

In Defense of the neo-Piagetian Approach to Modeling and Engineering Human-Level Cognitive Systems

John Licato and Selmer Bringsjord

licatj, selmer@rpi.edu
Department of Computer Science
Rensselaer Polytechnic Institute
Troy NY 12180 USA

Abstract

It seems that any robust cognitive system must have the capacity to: maintain and learn new concepts; believe propositions about its environment built out of these concepts, and out of what it perceives; reason about the propositions it believes, in order to among other things manipulate its environment, and learn new concepts. Given this list (**L**), it is hard to see how any attempt to build or simulate a cognitive system can avoid falling under what can be referred to as a *neo-Piagetian* approach. Unfortunately, the neo-Piagetian view of cognition relies on assumptions that have been challenged for years by Jerry Fodor, who purports to show through straightforward deduction that **L** requires all concepts to be innate. In this paper, after setting out Fodor's challenges, we refute them and, inspired by those refutations, sketch our solutions on behalf of those wanting to computationally model and construct cognitive systems, particularly with neo-Piagetian assumptions. The overall and guiding purpose of this paper is to encourage a larger and more explicit role for neo-Piagetianism in the engineering of cognitive systems, which as we will show, is necessary for both fields to respond to Fodor's challenges.

The fields of psychology, cognitive systems, and philosophy of mind overlap quite frequently; experts from any of these fields can usually speak intelligently on topics interesting to those in the other fields. It becomes easy, however, to forget the benefits of occasionally re-examining the alternate perspectives that seemingly tangential fields of study sometimes offer. When done right, mining these fields can bring revolutionary insights, but the mining must be done in a way that ensures the proper mixing of assumptions, especially when those assumptions contradict each other.

We will present some of the assumptions underlying one of the most important classes of psychological theories: neo-Piagetianism. After looking at some intimate connections between neo-Piagetianism and cognitive systems, it becomes clear that some of these shared assumptions face direct challenges from Jerry Fodor; challenges of sufficient power as to require that they are made explicit and addressed. We then show that Fodor's objections can be suc-

cessfully rebutted, and suggest that to do so, cognitive systems must answer Fodor's objections in a neo-Piagetian way.

Neo-Piagetianism and Cognitive Systems

It is a testament to the success of a theory that it becomes almost implicitly accepted by most practitioners in the relevant field. The core claims of Jean Piaget, the collection of which we will refer to in this paper as *Piagetianism*, are examples of this. Piaget faced criticism of some of his core claims, and largely in response to these a class of theories we now refer to as *neo-Piagetianism* arose to preserve Piaget's original conceptions (Meadows 2006; Case 1992). Though the theories falling under the umbrella of neo-Piagetianism may look quite different at first glance, it is generally agreed that they share a set of common postulates, among them:

- **Knowledge is structured, and that structure is important in determining the thoughts agents are capable of having.** We use concepts to perceive and understand the world, and these concepts behave as if they have a structure, a postulate the neo-Piagetians have inherited from Piagetianism (Case 1992). Furthermore, this structure must have a certain productivity and systematicity in order to express and generate combinations of concepts the agent may not have encountered before (we elaborate below, when we explaining compositionality).
- **Conceptual development involves a qualitative change in the underlying structure of concepts** (Case 1992). Development, then, is more than mere acquisition of data; it is a qualitative change in the structuring of that data. This is a difficult feature to model and understand, as computations are often defined over symbols with a relatively stable structure. Shultz & Sirois (2008), for example, draw distinctions between artificial neural networks that are static (using back-propagation and quantitative changes) and constructive (using cascade correlation and qualitative changes).
- **Conceptual development fundamentally involves the agent.** Learning is not a result of a passive agent being exposed to events. Rather, learning is an interactive process in which the agent is actively involved. This process involves responding to observation both by fitting the observed facts into existing structures (*assimilation*) and

modifying existing structures to fit the facts better (*accommodation*). This interaction between agent and environment as a part of the learning process has been described in CLARION as comportment (Sun 2002) and as an emerging approach among neo-Piagetians (Demetriou and Raftopoulos 2004).

- **There are limits to the expressivity of thought, which increases with age, experience, and maturation; these limits may be uneven across different domains of thought, and between individuals.** Whereas the previous three postulates are mostly restatements of Piaget's original views, this claim is the most characteristic of *neo-Piagetian* thought. Note that this is not simply stating that children can communicate better or think more efficiently as they age, which is obvious; instead, the idea is that there are upper limits to the *types* of thoughts children can think. Exactly how to define this upper bound is an active area of research; many believe that it pertains to working memory (Case 1992; Barrouillet and Gaillard 2011; Meadows 2006).
- **F: Humans, if neurobiologically normal, nurtured, and sufficiently educated, naturally develop a context-free deductive reasoning scheme at the level of elementary first-order logic** (Bringsjord, Noel, and Bringsjord 1998; Bringsjord et al. 2006). Though **F** is directly reflective of Piaget's position (e.g., for confirmation see (Inhelder and Piaget 1958)), **F** is not as uncontroversial as the preceding postulates, but it is important to include, for reasons we discuss shortly.

There are many more, but they are not relevant to the present discussion; see Case (1992) for a more thorough list.

Any cognitive architecture comes with assumptions about how human thought works (Sun 2004; Ehman, Laird, and Rosenbloom 2006), and the preceding paragraphs provide examples of how many of those line up with neo-Piagetianism. The preceding list should sound very familiar to designers of cognitive systems; indeed, whether done explicitly or not, many cognitive systems work on some subset of neo-Piagetian assumptions¹. To make this connection even more explicit, let us briefly analyze the current state of cognitive systems. It should be relatively uncontroversial that:

At the very least, a sufficiently robust cognitive system can (this we dub list **L**):

- maintain and learn new concepts;
- believe propositions about its environment built out of these concepts, and out of what it perceives; and
- reason about the propositions it believes, in order to, among other things, manipulate its environment and learn new concepts.

Since in our view the acceptance of **L** is a bare minimum for those aiming to build artificial cognitive systems, we

¹Drescher (1991) and Shultz (2004) each present architectures which are explicitly inspired by Piagetianism, the former in the symbolic approach and the latter in the connectionist (Drescher 1991; Shultz 2004)

will argue that it would be beneficial to encourage closer coöperation between the engineering of such systems and neo-Piagetianism; this of course entails the need to test the plausibility of the neo-Piagetian assumptions, and to refine working theories about how to implement some of the items in **L**. In short, it seems that understanding the assumptions of neo-Piagetians is important to any long-term research program in cognitive systems. Given this, it is crucial to acknowledge that these assumptions have been under attack for years, most prominently by Wason and Johnson-Laird on the one hand, and Fodor on the other.² In fact, according to Fodor, many of the neo-Piagetian assumptions lead to conclusions that are inherently *incompatible* with **L**. It thus stands to reason that Fodor's arguments must be rebutted before any neo-Piagetian/cognitive systems crossover can be achieved. We turn now to Fodor's objections.

Fodor's Objections

It is outside the scope of the present paper to provide a full discussion of Fodor's theory of concepts; for a more detailed understanding, we recommend (Fodor 1998; 1980; 2008). We will try to limit the discussion of Fodor to those parts directly relevant to neo-Piagetianism and cognitive systems. To understand this discussion, we first note that Fodor's program attempts to answer the deceptively simple question: *What exactly is a concept?* Fodor primarily employs three theses to answer this question:

Representational Theory of Mind The intentionality of natural languages derives from the intentionality of propositional attitudes, which derive their intentionality from mental representations, which are themselves composed of concepts.

Computational Theory of Mind (At least some) thoughts are computations defined over the syntax of mental representations.

Concept Atomism All lexical concepts (roughly, simple concepts referring to words like DOG or CAT) are atomic; they are not constructs of epistemic (experience-related) capacities, sensorimotor concepts, canonical definitions, logical constructs, inferential roles, or any other feature of concepts that violates compositionality.

We will discuss compositionality shortly. Note that throughout this paper we will be using notation similar to Fodor (2008), where entirely capitalized words refer to concepts, and italicized words are those things in the world to which concepts refer. To think the concept CAT is to think about the class of objects in the world we call *cats*. Single quotes refer to words. To make a statement including 'cats' is to refer to *cats* (Fodor 2008). We put complex concepts in parentheses, e.g. (BLACK CAT) is a complex concept built out of the lexical concepts BLACK and CAT.

²In the case of the former, the original attack on Piaget's claim that neurobiologically normal human beings develop into reasoners competent at the level of first-order logic comes via what is now known in countless books and articles at the 'Wason selection task;' see (Wason 1966). As to the latter, see (Fodor 1980), and most recently (Fodor 2008).

Fodor would classify the theories we are herein calling “neo-Piagetian” as “Concept Pragmatism” (Fodor 2008), as opposed to what he identifies as the “Cartesian” view of concepts. Essentially, the idea is that “thought is the internalization of action” is pragmatic, whereas “action is the externalization of thought” is Cartesian (Fodor 2008). Cognitive systems that acquire concepts through “bottom-up” processes are examples of Concept Pragmatism.

Fodor, however, claims it’s all happening in the wrong order. Imagine an agent *A* who, through some automatic low-level process, learns how to recognize normal instances of the concept DOG. *A* has thus acquired the “know-how” component of the concept DOG, which means something like ‘knowing how to behave towards’, or ‘knowing how to recognize’. Eventually, *A* associates whatever epistemic capacities he uses to recognize *dogs* with an internal symbol corresponding to DOG. Now, because *A* can form mental representations consisting of the concept DOG which refer to whatever epistemic capacities he has associated with that symbol, he has acquired the concept’s “know-what” component, which means *A* has the ability to include DOG in mental representations.

This is where most accounts of Piagetian and neo-Piagetian theories, which describe concepts as structural constructs out of epistemic capacities, end; the same goes for any cognitive architecture which postulates a similar process of concept learning. But Fodor argues that such a process is simply insufficient, as it does not explain compositionality, perhaps the most important restriction on theories of concepts and the one emphasized in (Fodor 2008).

Compositionality

The principle of compositionality has been expressed in many different forms (Fodor 1998; 2008; Robbins 2002; Prinz 2002; Connolly et al. 2007); we will use the version formulated in (Robbins 2002):

Compositionality Constraint The content of a complex (NON-LEXICAL) concept is exhaustively determined by the contents of its constituent concepts and the rules governing the combination of those constituents.

Compositionality is used as a convenient way to explain two other apparent properties of thought: **systematicity**, and **productivity** (Fodor 2008; Robbins 2002; Prinz 2002; Connolly et al. 2007). For an example of productivity, consider that if agent *A* can think of (represent mentally) concepts *C1* and *C2*, then *A* can think of a practically unlimited number of complex concepts which are combinations of the two; for example, (*C1* AND *C2*), (*C1* OR *C2*), (*C1* AND NOT *C2* BUT IF NOT *C1* THEN *C2*), etc. Similarly, due to systematicity, if *A* can represent a complex concept such as (*IF C1 THEN C2*), then *A* can also think (*IF C2 THEN C1*). Both of these properties of thought can be explained by the compositionality constraint as side effects of the compositional nature of concepts.

It is important to understand the distinction between features we use when working with concepts, and what features

are constitutive of conceptual structure. While it’s undoubtedly true that we use our concepts in a variety of ways, the goal here is to figure out what the minimal conditions are in order to possess a concept; we want to know what it means to be able to represent a concept and use it productively and systematically. For example, associations between concepts certainly exist, but the compositionality argument states that since associations between concepts do not compose in a systematic fashion, they are not part of the minimum conditions that need to be met in order to possess a concept.

If one accepts the discussion so far, we run into a problem – *feature emergence* violates compositionality. Let us take a brief look at stereotype theories of concepts, often suggested as an alternative to concept atomism (Connolly et al. 2007). Whereas concept atomism postulates that lexical concepts are fundamental and not constructs of epistemological capacities, stereotype theories propose that concepts are exemplar instances of a certain type, usually inferred through some statistical process (Fodor 1998). For example, a stereotypical PET is a small animal which is not aggressive, which will not eventually grow past a certain size, etc. Presumably, we can observe a certain number of fish and obtain the stereotypical description of them through a statistical average of their features. In the same way, a stereotypical FISH is one that is usually eaten, can be caught with worms at the end of a fishing hook, can be found in the wild, etc. According to compositionality, the content of the complex concept (PET FISH) should be exhaustively determined by the contents of PET and FISH. But there are some properties emergent in the (PET FISH) stereotype that were not present in either the PET or FISH stereotypes. A stereotypical (PET FISH) is a goldfish, which is stereotypically with an orange hue, an inch or two in length, etc. Neither of those properties were properties of the stereotypical PET (although a goldfish is a type of pet, any sort of statistical average of all pets would not yield the feature “orange hue”) or the stereotypical FISH. Even if one were to take the way the stereotypes for PET and FISH combine to create (PET FISH) and use that as a model for concept combination, it would not apply to other cases; (ENERGY DRINK) is an example. The stereotype for a complex concept is not *exhaustively* determined through the stereotypes of its constituent concepts, and therefore, Fodor concludes, stereotypes are not concepts.

One might object at this point: “What prevents an agent from drawing on the non-compositional adjacent properties of a concept (for example, associations between DOG and other concepts: Dogs usually have four legs, are hairy, etc.) during the concept combination process?” One reason is the productivity and systematicity of thought: If there is no relatively systematic way to combine and produce new concepts, how is it that one can form a complex sentence one has never spoken before (“I read a green book about a green alien while I was standing on one foot atop a tall table”), speak it to someone who has never heard that complex sentence before, such that the semantics of the sentence are clear to both speaker and hearer? Applications of the compositionality constraint to theories of concepts is the sort of argument that pervades Fodor’s writings, with perhaps the most promi-

nent use of the compositionality constraint in (Fodor 2008).

From the vantage point we have arrived at, we can see the challenge to cognitive systems theory: Any sufficiently robust cognitive system should, as is reflected in **L**, be able to reason about its propositions, which presumably involves performing computations on the concepts out of which those propositions are composed (assuming something like Computational Theory of Mind is right). Among these computations are productivity and systematicity. Therefore, cognitive systems must either accept compositionality and concept atomism, or explain productivity and systematicity in a way that avoids relying on compositionality.

Responding to Compositionality Recall that compositionality is proposed as a way to explain the apparent systematicity and productivity of thought. However, if one can explain systematicity and productivity without requiring that concepts are totally compositional, then perhaps stereotype theories of concepts can be valid after all. Such an approach is interesting to examine, since it can help to refine both cognitive systems and neo-Piagetian theories.

We start by weakening the compositionality requirement. As Prinz (2002) observed, “it is one thing to say that concepts *must be* compositional, and another to say that they *must be capable* of compositional combination [emphasis added]” (Prinz 2002). This is similar to the distinction of strong vs. weak compositionality observed by (Robbins 2002). Essentially, we abandon the idea that there is only one way that concepts can compose. Instead, concepts can be combined in a way that generates a range of possible interpretations for the ideal representation of the complex concept (hypotheses), which are then narrowed down by the agent.

How does this explain emergent properties in complex concepts like PET FISH? The properties of a stereotypical fish could be extracted from experiences with pet fish, and consistency checks could be done. For example, if when thinking of pet fish one constructs a representation in which a boy has a shark tank in his room, some of the consequences of this representation would have to be evaluated. How would he afford taking care of the shark? Where would he have the space to store it? — and so on. The challenge for implementations of cognitive systems here would be how to generate and answer such questions in reasonable time. The problem then reduces to explaining systematicity and productivity in a way that can replicate human-level performance.

Space does not allow us to give a full summary of the debate on compositionality; for more information see (Robbins 2002; Prinz 2002; Patterson 2005; Fodor 1998; 2000; Weiskopf 2009; Fodor 2008). For now, we can be reasonably satisfied with the assumption that a sufficiently robust cognitive architecture could use a consistency check, a search with the agent’s existing knowledge base/schema to find contradictions, and use the result of that check to evaluate the plausibility of the potential interpretations. This is a back-and-forth, interactive learning process characteristic of neo-Piagetianism; hence appeal to such a process highlights common ground between neo-Piagetianism and cog-

nitive systems theory. Also, it is important to note that such a process of “consistency-checking” may imply the existence of some innate, sufficiently expressive logic-based mechanism to do the checking (we explore this shortly).

Certainly, this approach is not revolutionary. Cognitive systems designers must specify the detailed mechanism by which concepts are generated and manipulated. But this is an example of a way in which cognitive systems theory and neo-Piagetianism can be mutually beneficial. Neo-Piagetianism postulates structured concepts, making it a prime target for Fodorian compositionality attacks. As we have stated, a detailed description of a process which would yield sufficient productivity and systematicity should be enough to escape this challenge. However, like classic Piagetianism, neo-Piagetianism tends to suffer from underspecification of the details of the mechanisms by which concepts combine and develop (Meadows 2006). It is this problem that cognitive systems and simulations can directly alleviate.

For example, Bello, Bignoli, & Cassimatis (2007) applied the Polyscheme (Cassimatis 2006) framework to the false belief task, which is traditionally believed to demonstrate that the ability for a child to represent the beliefs of others (sometimes called ‘second-order’ beliefs) doesn’t fully develop until around age 4 (Meadows 2006). Instead, Polyscheme was able to emulate the emergence of what appeared to be second-order beliefs by simply being made to shift its “cognitive focus of attention when asked to make predictions about the actions of others” (Bello, Bignoli, and Cassimatis 2007). In this way, cognitive systems appropriately focused on demonstrating the plausibility of neo-Piagetian cognitive structures in a real-time environment, keeping in mind the restrictions required by productivity and compositionality, can advance, refine, or refute neo-Piagetian theories.

The Circularity Objection to Concept Acquisition

However, the neo-Piagetian explanation needs a bit more work in order to answer another objection, namely, the claim that circularity infects accounts of concept possession. Consider the evaluation of a potential interpretation *i*. At some level, a mental representation must be constructed (a hypothesis) with *i* as a constituent part in order for the evaluative computation (which, remember, is defined over representations) to work. Since the process of hypothesis construction requires the ability to represent *i*, it cannot be the process by which the agent acquires the ability to represent *i*. If learning the concept *i* is the acquisition of the ability to represent *i*, then this cannot be the process by which *i* is learned, nor can it be any inferential process that presupposes the ability to represent *i*. Therefore, all concepts are unlearned; all concepts must be innate (Fodor 1980; 2008).

This circularity objection appears to threaten the acquisition of simple concepts, leading to the (in)famous Fodorian conclusion that “[t]he only coherent sense to be made of such learning models as are currently available is one which presupposes a very extreme nativism” (Fodor 2008). Before

we sketch out our solution, let us take a step back and summarize the thinking that got us here:

1. Thoughts are computations defined over mental representations.
2. Mental representations are built out of mental concepts.
3. Mental concepts correspond to real concepts in the world, and an agent possesses those concepts iff it can build mental representations using that concept.
4. An inferential process of concept acquisition requires the ability to represent the acquired concept (in order to represent it so that it can be evaluated), and is therefore circular.
5. Therefore, concepts cannot be acquired through inferential processes.

Clearly, the conclusion, 5., is incompatible with the assumptions in **L**: this is what has been referred to as **FP**, “Fodor’s Paradox” (Quartz 1993; Shultz 2004; Carey 2009; Margolis and Laurence 1999).³ The ability to understand a concept involves the ability to represent that concept mentally (assuming Representational Theory of Mind), and acquiring the ability to do so, according to most accounts of learning, involves a process of evaluating potential representations (hypotheses) and matching them up with available evidence. The process is then often repeated until a hypothesis sufficiently matches whatever criteria is set by the agent. But this hypothesis testing presupposes the ability to represent the concept, and so all accounts of concept learning are apparently circular.

Why must we assume that concept possession is a binary property? Is it not possible to encounter a concept, and through normal interactions with that concept slowly obtain a partial understanding of how to mentally represent that concept? Certainly the neo-Piagetian / Pragmatist idea of “knowing how” before “knowing what” is consistent with such an approach (Fodor 2008). If instead, concept acquisition is a process of discrete steps, where an inferential process takes each intermediate step to the next, resulting in a more accurate representation of the target concept, clearly Fodor’s paradox as described here would still apply. Each iteration is a process of hypothesis formation and testing on a small scale.

Alternately, a dispositional theory might then be suggested, where by some automatic means one learns to behave in accordance with a rule rather than having that rule as the subject of an intentional state. At first glance, it looks like this may avoid the circularity involved in hypothesis formation, but Fodor argues that this dispositional account of rule-following will not work. Given a rule *R*:

an account of rule-following that invokes behavioral intentions needs a story about what’s going on when an intention to behave in accordance with *R* is (part of) what explains behavior that does accord with *R*. Well, if the working commitment to RTM and CTM still holds, then what distinguishes following *R* from

mere action in accordance with *R* is that in the former case *R* is mentally represented (‘in the intention box’, as one says) and the mental representation of *R* is implicated in the etiology of the behavior that accords with *R*. But if *that’s* right, then only someone who is *already* able to mentally represent conjunction can intend to follow the rules that constitute the definition-in-use of AND. (Fodor 2008)

The entire class of Piagetian and neo-Piagetian theories relies on the belief that external behavior can be learned pre-symbolically and then turned into a concept or declarative symbol. Similar processes have been called “bottom-up learning” (Sun 2004), M-Sorting to I-Sorting (Weiskopf and Bechtel 2004), and Piaget’s semiotic function (Piattelli-Palmarini 1980). Fodor argues that this resulting symbol must then be subject to compositionality, and since the only way to explain the process by which disposition can map to a compositional symbol is one which presupposes the existence of the symbol, all such symbols must be innate. But, as we have mentioned, a process which produces interpretations of complex symbols in a systematic way which draws on whatever it knows about the concept should demonstrate a consistency that can be empirically compared to human interpretation of the same complex symbols, and would need to do so in order to satisfy the productivity and systematicity constraints.

The objection, then, assumes that the ability to represent a concept when properly activated is innate. If the ability to represent a concept *C* depends on the level of development of an agent’s relevant conceptual structures, then the neo-Piagetian response still looks valid—the reason children can’t represent most adult concepts or act in accordance with complex rules of behavior is because their conceptual structures have not yet developed to a degree of complexity sufficient to represent these rules. If we stick with the neo-Piagetian assumption that development involves an increasing complexity of these structures, then we may be able to safely straddle that line between two extremes: that all concepts are necessarily innate, and saying that all (declarative) thought is simply emergent from lower level, possibly biological, processes. Unsurprisingly, Fodor has another circularity argument which challenges that possibility as well.

The Circularity Objection to Bootstrapping

The objection can be understood in terms of expressibility. If thought is structured in such a way that it can be represented using symbols, and those symbols follow certain inference rules, then those symbols have a certain expressibility whose upper bound is equal to that of a formal logic *L*. If *L* has a certain expressibility, then by definition the concepts bounded by *L* cannot combine in any way to express concepts with complexity greater than *L*. For example, propositional logic cannot express certain statements in first-order logic (FOL), and deterministic finite automata (which are after all in correspondence to a formal logic whose expressibility is below that of FOL) cannot recognize all the

³Although Carey (2009) formulates the paradox similarly, **FP** as formulated here tries to focus more on its non-lexical role.

languages a Turing machine can.⁴

This problem, that of explaining the emergence of higher-level logics spontaneously out of lower-level logics, has been called the “bootstrapping problem” (Carey 2004; Rips, Asmuth, and Bloomfield 2006), and appears in other areas of conceptual development. Susan Carey attempted to explain how children learn the natural numbers using a bootstrapping process (Carey 2004), but was challenged by Rips et al. (2006), who used a circularity argument similar to the Fodorian ones we cite here. Whereas Carey supposed that children infer the inductive property of natural numbers from the memorization of a sequence of names (one, two, three...) and the ability to make analogies, Rips et al. note that given only that evidence, the standard natural number assumption is ambiguous with other possible hypotheses (such as counting modulo 10), and therefore there must be some sort of bias that prefers the inductive property as an explanation of the natural numbers. But if this is the case, then the bootstrapping process is unnecessary (Rips, Asmuth, and Bloomfield 2006), indeed it begins to look more like a Fodor (2008) explanation: that there is an innate concept of natural numbers and it is merely “activated” with experience.

But the Rips paper gave an example of a child who grew up in an environment where malicious aliens willingly taught the child the wrong numbering system (they only taught her how to count modulo 10), and this is a key point. Children do not learn in an isolated environment where they are left to work out each potential hypothesis, as if they were machines fed input and expected to output the correct answer. Instead, it is an interactive process where they test out each hypothesis, by observing the behavior of parents who correct or reward them based on the accuracy of each hypothesis. Again, the neo-Piagetian view is important to escape from the absurd conclusion of extreme nativism.

The idea that concepts develop by increasing in complexity, however, can mean something like the ability to represent more variables at once, a possibility which brings to mind the current research on the role of working memory in cognition (Barrouillet and Gaillard 2011). The expressibility of concepts whose complexity is bounded by the availability of working memory is a different type of limit. However, even in that case, the problem remains: How to explain the origin of logical expressibility?

It seems that only two solutions are possible. Either the child, in the beginning, already has concepts with expressibility equivalent to the maximum expressibility that humans can think, or the developmental process which modifies the conceptual structures has this same maximum expressibility. We must be careful here not to make another regress: When explaining how this process (which may or may not be one of Piaget’s functional invariants) knows how to modify the conceptual structures appropriately, it must be defined in a way that does not presuppose it already has and understands the concepts which it is modifying the structures to be able to represent. Otherwise, we push back the explanation of how

concepts are acquired to this modifying process, and then it is unclear what the need for the conceptual structures are in the first place.

Logic and neo-Piagetianism

Many of the criticisms of Piaget’s original theory centered on his usage of formal logic to describe thinking of children at different stages (Meadows 2006), attacks which led to the current consensus that Piaget was wrong to do so (Case 1992; Meadows 2006). The most successful of these attacks have been studies showing that thought does not conform to the Piagetian logical model as neatly as Piaget and Inhelder (at least initially) believed.⁵ As a result, many neo-Piagetians prefer to describe the qualitative features of the cognitive structures themselves as opposed to using formal symbols (Case 1992).

But we should not throw out the usefulness of formal logic to model mature human cognition just yet. A strong case can be made at least for the following:

F Humans, if neurobiologically normal, nurtured, and sufficiently educated, naturally develop a context-free deductive reasoning scheme at the level of elementary first-order logic.

See (Bringsjord, Noel, and Bringsjord 1998; Bringsjord et al. 2006) for a defense of **F**, the details of which we will not repeat here. Note that **F** does not necessarily imply an automatic emergence of FOL, as this would run contrary to empirical evidence. Rather, we defend a version tied to education. In true neo-Piagetian fashion, it may not be the case that education in formal logic would immediately pervade the entirety of the learner’s conceptual repertoire as Piaget supposed, but it almost certainly can allow for more effective conceptual development at least in some areas of thought. Presumably, one educated in formal logic would have an easier time incorporating FOL-level reasoning to develop his or her schema than one who has had no exposure at all to FOL.

However, recall we earlier mentioned that the functional invariant processes that govern the modification of conceptual structures must have the expressivity of at least first-order logic in order to know how to construct structures capable of representing statements in first-order logic, else they face charges of circularity à la Fodor. This seems unavoidable; whether concepts originate through inductive processes or are emergent from a connectionist architecture, the expressive power to understand and systematically produce them must come from *somewhere*.

The ICARUS cognitive architecture, for example, makes explicit use of conceptual inference as its most basic activity (Langley and Choi 2006). With a module containing built-in inference rules of sufficient expressibility, architectures like ICARUS successfully avoid this Fodorian objection. Drescher (1991), however, references this issue, but quickly brushes the issue aside. Any modern computer is technically only a deterministic finite automaton (or at most a linear bounded automaton), yet we consider it Turing-complete for

⁴Fodor leaves aside the logico-mathematics of the relationships and processes that can hold *between* logics, but this topic is outside the scope of the present paper. See (Bringsjord et al. forthcoming).

⁵See Meadows (2006) for a thorough summary.

many purposes (Drescher 1991). In the same way, a mind that lacks the expressive power of FOL may conceivably learn to behave in a manner resembling FOL thought; yet it is difficult to imagine such thought developing as quickly as it does in (properly educated) children without a preexisting ability to generate hypotheses with FOL's expressivity. Because of this reason, it may not be a good idea to dismiss the bootstrapping problem so quickly.

It seems straightforward to say that if most human thought is structured, and humans can understand and express thoughts in FOL, then the structures themselves must be capable of expressing at least FOL. This does not necessarily suppose the immediate pervasion of FOL into all aspects of a child's life immediately after the child learns it, as Piaget & Inhelder (1958) believed; nor does it require that all human beings inevitably reach the formal operations stage. It does, however, emphasize the importance of using FOL and higher logics in modeling normal adult cognition; it is an essential component of human-level thought, in general.

Conclusion: Lessons for neo-Piagetians and AI Researchers

In this short and non-technical paper, we make no claim to have carried out an in-depth analysis of Fodor's critique of the current understanding of concepts, nor to have provided a thorough account of all possible similarities and differences between cognitive systems theory and neo-Piagetianism. Such a task would be impossible in the scope of this paper, and is the topic of our ongoing research.

Yet there are some key and perhaps not insignificant points to take away from this paper. The first is that neo-Piagetianism and cognitive systems theory have a lot in common, and both stand to benefit from the other. Implementing psychological theories force them to come to terms with real world limitations and to refine or abandon unworkable ideas (Shultz and Sirois 2008), and cognitive systems can borrow fresh ideas from a class of theories it already shares much in common with. The second point is that neo-Piagetians (and the cognitive systems which share their assumptions) must accept a version of Piaget's **F** as we have formulated it, or they risk succumbing to the Fodorian paradoxes. We believe that both of these points represent new approaches to both the theory behind, and the implementation of, human-level cognitive systems.

A few more points to summarize in closing:

- The process of concept acquisition must be more closely analyzed, and neo-Piagetians must define it in a way that avoids the circularity objection, while preserving sufficient compositionality.
- Any cognitive system will need to define compositionality carefully. If accepting strong compositionality, how does it avoid the Fodor paradox and an appeal to nativism? If accepting a weaker version of compositionality, it must describe a process for understanding never-before-encountered complex concepts, and this process must not be so nondeterministic as to generate hypotheses that would not be encountered by a reasonable hu-

man being. The distribution and nature of these hypotheses can presumably be empirically verified by comparing to human-generated hypotheses.

- Cognitive systems cannot commit the fallacy of having a more powerful logic spontaneously generate from a less expressive one, without explaining which process mediates the increase in expressivity, and if it does so it must ensure that this process is defined in a way that does not lead to a homuncular regress.

As far as we can tell, nobody has ever attempted a Fodorian artificial cognitive system, and for good reason. If concept atomism is compatible with strong compositionality, then it presumably explains deterministically the process by which two concepts combine to form a (possibly unfamiliar to the thinker) complex concept C. (However, it would be quite interesting to see if a system that came with a large repertoire of pre-defined concepts, and which then learned using an analogical mechanism, could demonstrate general intelligence.) If a concept is atomic, and not at least partially composed of some subset of inferential properties or facts connected to other concepts, then what is left? What is the point of a concept that one can think about, but not know how to behave in regards to, not know how to draw inferences about, and not associate with any other concepts? Fodor's reductionist program may have oversimplified via too strong a version of compositionality.

Yet, the compositionality requirement cannot be ignored. It is the systematicity and productivity of thought which Fodor wanted to explain, and he correctly brought attention to the details of the fundamental structure of concepts as containing the key to understanding them. We feel that neo-Piagetianism and the work done under this banner offers insight into these structures — insight that should be harvested by those building cognitive systems.

References

- Barrouillet, P., and Gaillard, V. 2011. *Cognitive Development and Working Memory : A Dialogue between Neo-Piagetian Theories and Cognitive Approaches*. New York, USA: Psychology Press.
- Bello, P.; Bignoli, P.; and Cassimatis, N. 2007. Attention and Association Explain the Emergence of Reasoning About False Beliefs in Young Children. In *Proceedings of ICCM 2007 - Eighth International Conference on Cognitive Modeling*, 169–174.
- Bringsjord, S.; Shilliday, A.; Taylor, J.; Bello, P.; Yang, Y.; and Arkoudas, K. 2006. Harnessing Intelligent Agent Technology to Superteach Reasoning. *International Journal* 2:88–116.
- Bringsjord, S.; Taylor, J.; Wojtowicz, R.; Arkoudas, K.; and van Heuvelen, B. forthcoming. Piagetian Roboethics via Category Theory: Moving Beyond Mere Formal Operations to Engineer Robots Whose Decisions are Guaranteed to be Ethically Correct. In Anderson, M., and Anderson, S., eds., *Machine Ethics*. Cambridge, UK: Cambridge University Press.

- Bringsjord, S.; Noel, R.; and Bringsjord, E. 1998. In Defense of Logical Minds. In *Proceedings of the 20th Annual Conference of the Cognitive Science Society*, 173–178. Mahwah, NJ: Lawrence Erlbaum Associates.
- Carey, S. 2004. Bootstrapping & the Origin of Concepts. *Daedalus* 133(1).
- Carey, S. 2009. *The Origin of Concepts*. New York, New York, USA: Oxford University Press, USA.
- Case, R. 1992. Neo-Piagetian theories of child development. In Sternberg, R. J., and Berg, C. A., eds., *Intellectual Development*. New York, New York, USA: Cambridge Univ Press. chapter 6, 161–196.
- Cassimatis, N. 2006. A Cognitive Substrate for Achieving Human-Level Intelligence.
- Connolly, A. C.; Fodor, J. A.; Gleitman, L. R.; and Gleitman, H. 2007. Why stereotypes don't even make good defaults. *Cognition* 103(1):1–22.
- Demetriou, A., and Raftopoulos, A., eds. 2004. *Cognitive Developmental Change : Theories, Models and Measurement*. Cambridge, UK: Cambridge Univ Press.
- Drescher, G. L. 1991. *Made-Up Minds : A constructivist Approach to Artificial Intelligence*. The MIT Press.
- Ehman, J. F.; Laird, J.; and Rosenbloom, P. 2006. A Gentle Introduction to Soar, an Architecture for Human Cognition: 2006 Update.
- Fodor, J. A. 1980. *The Language of Thought*. Cambridge, Massachusetts: Harvard University Press, 2 edition.
- Fodor, J. 1998. *Concepts : Where Cognitive Science Went Wrong*. New York, New York, USA: Oxford University Press, Inc. New York, NY, USA.
- Fodor, J. A. 2000. *In critical condition : polemical essays on cognitive science and the philosophy of mind*, volume 75 of *Representation and mind*. MIT Press.
- Fodor, J. A. 2008. *LOT 2: the language of thought revisited*. Oxford Univ. Press.
- Inhelder, B., and Piaget, J. 1958. *The Growth of Logical Thinking from Childhood to Adolescence*. New York, NY: Basic Books.
- Langley, P., and Choi, D. 2006. A Unified Cognitive Architecture for Physical Agents. In *Proceedings of the Twenty-First National Conference on Artificial Intelligence*. Boston, MA: AAAI Press.
- Margolis, E., and Laurence, S., eds. 1999. *Concepts: Core Readings*. MIT Press.
- Meadows, S. 2006. *The Child as Thinker*. New York, New York, USA: Routledge, 2 edition.
- Patterson, D. 2005. Learnability and Compositionality. *Mind and Language* 20(3):326–352.
- Piaget, J., and Inhelder, B. 1958. *The Growth of Logical Thinking*. New York, New York, USA: Basic Books, Inc.
- Piattelli-Palmarini, M., ed. 1980. *Language and Learning : The Debate between Jean Piaget and Noam Chomsky*. Cambridge, Massachusetts: Harvard University Press.
- Prinz, J. 2002. *Furnishing the mind : concepts and their perceptual basis*. MIT Press.
- Quartz, S. R. 1993. Neural networks, nativism, and the plausibility of constructivism.
- Rips, L. J.; Asmuth, J.; and Bloomfield, A. 2006. Giving the boot to the bootstrap: How not to learn the natural numbers. *Cognition* 101(3):B51–B60.
- Robbins, P. 2002. How to Blunt the Sword of Compositionality. *Noûs* 36(2):313–334.
- Shultz, T. R., and Sirois, S. 2008. Computational Models of Developmental Psychology. In Sun, R., ed., *The Cambridge Handbook of Computational Psychology*. New York, New York, USA: Cambridge Univ Press. chapter 16, 451–476.
- Shultz, T. R. 2004. *Computational Developmental Psychology*. Cambridge, Massachusetts: The MIT Press.
- Sun, R. 2002. *Duality of the Mind: A Bottom Up Approach Toward Cognition*. Lawrence Erlbaum Associates, Mahwah, NJ.
- Sun, R. 2004. Desiderata for cognitive architectures. *Philosophical Psychology* 17(3):341–373.
- Wason, P. 1966. Reasoning. In *New Horizons in Psychology*. Hammondsworth, UK: Penguin.
- Weiskopf, D. a., and Bechtel, W. 2004. Remarks on Fodor on Having Concepts. *Mind and Language* 19(1):48–56.
- Weiskopf, D. a. 2009. Atomism, Pluralism, and Conceptual Content. *Philosophy and Phenomenological Research* 79(1):131–163.