they can be found in phrase structure trees, etc. But deciding how to set linguistic parameters involves using this information to reason about the input; e.g., locating phrase boundaries and determining what sorts of words fall on those boundaries (Juszyk, 1997, Ch. 6). Here the child can be seen as entertaining hypotheses concerning what sort of language she is learning, and aiming to achieve a certain confidence level in these judgments prior to turning her linguistic switches. Once these few parameters are set, however, the correct grammar simply 'pops out' without any process of theory construction or further inference.[9] In Piattelli-Palmarini's (1989) useful terms, setting parameters simply *selects* one of a relatively small set of pre-specified possible grammars that are encoded by the switch settings. In a similar way, once an object is chosen for imprinting, a set of attitudes and behaviors are, as Lorenz puts it, *released*.

What these cases suggest is that many acquisition processes are hybrids, involving a complex interplay between learning and innate information. With these preliminaries on learning out of the way, I turn now to the processes at work in the specific case of concept learning.

[9] This point is made by Jackendoff (1997), who notes that within the Universal Grammar itself there may be both (i) mechanisms for constraining the search space of possible grammars, and (ii) learning mechanisms that settle on a particular grammar given the linguistic evidence. However, he notes that most principles-and-parameters advocates have aimed to '[constrain] the search space of possible grammars, so that the complexity of the learning procedure can be minimized' (p. 6). In the limiting case, nothing resembling learning happens at all. The switches totally determine the grammar. But, as he goes on to note, this is not the only possibility. So there may be space for learning procedures even within UG itself.

4. The rationality of concept acquisition

Presented in the most general terms, the process of concept learning has two stages. In the first stage, the learner arrives at a judgment to the effect that a certain category (an individual, property, or kind) exists. This category is the one for which the subject is endeavoring to learn a concept. These judgments have the following form:

(E) There exists such-and-such a type of thing.

Existential, or *E-type*, judgments characterize the category to be learned by the properties that its instances have manifested in the learning situation. In the second stage, the learner produces a novel representation that has the function of picking out the category that is hypothesized to exist in the E-type judgment. I call the psychological operation of producing a new concept ‘coining’. These productions can be represented as having the form:

(C) Call the F (/that F) ‘G’.

Here F is a description of the type of thing hypothesized to exist in the prior E-type judgment, and G is a previously unused mental symbol of the appropriate syntactic type. The distinctive mental act that is expressed by instances of (C) is the act of coining a concept.

Coining can be described in terms of its psychological, semantic, and epistemic functions. Psychologically, coining is a basic cognitive operation that functions to produce new primitive concepts. This operation shouldn’t be confused with other basic operations, such as inference. When we coin a concept G, it isn’t *inferred* from anything—it can’t be, since G wasn’t part of one’s conceptual repertoire prior to being coined, and inferences only take you from concepts that you already possess to other concepts you already possess. It is for this reason that an inductivist conception of learning makes it *inevitable* that there is no such thing as learning a concept: inductivism only recognizes evidential relations among propositions such as entailment, inductive or abductive degree of support, and so on, and these are grounded in the notion of various sorts of inference.

Semantically, coining functions to introduce a new concept as a label for a category that is as yet only indirectly represented. An *indirect representation* of a category G is a descriptive or demonstrative representation that is satisfied by G, e.g., 'the property, whatever it is, that the speaker is referring to'; 'the thing, whatever it is, that causes those observations'; 'that kind of creature'. Indirect representations have the form 'the F' or 'that F'.[10] These representations provide an initial cognitive fix on the individual or property for which the thinker is learning a dedicated concept.

Coining produces a concept that *directly* represents whatever satisfies an indirect representation. By direct representation I mean that the new concept represents its content just as such, not under any particular descriptive or demonstrative mode of presentation.[11] The semantic function of coining, then, can be expressed as:

(S) $\mathit{Rep}_D(\text{G}) = \mathit{Rep}_I(\text{THE F})$ [or $\mathit{Rep}_I(\text{THAT F})$]

[10] While these expressions indirectly represent entities, they are at the same time direct representations of (respectively) the property of being the unique F, or the property of being the local or currently demonstrated F. Direct and indirect representation can come apart, since being G is typically not the same thing as being the F; see Section 6.1 for discussion.

[11] This isn't to imply that primitive concepts represent their content under *no* mode of presentation. Rather, it is to draw a contrast between complex representations—which *necessarily* represent their contents in a certain way, as fixed by their compositional structure—and primitives, which, insofar as they lack any such structure, don't. This distinction is largely independent of one's view about what actually determines a primitive concept's MOP. To see this point, consider a conceptual role-based theory, on which a primitive's MOP is determined by certain of its inferential liaisons. It might be that when a primitive is first introduced, the only such liaison that it possesses links it to its introducing descriptive information. However, we shouldn't assume that this link is inviolable. Later it might acquire more links to other bodies of information, and might even lose its link to the initial description. This is particularly likely if, as I have been emphasizing, the properties that occasion a concept's being coined are often fleeting ones. So the only view that is strictly inconsistent with the claim that primitive concepts can present properties in a way that differs from how they are presented by these descriptions is a view on which primitive concepts are *permanently* tied to the MOP given by their introducing description. But without further argument, I can see little reason to adopt this view. These descriptions may even be long forgotten for most of our concepts. Conceptual role theories of MOPs can, then, accept that primitives may later acquire different MOPs than they begin life with. Complex concepts, however, can't change their constituents—which are at least partial determinants of their MOP—without changing their identity. For more on content and modes of presentation, see Weiskopf (forthcoming-a).

where G is the newly coined concept, THE F/THAT F is its reference fixer, Rep$_{\mathrm{D}}$ is a function from concepts to what they directly represent, and Rep$_{\mathrm{I}}$ is a function to what they indirectly represent. Suppose also that we have the definition of indirect representation:

(Def. of Rep$_{\mathrm{I}}$) Rep_I(THE F) = α iff α satisfies Rep_D(THE F)

And suppose finally that we have both of the following:

(i) Rep_D(THE F) = <the F>

(ii) <G> satisfies <the F>

Then we can conclude that (iii) Rep_I(THE F) = <G>, by the definition of Rep$_{\mathrm{I}}$ plus (i) & (ii), and hence (iv) Rep_D(G) = <G>, by (S) and (iii).

There are three possible semantic relationships that might obtain between an indirect representation THE F (or THAT F) and a newly coined concept G. First, G has as its content the non-rigid description <the F>. Second, G has as its content the rigidified description <the actual F>. Third, G rigidly represents as its content the entity that satisfies <the F> at the time that G is coined. On the first and second options, coined concepts simply re-package descriptive content in a new simple symbol. I am advocating the third option: the semantic function expressed by (S) produces a rigid, non-descriptive concept from a non-rigid descriptive one.[12] A virtue of this approach is that it allows the acquisition of representations that *directly* represent new entities,

[12] A similar proposal is explored by Rey (1992), although he doesn't distinguish among these possibilities. He suggests that we imagine the human language of thought to contain an operator like Kaplan's 'dthat' for rigidifying definite descriptions, and that new concepts might be introduced by combining 'previously uninterpreted predicates' (p. 323) with such rigidified descriptions. Note, though, that on my account the input description need not contain any such symbol as DTHAT[THE F] or THE ACTUAL F. Concepts such as DTHAT and ACTUAL may be more sophisticated intellectual achievements. The inputs to coining, then, are descriptions and demonstratives that need not contain references to the actual world. The semantic function of coining captures the rigidifying effect of these concepts *without* requiring that the concept learner possess them.

and hence allows for acquiring new representational content. An effect of learning G is to make thoughts with the content <…G…> accessible where previously they were inaccessible.

Finally, what is the epistemic function of coining concepts? It can be helpful here to consider analogies with natural language and other public representational systems. Words can be regarded as pairings of sounds (or orthographic forms) with meanings. Similarly, concepts can be regarded as pairings of vehicles and contents. The vehicles of thought are symbolic structures realized in patterns of neural activity, rather than sounds or shapes on surfaces. So learning a new concept involves acquiring the ability to token a new symbol-content pairing.

Natural language provides many examples of coining new terms. We can introduce terms to talk about things whose precise identity isn't yet known. Scientists do so when they coin terms for the causes of phenomena, or for putative entities and mechanisms, early in their investigations. Detectives do so as well when they coin descriptive names for criminals ('Zodiac', 'Jack the Ripper').[13] In each of these cases, we have reason to think that something (an individual or property) exists, and we introduce a new mode of referring to that thing (a new sound-form assigned to a syntactic category). Just what is required for successful reference fixation in these cases is a matter of debate. At a minimum, it matters whether what is being named is the causal source of the naming and whether certain uniqueness conditions are satisfied (there should be only one underlying cause for the effects observed, for example). In the case of terms introduced by ostension it matters whether the subject is well-positioned to demonstrate the individual or property being named. Failing to meet these conditions may result in introducing a term with empty reference.

[13] See Jeshion (2004) and Reimer (2004) for recent discussion of the semantics of descriptively introduced names.

Coining a new concept, like coining a term, uses a description of the target object or category in order to fix reference initially. A variety of different sorts of descriptions might be used to introduce new concepts. These descriptions may use any sort of information available, including information about natural language (see Sec. 5). Teaching a child the concept FRAGILE, for instance, might involve calling a number of perceptually dissimilar objects 'fragile'. The child's FRAGILE concept might, then, have its content fixed initially by a description such as THE PROPERTY TEACHER IS CALLING 'FRAGILE' (where the quoted word is a word of English). Entertaining such a description requires the ability to think about teacher, her intentional verbal acts, and her noises (for evidence that young children have these abilities, see Bloom, 2000). The child's FRAGILE concept then represents the property that satisfies the description, namely the property of being fragile.

Other descriptive information might be causal, rather than linguistic. In the case of causal-theoretical concepts, such descriptions might have the form, e.g., THE KIND OF THING THAT CAUSES THESE EFFECTS. Here the new concept refers to the kind that description picks out (if any). Children may readily be able to think about how things causally interact with their one another; e.g., a child playing with a new toy might be able to think about THE KIND OF THING THAT MAKES THE MACHINE LIGHT UP. In the context of playing with the toy, a concept of that sort of thing (say a certain kind of game piece) might be important to acquire.

Finally, demonstratives can also fix the reference of newly coined concepts by tying them to particular samples, e.g., in the presence of Play-Doh THAT KIND OF SUBSTANCE might serve to fix the reference of a newly coined PLAY-DOH representation. The availability of such quantificational apparatus or mental demonstratives is enough to fix the reference of newly coined concepts. Presumably, given that we ordinarily lack detailed information about the things we are describing and demonstrating when we coin concepts, situational factors such as being causally connected to the target property in the right way play a significant role in reference fixation as well.[14]

What is the purpose of introducing new terms in public language? One goal is to facilitate more efficient communication. We could certainly refer to many things for which we have distinct words by using long and elaborate descriptions or (if samples are present) demonstratives. But doing so would be needlessly cumbersome. It would be preferable to introduce a symbol that is easily produced and remembered, particularly when we are dealing with things we believe to be stable and recurring parts of the environment.

A more important purpose turns on the fact that descriptions and complex demonstratives usually pick out categories by their *contingent* properties, e.g., ‘the sort of thing that causes such-and-such observations’, ‘the sort of flower on the table yesterday’, ‘that [pointing] kind of gadget’. These contingent properties are fleeting. They can be had by a category or individual at one time, and be absent at a later time. This dependence on situational factors puts them at a disadvantage as stable devices for representing categories. All things considered, it is better to have relatively *situation-independent* ways of representing categories of interest. New terms—proper names, nouns, verbs, and so on—can represent their referents in a way that abstracts from these potentially idiosyncratic features.

This leads to a parallel explanation for why we are led to create new concepts. Many things are likely to be encountered only once—particular cars, rocks, trees, clouds, and so on. Unless they are of some more lasting significance, we do not name these particulars. Their later

[14] As Devitt & Sterelny (1999) argue, both causal chains and descriptions may be required for introducing a term with a certain reference.

re-identification is of no importance, and it isn't necessary to gather further information about them under a single heading in memory. Many classes of things are similarly unimportant. Although they recur in our experience, there is nothing productive to discover about them: they have no hidden nature, they produce no significant or striking body of effects, they are not attended to by our conspecifics, and so on. We *can* pick them out, notice them, and describe them, if we choose. But we do not tend to form and store concepts dedicated to representing these sorts of things, even if we *could* do so.

On this view, we create new primitive concepts when we encounter some category that is of importance or interest to us, or which we have reason to think will be significant in the future. Once we create such concepts, we can use them in a variety of cognitive tasks, including the formation of new hypotheses about these categories. If you have an indication that nearby there is a category of interest, you have reason to begin collecting information about it, forming hypotheses about it, seeing if you can detect other instances of it, and so on.[15] Having a concept dedicated to representing a particular category provides a useful heading in memory under which incoming information about a category can be organized (a point emphasized in Millikan, 2000).

Coining new primitive concepts gives us a way of representing a category just *as such*, that is, without mentioning any properties that it might have. New primitives gain a significant degree of independence from the complex descriptions that are used to fix their reference initially, since newly formed concepts can enter into different causal and semantic relations than

[15] This account of our epistemic warrant for introducing new concepts is closely paralleled by Kroon's (1985) account of epistemic warrant for introducing new terms. Kroon emphasizes that term introduction is warranted when we believe that there is some entity that deserves being singled out by a term of its own, and when we take ourselves to be capable of determining further truths about the object to be named. This latter condition, which Kroon calls the Fact-finding condition, dovetails nicely, on the present account, with the emphasis on the role of newly coined concepts in generating hypotheses.

their reference-fixers could have. Given the contingency of many of these descriptions, this is a significant advantage.

So the primary rationale for creating new primitive concepts stems from the fact that they are devices for representing categories simply as such, as opposed to representing them under a particular complex mode of presentation. Complex symbols derive their causal and semantic properties from, *inter alia*, their constituents. Primitive representations represent their contents independently from these complex, contingent modes of presentation. A contingent representation of a category may not be useful if circumstances change: VODKA and MARY'S FAVORITE DRINK represent the same category now, but if Mary becomes a teetotaler (or is fickle about her drink preferences), they won't. If such a description is the only way you have of representing vodka, you run the risk of losing that representational ability as the world changes. Representing vodka as such blocks this possibility.

That new concepts are created as vehicles for thinking about categories that are in some sense or other *of interest* drives home the fact that concept learning mechanisms are responsive to rational considerations such as the probability that the category encountered is one that will be significant in future interactions, will provide a wealth of empirically discoverable effects, is likely to be of interest to your conspecifics, and so on. This further highlights the differences between these processes and triggering. The central image that triggering theorists favor is that of the organism patiently awaiting the occurrence of the right sort of stimulus to release a concept: as Fodor poetically puts it, in order to have new ideas 'you expose yourself to the play of experience, and you keep your fingers crossed' (Fodor, 1981, p. 314).

More than mere exposure is required, though. Just observing a set of correlated properties won't produce a new concept if a person doesn't think that those correlations have a common causal source. This requires making an E-type judgment, and making such judgments might not be trivial. It might not even be enough for that cluster of properties to be a *novel* one, since one and the same kind of thing can present itself in different ways on separate occasions, and you need to be able to discern whether this new cluster of properties indicates a new category, or just an old category presenting itself in a new way. E-type judgments require reasoning, not simply experience.

Finally, it should be clear that coining is a learning process, in the adaptationist sense. In coining a concept one gains a new ability to represent a category that was previously only indirectly representable, and this change in one's cognitive abilities comes about as the result of an epistemically sensitive psychological process with relatively persistent effects.

INSERT FIGURE 1 HERE

The possible mechanisms of concept acquisition, then, can be depicted as in Figure 1. I now turn to describing one particular application of this general picture of coining.

5. Learning concepts through language

It's a familiar, if much disputed, claim that one resource for extending our inner system of representations is language. Some have made the strong claim that natural language just *is* the vehicle of fully conceptual thought (Carruthers, 2002). We need not go this far, however. A weaker hypothesis advanced by the psychologist Sandra Waxman is that words function as 'invitations to form categories' (Waxman, 1999, p. 269). I would modify this slogan slightly:

words can serve as invitations to form *concepts*. Some recent developmental research suggests that this proposal might be on the right track.

Infants do not always spontaneously individuate objects by their kinds. If shown two objects belonging to different kinds (e.g., a toy duck and a ball) moving one at a time from behind an occluding barrier, infants may display no surprise if the occluder is dropped to reveal only a single object. Infants don't seem to conceive of the moving objects in terms of the kind to which they belong, hence they don't distinguish between them sufficiently to be surprised when only one object is revealed. This peculiar limitation can be overcome even in young infants, though. Xu (2002) showed that 9-month olds succeed in distinguishing a toy duck from a ball when each is given the appropriate, different verbal label, but don't do so when either no label is given or a single label (e.g., 'a toy') is given for each. Most interestingly, this effect disappears when two tones are used in place of words, as well as when humanlike noises of emotional expression are used. This suggests that language plays a *distinguished* role in learning to individuate objects by their kind.

So it appears to make a difference whether you give children a word when they are trying to learn certain categories. Superordinate concepts provide further evidence of this phenomenon. Superordinates are concepts such as ANIMAL, TOY, FURNITURE, VEHICLE, KITCHEN UTENSIL, etc. They usually represent groups of objects that have little in common perceptually. Kitchen utensils and animals tend to be perceptually heterogeneous. Seeing them as belonging to a common kind, or as falling under a common concept, requires some effort or insight beyond noticing surface similarities. Three-year old children (and even five-year olds in some tasks) evince great difficulty when asked to classify by superordinate membership (Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976).

However, learning superordinates is facilitated by giving children a novel noun, as Waxman and her colleagues show (Waxman, 1990; Waxman & Gelman, 1986). In these studies, preschool children were asked to sort pictures of various objects such as animals and food in either a No Word or a Novel Noun condition. Children who were not given a verbal label had difficulty sorting together pictures of objects that belong to a common superordinate category, even while they did well at sorting together 'basic' level objects (fish, dogs, etc.). By contrast, children given a novel noun were able to sort together perceptually dissimilar objects at the more abstract superordinate level. A similar result was shown by Nazzi & Gopnik (2001), who presented children with triads of perceptually distinct objects, giving two of them one name and the remaining object a different name. Twenty-month olds tended to pair objects with the same name together, while 16-month olds did not. Lexical items, then, can facilitate going beyond perceptual similarities and dissimilarities even in very young children.

What is the cognitive role of words in this learning situation? One possibility is that the child reasons as follows.[16] First, the child is aware that instances I_1, I_2, & I_3 are not very similar to each other. Nevertheless, the teacher called all of those instances 'blickets'. Suppose the child believes a general principle to the effect that when an adult calls a group of instances by a common word, then probably there is some important property they have in common. This entitles her to judge:

(E) There is some important property these instances have in common.

To believe (E) is to believe that a certain category of interest exists in the environment. Believing that such a category exists might lead the child to stipulate:

[16] I don't mean to imply that this reasoning is conscious or that the child could articulate it in precisely the way laid out here, only that it gives a more or less accurate characterization of some of the child's psychological state as she tries to figure out how to respond in the experimental situation.

(C) Call the property the teacher is talking about BLICKET.

In (C), the child coins a new label for the property the adult is indicating that the observed instances have in common, whatever that property might be. Presumably at this stage the child need have no detailed notion in mind of what the property *is*, beyond the fact that the trusted adult speaker has indicated to her that it is present. She only needs to recognize that her attention is being drawn to a commonality that she was not previously aware of.

Coining such a label might have two further psychological effects on the child. First, having a common feature increases the similarity of the observed instances, which in turn increases the likelihood that they will be categorized together. So if the child now represents these instances as being blickets, she represents them as being more similar to each other than they seemed previously. Second, it might encourage the child to search for ways to flesh out what she knows about the as-yet poorly understood feature. Being able to represent blickets as such enables the child to formulate hypotheses about precisely what sort of property blickethood is. These two effects could then account for the fact that children in the novel word condition perform better than those in the no-words condition. On this picture, words can facilitate categorization because they induce subjects to coin new concepts.

This process is anything but brute-causal. The child needs first to recognize that there is no basis that she has for classifying the instances together. The child's justification for introducing the novel concept comes from the principle connecting new words used by knowledgeable speakers with the existence of unobserved properties. If children are trusting of informants, then when their categorization judgments conflict with the use of a word by surrounding adults, they should hypothesize that there is something the informants know that they don't about the to-be-categorized objects. Given the judgment (E) that there exists some such property, (C) just involves introducing a new label for it. Having instantiated this label, the child can formulate hypotheses about the things that have the property in question. This in turn helps to explain how children might succeed at the sorting task when a word is present, but not without it.

6. Objections

6.1. Bootstrapping and conceptual content

This account of concept learning shows that the acquisition of indefinitely many concepts is possible if we suppose that some concepts are present initially. In defending a similar position, Landy & Goldstone (2005) say:

> Learning, we think, often involves the construction of psychological units that were unavailable before learning, and which transform the expressive capacities of the agent's psychological language. Cognitive processes like chunking alter what is tractably expressible, but we postulate processes that alter and extend what is expressible in principle in the cognitive language (i.e., using the basic terms and combinatorial rules of the agent). (p. 346)

In effect, if you have *some* vocabulary available in your language of thought, you can use it to learn a much wider vocabulary. It is, then, a kind of bootstrapping model. Given some initial concepts to work with, many more can be acquired by serving as the inputs to processes which expand the primitive lexicon of the conceptual system itself.

One might object to this story as follows. Coining mechanisms may show how new concepts can be introduced. But they don't constitute a way to genuinely extend the expressive power of the conceptual system. So, while they might provide a way to learn new concepts, no new representational *contents* are being acquired, only ways of repackaging old content in new *vehicles*.[17] This objection raises deep issues about how 'expressive power' is to be understood. I will sketch a reply that depends on one understanding of expressive power, and then offer some reasons for thinking that this is the appropriate understanding.

Suppose a child sees Daddy shaving with a razor each morning, and suppose that she can entertain thoughts involving the descriptive concept THE KIND OF THING DADDY USES ON HIS FACE EVERY MORNING. Having seen Daddy using that same sort of thing each day, she coins the symbol RAZOR to represent that sort of thing. This new symbol represents razors in the event that they are what falls in the extension of the reference-fixing descriptive content <the kind of thing Daddy uses on his face every morning>. The new symbol doesn't represent that descriptive content itself, only what satisfies it. This is just what is meant by saying that this description functions only to fix the reference of the new concept. So the child has what I called above an indirect way of representing razors. Nothing about the example is specific to descriptions; the child could also have represented razors by a demonstrative such as <that kind of thing>. Why, then, isn't this enough for the child to already have thoughts with the content <razor>?

Assume that general concepts have properties as their contents. So RAZOR represents the property <razor>, which has as its extension {x | x is a razor}. The child encounters a particular razor and describes it using a representation having the complex content <the kind of thing Daddy uses on his face every morning>. This complex descriptive content happens to have as its (present) extension the set of razors, since those are the kind of thing Daddy shaves with. But despite the fact that they coincide in extension, RAZOR and the description do not have the same content, since they represent *different* properties. The property of being a razor isn't the same as

[17] Viger (2005) notes that the notion of expressive power is central to the dialectic here. He argues that natural language can expand the expressive power of mentalese by introducing logical operators into it. He seems, however, to agree with Fodor that the expressive power of a creature's repertoire of predicate concepts cannot be increased. This latter point is the one that I take issue with here.

the property of being the thing Daddy uses on his face. The concepts are co-extensive but not content-identical. When the child acquires the ability to represent <razor>, then, she acquires the ability to directly represent a property that she couldn't before, even if she was able to represent that property's extension by using one of its contingent aspects to pick it out. This new ability constitutes an increase in the expressive power of the conceptual system, if the expressive power of a system is defined in terms of the properties that can be directly represented using the vocabulary of the system.

Putting the same point somewhat differently, there is a difference between being able to *refer to* a content-element (e.g., a property, kind, or individual) and being able to *entertain propositions containing it*. Indirect representations of properties deliver the former, while direct representations deliver the latter. So descriptive concepts such as THE F directly represent <the F>, while if <G> is what satisfies <the F>, the newly coined concept G would directly represent <G>, and would contribute that property (its content) to any thoughts containing it. The concept THE F, by contrast, contributes the property of being the F as its content. And as I've noted, <G> and <the F> are typically distinct properties. Given that coining operates to promote indirectly represented properties to being the (directly represented) semantic content of newly created mental representations, it counts as a way of expanding one's range of thinkable contents.

One reason to favor this notion of expressive power is the idea that what a system can express is as a whole a matter of the contents expressible using its vocabulary; for example, the propositions that it can represent. If we suppose propositions are Russellian, composed from properties and individuals, then the atomic propositions that can be entertained are determined by the properties and individuals that are expressed by the conceptual vocabulary. Two atomic Russellian propositions differ if they differ in either the individuals or properties that they contain, even when those properties are co-extensive. If we allow that new concepts express new properties that are contingently co-extensive with the indirect representations that fix their reference, we can still allow that these new concepts expand the expressive power of the conceptual system, since they expand the range of propositions that we can potentially entertain.

A second reason for construing expressive power in terms of the properties that the system directly represents comes from an example. Imagine that a person could demonstratively identify each instance of a category, for example the set of all cats. Such a person could then form the disjunction of all those demonstratives, e.g., THAT$_1$ OR THAT$_2$ OR THAT$_3$ OR … (where each THAT$_N$ is a separate demonstrative referring to a separate individual cat). This disjunction might pick out the same extension as the concept CAT, but being able to entertain that complex disjunction isn't sufficient for having that concept despite this coextensiveness. This example again suggests that expressive power needs to be more finely individuated than just referential abilities.

Representational systems containing quantifiers and demonstratives automatically have rather extensive referential abilities. Assuming that they also have the ability to represent kinds as such (i.e., they have the concept KIND) plus many contingent properties of such kinds is granting them a great deal. So to the extent that such systems can gain in expressive power, these gains must come in the properties that can be represented as such by the vocabulary of the system. These considerations from the theory of content suggest that this conception of expressive power is the appropriate one.

6.2. The scope of cognitive explanations

Jerry Samet notes that '[i]ntuitively, concepts are the building blocks of thought. But this means that they are prior to thoughts in just the way that bricks are prior to brick walls: you need the bricks before you start to build' (Samet, 1986, p. 583). In a similar vein, Antony says 'the fact that *cognitive* explanation has to stop short of the raw materials of cognition is one of the central lessons rationalists have always tried to get across' (2001, p. 213). On some readings, this is plausible. Perhaps no cognitive explanation will account for why there are any cognitive systems at all. We might need evolutionary biology or neuroscience for that. Perhaps no cognitive explanation could explain why certain particular cognitive states come about (e.g., intrusive thoughts). We might again need to descend to lower levels to explain these events. And certainly no cognitive explanation of the presence of a unit of 'raw material' can employ that very unit. But it doesn't follow that an explanation for the presence of some piece of raw material can't be couched in terms of the employment of *other* cognitive materials.

However, perhaps we can construct an argument for the claim that no cognitive processes can produce new cognitive materials. Pylyshyn (1984) distinguishes between states and processes that operate at the *cognitive* level, and those that are part of the *functional architecture* of a system. Functional architecture determines 'what operations are primitive, how memory is organized and accessed, what sequences [of representations] are allowed, what limitations exist on the passing of arguments and on the capacities of various buffers, and so on' (p. 92). A cognitive process is a set of representational states (and transitions among them) that takes place within the resources and constraints provided by the functional architecture. The central characteristic of functional architecture is its cognitive impenetrability: architectural states and processes cannot be influenced by an organism's beliefs, desires, intentions, and other intentional states. 'The behavior of the putative, primitive operation must itself not require a semantic level explanation. In other words, in explaining the behavior of the hypothesized primitive, there must be no need to appeal to goals, beliefs, inferences, and other rational principles […]' (p. 133).

Functional architecture can change as a result of neurobiological factors such as environmentally driven neuronal plasticity, trauma, and rewiring due to normal maturation. But it cannot change as a result of any psychological process. This is implied by the fact that it doesn't change as a result of 'rational principles'. Further, the functional architecture fixes the set of primitive operations and representations available to the system. So any changes in the representational resources of a creature cannot be the product of psychological states, or rational activity more generally.[18] The dilemma is: either (1) new concepts arise only from the non-rational operation of the functional architecture; or (2) some architectural changes can be induced directly by psychological states.

Rupert (2001) proposes that the mechanisms that produce new concepts are themselves part of functional architecture. This is, in effect, to adopt option (1) and deny that concept learning is a psychological process in the genuine sense. I have given some reasons in Section 4 to suppose that, in its sensitivity to broad epistemic constraints, concept learning is more plausibly thought of as a rational psychological process. It seems to violate Pylyshyn's neat separation of levels to say that there are neurobiological (non-semantic) processes that are sensitive to a range of evidential factors as such.

An answer in the spirit of option (2) might run as follows. Some facts about the architecture are immutable, but perhaps certain of its aspects allow for relatively constrained sorts of change. Suppose that the conceptual system is a language-like system of representation in at least this respect: there are various classes of expressions corresponding roughly to the syntactic classes available in natural languages. So there might be mental analogues of nouns,

[18] At least, not the *direct* product. Intentional acts like getting drunk or hitting myself in the head with a hammer can certainly change the structure of my functional architecture—not always for the better.

verbs, adjectives, and so on. The mental lexicon plus the syntactic and semantic rules together determine the range of possible propositional thoughts.

Now we can define two possible senses in which this architecture might be ‘fixed’:

Strong immutability: No cognitive process can alter either the mental lexicon or the syntactic/semantic rules;

Weak immutability: No cognitive process can alter the syntactic/semantic rules.

On strong immutability, the mental lexicon itself is completely outside the influence of cognitive processes. On weak immutability, however, changes in the mental lexicon itself are permitted, so long as they do not also involve changes to the syntactic and semantic rules of the conceptual system. It might be permissible to add a new ‘noun’ to the body of available nominal concepts; but if the system lacked, say, any syntactic rules for inserting adverbial concepts into thoughts, no adverbial concepts could be added to the system, since their deployment would require altering the syntactic/semantic rules themselves. Weak immutability entails that no changes to the possible logical form of thoughts can be introduced, but this leaves open the possibility of changing the range of particular concepts that are inserted into those logical forms.

The mechanisms I have been describing are consistent with weak immutability. They produce new symbols belonging to previously existing syntactic types. At the heart of these mechanisms is a basic mental operation that produces a new primitive symbol and assigns as its semantic content whatever satisfies a description or demonstrative. If this operation itself is a mental primitive, its functioning cannot receive further psychological explanation. But that doesn’t entail that creating new concepts isn’t a psychological process, since the deployment of this primitive operation is under the control of a range of broadly epistemic considerations.

If the inferential processes in a system are only sensitive to the syntactic types of symbols (e.g., they have something like the form: [Adj N] → N), then adding new members to each syntactic type won't require adding new inferential rules. The same applies to the case of the combinatorial rules that form complex expressions from primitives. Generally, as long as no representational primitives require *item-specific* inferential or syntactic rules, adding new primitive symbols won't necessitate changes to the rules of the cognitive system. (Or, alternatively, if there are primitive item-specific inferential rules, the system can't learn any new ones.) The sorts of changes produced by adding new symbols, on this account, don't require wide-ranging modifications to the system's resources or inferential processes. So allowing for weak immutability seems to be a comparatively minor concession.

7. Conclusions

To return to the opposing views sketched at the outset of this discussion, it is clear that compositionalists and atomists present us with an unappetizing dilemma. Compositionalists offer too impoverished a conceptual basis, while atomists enrich the set of conceptual primitives but offer an implausible view of their acquisition. These views agree in offering an essentially fixed vocabulary, disagreeing only on what sorts of concepts are part of that fixed vocabulary.

I don't take a stand here on how rich the initially specified set of triggered concepts is, or even on what sorts of concepts make it up. Rather, I've argued that the common notion of a fixed conceptual lexicon is a mistaken one. Of course, jettisoning the fixed-vocabulary assumption does not settle the debate between compositionalists and atomists. Rather, it offers a solution to a problem they have in common. However large the triggered basis is, and however extensive the available combinatorial operations are, human minds come equipped with several mechanisms for producing new representations that allow them to tune their cognition and behavior more finely to the environment. This form of tuning should be seen as learning in a perfectly straightforward sense: the adaptation of a creature's cognitive resources to the environment and the tasks it has to carry out. And this broadened conception of learning, plus the existence of a variety of concept-learning mechanisms that satisfy it, provide us with what we desired from the outset, namely a solution to the Acquisition Problem.[19]

References

Antony, L. M. (2001). Empty heads. *Mind and Language, 16*, 193-214.

Ariew, A. (in press). Innateness. In M. Matthen & C. Stephens (Eds.), *Handbook of the Philosophy of Biology*.

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences, 22*, 577-609.

Bateson, P. (1990). Is imprinting such a special case? *Philosophical Transactions of the Royal Society of London B, 329*, 125-131.

Bateson, P. (2000). What must be known in order to understand imprinting? In C. Heyes & L. Huber (Eds.), *The Evolution of Cognition* (pp. 85-102). Cambridge: MIT Press.

Bloom, P. (2000). *How Children Learn the Meanings of Words*. Cambridge: MIT Press.

Bolhuis, J. J. (1991). Mechanisms of avian imprinting: A review. *Biological Reviews, 66*, 303-345.

Carruthers, P. (2002). The cognitive functions of language. *Behavioral and Brain Sciences, 25*, 657-674.

Cowie, F. (1999). *What's Within? Nativism Reconsidered*. Oxford, UK: Oxford University Press.

Devitt, M., & Sterelny, K. (1999). *Language and Reality* (2nd ed.). Cambridge, MA: MIT Press.

Fodor, J. (1975). *The Language of Thought*. Cambridge, MA: Harvard University Press.

Fodor, J. (1981). The present status of the innateness controversy. In *Representations* (pp. 257-316). Cambridge, MA: MIT Press.

Fodor, J. (1994). Concepts: A potboiler. *Cognition, 50*, 95-113.

Fodor, J. (1998). *Concepts*. Oxford, UK: Oxford University Press.

Fodor, J. (2001). Doing without *What's Within*: Fiona Cowie's critique of nativism. *Mind, 110*, 99-148.

Gardner, T. J., Naef, F., & Nottebohm, F. (2005). Freedom and rules: The acquisition and reprogramming of a bird's learned song. *Science, 308*, 1046-1049.

Gibson, J. J., & Gibson, E. J. (1955). Perceptual learning: Differentiation or enrichment? *Psychological Review, 62*, 32-41.

Goldstone, R. L. (1998). Perceptual learning. *Annual Review of Psychology, 49*, 585-612.

[19] I would like to thank Edouard Machery, Rob Rupert, Robert Thompson, Chase Wrenn, and Tad Zawidzki for comments on earlier versions of this material. Thanks also to audience members at George Washington University, Rice University, UC Davis, University of South Florida, and York University, where versions of this paper were also read. Finally, an anonymous referee for this journal provided extremely thoughtful and detailed comments that spurred many improvements to the paper, for which I'm grateful.

Hampton, J. A. (1995). Similarity-based categorization: The development of prototype theory. *Psychologica Belgica, 35*, 103-125.

Jackendoff, R. (1983). *Semantics and Cognition*. Cambridge: MIT Press.

Jackendoff, R. (1992a). What is a concept, that a person may grasp it? In *Languages of the Mind* (pp. 21-52). Cambridge: MIT Press.

Jackendoff, R. (1992b). Word meanings and what it takes to learn them. In *Languages of the Mind* (pp. 53-67). Cambridge, MA: MIT Press.

Jackendoff, R. (1997). *The Architecture of the Language Faculty*. Cambridge: MIT Press.

Jeshion, R. (2004). Descriptive descriptive names. In M. Reimer & A. Bezuidenhout (Eds.), *Descriptions and Beyond* (pp. 591-612). Oxford, UK: Oxford University Press.

Juszyk, P. (1997). *The Discovery of Spoken Language*. Cambridge: MIT Press.

Katz, J. J. (1992). The new intensionalism. *Mind, 101*, 689-719.

Kroon, F. W. (1985). Theoretical terms and the causal view of reference. *Australasian Journal of Philosophy, 63*, 143-166.

LaBerge, D. (1973). Attention and the measurement of perceptual learning. *Memory and Cognition, 1*, 268-276.

Landy, D., & Goldstone, R. L. (2005). How we learn about things we don't already understand. *Journal of Experimental and Theoretical Artificial Intelligence, 17*, 343-369.

Laurence, S., & Margolis, E. (2002). Radical concept nativism. *Cognition, 86*, 25-55.

Levinson, S. C. (2003). Language and mind: Let's get the issues straight! In D. Gentner & S. Goldin-Meadow (Eds.), *Language in Mind: Advances in the Study of Language and Thought* (pp. 25-46). Cambridge: MIT Press.

Margolis, E. (1998). How to acquire a concept. *Mind and Language, 13*, 347-369.

Miller, G. A., & Johnson-Laird, P. N. (1976). *Language and Perception*. Cambridge: Harvard University Press.

Millikan, R. G. (2000). *On Clear and Confused Ideas*. Cambridge, MA: Cambridge University Press.

Nazzi, T., & Gopnik, A. (2001). Linguistic and cognitive abilities in infancy: When does language become a tool for categorization? *Cognition, 80*, B11-B20.

Nosofsky, R. M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General, 115*, 39-57.

Piattelli-Palmarini, M. (1989). Evolution, selection, and cognition: From "learning" to parameter setting in biology and the study of language. *Cognition, 31*, 1-44.

Prinz, J. (2002). *Furnishing the Mind*. Cambridge, MA: MIT Press.

Pustejovsky, J. (1995). *The Generative Lexicon*. Cambridge: MIT Press.

Pylyshyn, Z. W. (1984). *Computation and Cognition*. Cambridge, MA: MIT Press.

Reimer, M. (2004). Descriptively introduced names. In M. Reimer & A. Bezuidenhout (Eds.), *Descriptions and Beyond* (pp. 613-629). Oxford, UK: Oxford University Press.

Rey, G. (1992). Semantic externalism and conceptual competence. *Proceedings of the Aristotelian Society, 82*, 315-334.

Roelofs, A. (1997). A case for nondecomposition in conceptually driven word retrieval. *Journal of Psycholinguistic Research, 26*, 33-67.

Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology, 8*, 382-439.

Rupert, R. (2001). Coining terms in the language of thought: Innateness, emergence, and the lot of Cummins' argument against the causal theory of mental content. *Journal of Philosophy, 98*, 499-530.

Samet, J. (1986). Troubles with Fodor's nativism. In P. French, T. E. Uehling, Jr., & H. Wettstein (Eds.), *Midwest Studies in Philosophy: Studies in the Philosophy of Mind* (Vol. 10, pp. 575-594). Minneapolis, MN: University of Minnesota Press.

Schank, R. C., Collins, G. C., & Hunter, L. E. (1986). Transcending inductive category formation in learning. *Behavioral and Brain Sciences, 9*, 639-686.

Schyns, P. G., Goldstone, R. L., & Thibaut, J.-P. (1998). The development of features in object concepts. *Behavioral and Brain Sciences, 21*, 1-54.

Staddon, J. E. R. (2003). *Adaptive Behavior and Learning*. Retrieved 5/1/2007, from http://psychweb.psych.duke.edu/department/jers/abl/TableC.htm.

Sterelny, K. (1989). Fodor's nativism. *Philosophical Studies, 55*, 119-141.

ten Cate, C. (1994). Perceptual mechanisms in imprinting and song learning. In J. A. Hogan & J. J. Bolhuis (Eds.), *Causal Mechanisms of Behavioural Development* (pp. 116-146). Cambridge: Cambridge University Press.

van Kempen, H. S. (1996). A framework for the study of filial imprinting and the development of attachment. *Psychonomic Bulletin & Review, 3*, 3-20.

Viger, C. (2005). Learning to think: A response to the *Language of Thought* argument for innateness. *Mind and Language, 20*, 313-325.

Waxman, S. R. (1990). Linguistic biases and the establishment of conceptual hierarchies. *Cognitive Development, 5*, 123-150.

Waxman, S. R. (1999). The dubbing ceremony revisited: Object naming and categorization in infancy and early childhood. In D. L. Medin & S. Atran (Eds.), *Folkbiology* (pp. 233-284). Cambridge, MA: MIT Press.

Waxman, S. R., & Gelman, R. (1986). Preschoolers' use of superordinate relations in classification and language. *Cognitive Development, 1*, 139-156.

Weiskopf, D. A. (forthcoming-a). Atomism, pluralism, and conceptual content. *Philosophy and Phenomenological Research*.

Weiskopf, D. A. (forthcoming-b). Concept empiricism and the vehicles of thought. *Journal of Consciousness Studies*.

Xu, F. (2002). The role of language in acquiring object kind concepts in infancy. *Cognition, 85*, 223-250.

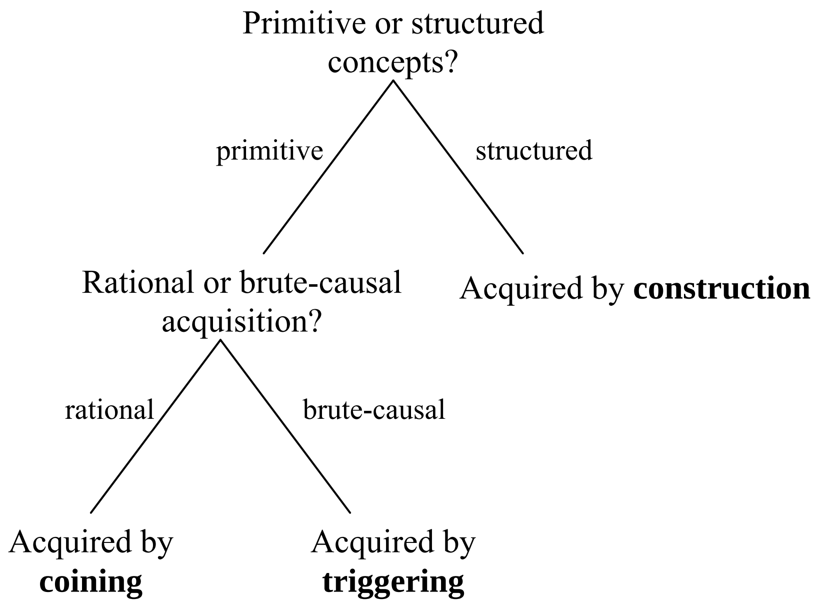
Primitive or structured concepts?
primitive
structured
Rational or brute-causal acquisition?
Acquired by **construction**
rational
brute-causal
Acquired by **coining**
Acquired by **triggering**

Figure 1: Map of possible ways to acquire concepts