ELSEVIER

COGNITION

Original Articles

Can bootstrapping explain concept learning?

CrossMark

Jacob Beck

Department of Philosophy & Centre for Vision Research, York University, Toronto, Ontario, Canada

ARTICLE INFO

ABSTRACT

Susan Carey's account of Quinean bootstrapping has been heavily criticized. While it purports to explain how important new concepts are learned, many commentators complain that it is unclear just what bootstrapping is supposed to be or how it is supposed to work. Others allege that bootstrapping falls prey to the circularity challenge: it cannot explain how new concepts are learned without presupposing that learners already have those very concepts. Drawing on discussions of concept learning from the philosophical literature, this article develops a detailed interpretation of bootstrapping that can answer the circularity challenge. The key to this interpretation is the recognition of computational constraints, both internal and external to the mind, which can endow empty symbols with new conceptual roles and thus new contents.

If concepts are the constituents of thought, the thoughts we can think are limited by the concepts we can have. It thus matters a great deal whether humans can learn new concepts throughout their lifespans. Enter Susan Carey's magnum opus, *The Origin of Concepts*,[1] which develops some of Quine's metaphors about concept learning into a full-blown account that Carey calls *Quinean bootstrapping*. While *Origin* is widely regarded as a *tour de force*, commentators disagree about whether Quinean bootstrapping (hereafter: bootstrapping) manages to do the work Carey requires of it.[2] In part, this disagreement traces to the difficulty of the task. There are powerful reasons, tracing to Plato's *Meno* and honed in recent decades by Jerry Fodor, to think that concept learning is impossible. But the disagreement also exists because *Origin* is, at least in places, a difficult text to interpret. It is not always clear just what Carey takes bootstrapping to be. In his review of *Origin*, Fodor (2010, p. 8) puts the point this way:

> Reading Susan Carey's book feels a little bit like coming in at the middle of a movie: you can sort of figure out what's going on, but you wouldn't bet the farm that you've got it right.

Fodor, of course, is an outspoken critic of Carey's. But it is not only her critics who have trouble pinning her down. Carey charges many of her would-be allies with misinterpreting her as well.[3]

This paper aims to charitably elucidate Carey's account of bootstrapping—to rewind to the start of her movie and play it back in slow motion, pausing at key points with new distinctions and clarifications. In so doing, I will not defend every aspect of bootstrapping, but I will defend it from one prominent line of criticism. Multiple critics allege that bootstrapping cannot explain how new concepts are learned without circularly presupposing that learners already have those very concepts. Drawing on developments from the philosophical literature on concept learning, I will show how it can.

One caveat. Although the resulting account of bootstrapping is inspired by Carey's writings, I am far from certain that she would endorse every aspect of it. Thus, while I will attribute the account to Carey, one might more cautiously view it as one promising way of developing her views.

1. Introducing bootstrapping

Carey is not interested in just any type of concept learning. A thinker that possesses the concepts FEMALE and FOX, and then learns

[1] Hereafter: *Origin*. Unless otherwise noted, all page references are to this book (Carey, 2009a). Carey's mature account of bootstrapping is also summarized and developed in several articles (Carey, 2004; Carey, 2009b; Carey, 2011a; Carey, 2014).

[2] For broadly sympathetic commentaries, see Shea (2009), Margolis and Laurence (2008), Margolis and Laurence (2011) and Piantadosi, Tenenbaum, and Goodman (2012). For more critical commentaries, see Fodor (2010), Rey (2014), Rips, Asmuth, and Bloomfield (2006), Rips, Asmuth, and Bloomfield (2008) and Rips and Hespos (2011).

[3] Inspired by *Origin*, Piantadosi et al. (2012) present a computational model of bootstrapping, but Carey (2014) objects that it is not really a model of *bootstrapping* (see §4.2 below). See also Carey's (2011b, p. 162) reply to Shea's (2011) interpretation of bootstrapping, as well as the other exchanges between Carey and her commentators in that volume.

that a VIXEN is a FEMALE FOX, arguably learns the new concept VIXEN.[4] It is doubtful, however, that the thinker alters her expressive power since VIXEN has the same content as FEMALE FOX. Such an episode thus would not count as bootstrapping.

Nor is Carey interested in the learning of a single new primitive concept, as when you meet John Doe for the first time and thereby acquire the concept JOHN DOE. Such cases arguably count as increasing one's expressive power (now you can think thoughts about John Doe; before you couldn't), but Carey does not include them in her discussion of bootstrapping.

Rather, Carey is occupied by cases wherein thinkers learn a batch of new concepts all at once that are at least partially inter-defined, such as concepts of positive integers (pp. 287–333), concepts of rational numbers (pp. 344–359), and concepts of physical entities such as matter, weight, and density (pp. 379–411). To relay the difficulty of such episodes of concept learning, Carey deploys the bootstrapping metaphor. But whereas hoisting oneself by one's own bootstraps is literally impossible, Carey believes that learning these concepts is merely difficult.

Three theses form the core of Carey's account of concept learning. First:

Discontinuity: Over development, thinkers acquire new batches of concepts that alter their expressive power.

One type of discontinuity involves a pure *increase* in expressive power, whereby the thoughts one could think prior to the bootstrap form a proper subset of the thoughts one can think after the bootstrap. For example, while many two year olds can recite a portion of the count list ("One, two, three, … "), they don't seem to know what the words in the list mean. If asked for *n* pennies from a pile, or to point to the card with *n* fish, they will respond with a random number of pennies or point to a random card. Moreover, their failures consist of more than ignorance of language. While further experimental probing reveals evidence of representations with quantitative content—including *analog magnitude* representations of approximate numerosities, *object file* representations that track the numerical identity of individual objects in parallel as their spatiotemporal position changes, and *natural language quantifiers* that are a part of each child's universal grammar—Carey makes a persuasive case that these representations all lack the expressive power to represent the integers. Carey concludes that two year olds lack the representational resources to think about the integers. Four year olds, by contrast, have those resources; they succeed on the point-to-a-card and give-me-*n* tasks.[5]

When children first memorize the count list, it serves as a mere placeholder structure. It encodes serial order ("three" comes after "two," which comes after "one"), but the nature of that order is not defined for the children. It's as though they were saying "eeny, meeny, miny, mo." Nevertheless, Carey maintains that this placeholder structure plays a crucial role in explaining how children acquire integer concepts, and that similar placeholder structures play an essential role in other episodes of concept learning.

Placeholder: Placeholders play an important role in generating conceptual discontinuities.

In defense of Placeholder, Carey argues that people who lack the relevant placeholder structures often fail to acquire new networks of concepts. For example, children who grow up in linguistic communities without a count list never become cardinal-principle knowers (pp. 302–4). Moreover, intelligent animals that lack language, such as chimpanzees, can laboriously learn precise integer concepts piecemeal, but never seem to extrapolate beyond those concepts to induce concepts of further positive integers (pp. 329–33). However, an African Gray Parrot that first learned "seven" and "eight" as mere placeholder terms was able to infer their cardinal meanings upon learning their serial locations in an ordered count list (Pepperberg & Carey, 2012). Finally, Carey observes that curriculum interventions that place an emphasis on placeholder structures outperform other curriculum interventions in generating conceptual change (pp. 479–84).

Carey's third thesis is:

Bootstrapping: There is a learning process called *bootstrapping* that draws on placeholders to bridge conceptual discontinuities.

Carey explicitly takes bootstrapping to involve not just a description of succeeding discontinuous conceptual systems, but a *learning process* that explains *how* thinkers get from the first conceptual system to the second. In support of this contention, Carey maintains that children must *somehow* manage to use placeholders to bridge conceptual discontinuities, and that it's hard to believe that learning isn't involved given that conceptual discontinuities are often bridged as a result of instruction and study, with success predicted by the particular curriculum that one's teachers adopt. Of course, any learning process must be *psychologically realistic*. Thus, a bootstrapping explanation of integer concepts must only appeal to representational resources that we are justified in believing that children actually have, such as analog magnitude and object file representations.

Carey's parade case of bootstrapping involves the acquisition of natural number concepts such as THREE, SEVEN, and TEN (pp. 287–333). The bootstrap begins when children memorize the count list placeholder system, typically by age two. Some months later they become "one-knowers." They'll give you one penny or point to the card with one fish, but respond randomly with larger values. Six to eight months later children become "two-knowers." They succeed on the give-me-a-number and point-to-a-card tasks for one or two, but no more. Several months later they become three-knowers, and then sometimes four-knowers. According to Carey, children at this stage have learned to use their object file systems to place models stored in long-term memory in one-to-one correspondence with objects in the world, and to associate such states of one-to-one correspondence with the first four number words. So they know that there is "one" object when the object is in one-to-one correspondence with a model of a singleton in long-term memory {i}; that there are "two" objects when the objects are in one-to-one correspondence with a model of a pair of individuals in long-term memory {j, k}; and so on, up to four (the upper bound of the object file system). Carey calls children at this stage "subset-knowers" and calls the system they use "enriched parallel individuation." Finally, by three-and-a-half or four years of age, children assign meanings to the remainder of the terms in their count list. According to Carey, this happens when children notice a "critical analogy":

The critical analogy that provides the key to understanding how the count list represents number is between order on the list and order in a series of sets related by an additional individual.

---

[4] Carey (p. 5) takes concepts to be mental representations with semantic as well as non-semantic properties that serve as the constituents of thoughts, including beliefs. This minimal characterization should suffice for our purposes.

[5] Carey also recognizes a second type of discontinuity, *conceptual change*, which not only involves thoughts that one can think after the bootstrap that one could not think before the bootstrap, but also thoughts that one could think before the bootstrap that one cannot think after the bootstrap.

This analogy supports the induction that any two successive numerals will refer to sets such that the numeral further along in the list picks out a set that is one greater than that earlier in the list. (p. 477)

Only at this point do children become "cardinal principle-knowers," and thus associate the appropriate integer with each word in their count list.[6] Carey takes this episode of bootstrapping to be typical insofar as it involves two essential stages (pp. 306–7): the construction of a network of symbolic placeholders that are directly related to one another, but not initially mapped onto preexisting concepts; and the subsequent interpretation of those placeholders through non-demonstrative modeling processes such as analogical mapping, thought experimentation, limiting case analysis, induction, and abduction.

2. Two challenges for bootstrapping

2.1. The circularity challenge

Drawing on Fodor's well-known skepticism about concept learning (Fodor, 1975, 1980, 2008), Fodor (2010) and Rey (2014) charge Carey with failing to explain how bootstrapping could increase a thinker's expressive power. Carey, recall, leans on "modeling processes" such as induction, abduction, analogy, limiting case analyses, and thought experimentation (pp. 307, 418, 476). Because these processes "are not deductive" (p. 307), it is tempting to think that they can "go beyond" their inputs and undergird the sorts of conceptual leaps that interest Carey. But Fodor and Rey argue that any such impression is illusory: Carey's modeling processes cannot explain how placeholders are endowed with new contents without circularly presupposing that the thinker possesses concepts with those very contents.

Consider induction, which "leaps" from a finite set of observations to a conclusion about unobservables. The conclusion clearly goes beyond the inputs in the sense that the conclusion does not deductively follow from the inputs. However, the conclusion itself can still be couched in terms of concepts that are available to the thinker. Induction tells you how to transition from 1000 sightings of green emeralds to the conclusion that all emeralds are green, but it doesn't tell you where to get the concepts ALL, EMERALD, or GREEN. In other words, induction selects an ampliative conclusion from your hypothesis space, but it doesn't generate or expand your hypothesis space; it presupposes that space. Likewise, reasoning by analogy can help one to draw novel conclusions about a target domain by comparing it to a more familiar domain, but it presupposes that one has the concepts with which to characterize the familiar domain. Thus, it can be useful to reason about electric circuits by comparing them to hydraulic systems, but only if you have the conceptual resources to characterize the properties of hydraulic systems. A similar circularity worry can be raised about the other processes on Carey's list.

This worry can be made concrete by recalling Carey's appeal to the "critical analogy" that "supports the induction that any two successive numerals will refer to sets such that the numeral further along the list picks out a set that *is one greater than* that earlier in the list" (p. 477, emphasis added). Rey (2014, p. 117) objects, "But here 'is one greater than' expresses the very concept of SUCCESSOR whose acquisition Carey is trying to explain. In the first place, one can ask how this concept even *occurs* to the child." Rey concludes that the child must *already* have the concept SUCCESSOR to entertain the analogy that he takes Carey to credit with generating that concept.

2.2. The deviant-interpretation challenge

In his critique of Carey's account of bootstrapping, Rey (2014) presses a second challenge. Why don't children interpret the placeholders in a deviant or non-natural way? Recalling Goodman's (1955) famous grue paradox and Kripke's (1982) famous discussion of Wittgenstein on rule following, he writes:

> Specifically, Carey's appeal to 'induction' as the way the child projects beyond the first three members of the sequence of cardinal numbers is directly prey to Goodman's point: simply observing the correspondence between the first three cardinal sets and first three ordinals (or any sequence of sets and finite sequence of terms) is compatible with an infinite number of functions compatible with the finite data the child has seen, e.g. *à la* Goodman, $Y = X + 1$ IF X IS ENCOUNTERED BEFORE 3000 CE, OR OTHERWISE 2, or, to take the kind of example discussed by Kripke (1982) in his exposition of Wittgenstein, $Y = X + 1$ UNLESS $X = 57,453$, IN WHICH CASE $Y = 2$,—and, again, and so on.
>
> [(Rey, 2014, p. 120)]

As Rey goes on to note, Goodman was concerned with the normative question of how good inductions ought to proceed whereas Carey is instead concerned with the descriptive question of what it is about children's psychology that makes their inductions actually proceed as they do. But Rey worries that Carey fails to answer even the descriptive question.

2.3. Setting aside the deviant-interpretation challenge

The circularity and deviant-interpretation challenges are independent of one another. The deviant-interpretation challenge assumes that thinkers have the conceptual resources with which to formulate the hypothesis that (say) "eleven" refers to eleven, but asks why children endorse that hypothesis over the various deviant hypotheses that compete with it. The circularity challenge, by contrast, questions how children could even formulate the hypothesis that "eleven" refers to eleven in the first place.

Unfortunately, the circularity and deviant-interpretation challenges are not always clearly distinguished, sowing confusion and engendering crosstalk between bootstrapping's critics and defenders. For example, in their critique of bootstrapping Rips et al. (2006) imagine two children who have memorized the count list up to "nine" as an uninterpreted placeholder structure, but then interpret it in two different ways. One child interprets the count list in the standard way, as embodying the successor relation. But the other interprets it as embodying a non-standard cyclical system according to which "one" refers to sets with 1 or 11 or 21... objects, "two" refers to sets with 2 or 12 or 22... objects, and so on. Rips et al. object that Carey fails to explain why children generalize to the integers rather

---

[6] Some researchers have challenged Carey's interpretation of the experimental data. For example, Spelke (2011) argues that Carey underestimates the role that analog magnitude representations play in children's acquisition of natural number concepts, Gelman (2011) argues that Carey underestimates children's innate capacity to represent discrete and continuous quantities precisely, and Davidson, Eng, and Barner (2012) present evidence that children don't truly understand the successor relation until some time after they become cardinal principle-knowers. For our purposes, we can set such empirical worries to one side. Our question is not whether Carey has the details of any particular episode of bootstrapping just right, but whether she has succeeded in isolating a general learning process that has the potential to explain how conceptual discontinuities are bridged.

than to the deviant cyclical system. This makes it sound like they are concerned with the deviant-interpretation challenge. But Rips et al. explicitly insist that their worry is not merely a rehash of Goodman and Kripke's concerns since it shows not just that the child's concepts are "ambiguous between rival numeral systems" but that they are "completely vague outside the counting range 'one' to 'nine'" (2006, p. B58). While Rips et al. don't elaborate on what they mean by vagueness (is there supposed to be a Sorites paradox lurking here?), they may mean that the child's concepts lack the expressive power to precisely formulate either numerical system, in which case their real concern is the circularity challenge, not the deviant-interpretation challenge. This interpretation is supported by their suggestion that children would need something like the Peano axioms in order to formulate the correct hypothesis, but would then "already have the natural number concept" (2006, p. B59). Perhaps motivated by these remarks, Margolis and Laurence (2008) implicitly interpret Rips et al. (2006) as pressing the circularity challenge. They object that Rips et al. underestimate the conceptual resources that children begin with—resources that, they argue, are sufficient to formulate the proper induction. Rips et al. (2008) in effect respond that Margolis and Laurence don't answer the deviant-interpretation challenge; even if children have the concepts they need to formulate the proper induction, we are left with no explanation of why they choose that induction rather than one of the many deviant inductions that they also have the conceptual resources to formulate. So perhaps Rips et al. *do* mean to be raising the deviant-interpretation challenge. If so, however, it remains unclear how their version of this challenge is meant to differ from a descriptive version of the traditional problem associated with Goodman and Kripke.[7]

In any case, I want to set the deviant-interpretation challenge aside. While it's a deep and stubbornly difficult problem, it's not specific to theories of *concept* learning but rather afflicts all theories of inductive learning, including theories that seek only to explain how new *facts* are learned. Even the most ardent concept nativist would still need to explain why we infer that all emeralds are green rather than grue. It hardly seems fair, then, to reject Carey's account of concept learning on the grounds that it includes—as one of its elements—an appeal to induction. The situation would perhaps be different if induction formed the totality of Carey's account of bootstrapping, but there is much more to it than that. As will become clear later, the most innovative elements of bootstrapping—including the elements that enable it to answer the circularity challenge—are independent of its appeal to induction. The remainder of this paper will thus set the deviant-interpretation challenge to one side in order to focus on the circularity challenge.

3. Two interpretations of bootstrapping

To understand how Carey addresses the circularity challenge, some distinctions are needed. If a person is currently using a concept, I will say that she is *deploying* the concept. For example, when you actively think that the cat is on the mat, you deploy the concept CAT. By contrast, a concept is *available* to a person just in case she could deploy it in reasoning, thinking, categorizing, remembering, and other cognitive processes without much effort, simply by endogenously shifting her attention. Thus, if you are busy thinking about the cat being on the mat, then presumably you are not deploying your concept PENGUIN. Nevertheless, the concept is available to you. You could deploy it simply by initiating the appropriate shift in your attention. Precisely how many concepts a person can deploy at any one time is open to debate, but will depend on such factors as the capacity of working memory and the extent to which occurrent thought operates serially or in parallel. Whatever the number, it is surely quite small in comparison to the number of concepts that are available to a person.

We can further recognize a class of *latent* concepts that are unavailable yet expressible in terms of concepts that are stored in the thinker's mind. One way for a concept to be latent is for it to be *unobviously composed* from one's available concepts. Consider the concept BURSE, which is satisfied by an object just in case it is either green and circular, or blue and enclosed by a prime number of sides, or red and preceded in presentation by a yellow triangle. We can suppose that each of the constituent concepts in this definition—GREEN, CIRCULAR, OR, AND, PRIME, … —are available to you. But of course you do not normally go around categorizing things in this peculiar way, and would find it rather difficult to do so. This gerrymandered combination of concepts isn't available for you to deploy in cognition even though each constituent is. It takes more than an easily executed endogenous shift in attention for you to put the individual concepts that constitute BURSE together and deploy them. As a result, BURSE counts as latent for you (though it might become available with sufficient practice).[8]

There is a second way for a concept to be latent. It can be fully formed and stored in the mind, but isolated from general cognitive processes such as reasoning and categorizing. For example, perhaps the concept GRAVITATIONAL MASS was stored inside an innate module in your head, awaiting the proper inputs to be released (presumably inputs you received in your first physics class). Following Rey (2014), we can say that such concepts are *lying in wait*.

Finally, I'll say that a concept is *foreign* just in case it is neither available nor latent. For example, the concept GRAVITATIONAL MASS is presumably foreign for Fido the dog. It is not available to Fido, nor composable from concepts that are available to Fido, nor lying in wait in some module of Fido's. Presumably there are likewise concepts that are currently foreign for you, though of course you cannot say what they are. Whether any concepts that are currently available to you were at one point foreign for you is controversial. For instance, it is controversial whether GRAVITATIONAL MASS was foreign for you prior to your first physics class, or merely latent in you.

---

[7] Rips et al. (2008, p. 944) again try to distance their objection from traditional worries about induction such as Goodman's by appealing to vagueness.

> The usual problem is justifying one possible inductive conclusion over another (e.g., linear extrapolation of data vs. polynomial extrapolation of the same data). By contrast, the problem with the Induction is that its conclusion is entirely vague about the continuation and hence consistent with many different correlations between numerals and set sizes. It would take a separate inference of a different kind to decide among these possibilities.

Here Rips et al. seem to take the usual problem of induction to concern choosing between *two* options (say, green vs. grue) and their worry to concern choosing among *many* options. But this can't be right since the usual problem includes infinitely many options as well. We can define "grue₁" as green before 3000 and blue afterwards; "grue₂" as green before 3001 and blue afterwards; etc. Goodman wondered why the hypothesis that all emeralds are green should be preferred to alternative hypotheses containing any one of these related predicates. There wasn't one particular date that exercised him!

---

[8] Not all complex concepts will count as latent. For example, the concept HAPPY MAN is complex, but for typical human thinkers it is available, not latent. There may be individual differences. Compositions that are obvious to you might be unobvious to me, and vice versa. There will also surely be borderline cases that are difficult to classify as obvious or unobvious (even for a thinker at a time), though for our purposes it is sufficient that the distinction admits of many clear cases on either side.

With the distinctions between available, latent, and foreign concepts at hand,[9] we can now distinguish a thinker's *available expressive power*, the expressive power of her available concepts, from a thinker's *total expressive power*, the expressive power of her available *and* latent concepts combined, and recognize an ambiguity in Discontinuity.

> Modest Discontinuity: Over development, thinkers acquire (i.e., make available) new batches of *latent* concepts that alter their *available* expressive power.

> Radical Discontinuity: Over development, thinkers acquire (i.e., make available) new batches of *foreign* concepts that alter their *total* expressive power.

We can consequently distinguish two versions of Bootstrapping.

> Modest Bootstrapping: There is a learning process called *bootstrapping* that draws on placeholders to bridge *modest* conceptual discontinuities.

> Radical Bootstrapping: There is a learning process called *bootstrapping* that draws on placeholders to bridge *radical* conceptual discontinuities.

We thus reach a crucial interpretive fork: Is bootstrapping supposed to be modest or radical?

After considering and rejecting two versions of Modest Bootstrapping in Section 4, I'll develop and defend a version of Radical Bootstrapping in Sections 5–6.

4. Modest bootstrapping

4.1. Modest bootstrapping and the circularity challenge

The circularity challenge asks how bootstrapping can alter one's expressive power without circularly presupposing that one had concepts with the new expressive power all along. We can now see that there are two versions of this challenge, one that asks how bootstrapping can alter a thinker's *total* expressive power and a second that asks how bootstrapping can alter a thinker's *available* expressive power. Modest Bootstrapping does not endeavor to answer the former version. In fact, it is typically endorsed precisely because its proponents judge it impossible to learn foreign concepts. But proponents of Modest Bootstrapping do offer explanations of how a thinker's available expressive power can be altered without circularity. And they correctly point out that a change in available expressive power is nothing to scoff at. When you increase your available expressive power, there are thoughts

that are newly available for you to deploy. There is thus a real psychological difference between a person who has a concept only latently and a person who has that same concept available for deployment. An emphasis on total expressive power masks this real cognitive change by blurring the distinction between relabeling a concept that was already available (e.g., by introducing EMAIL to stand for ELECTRONIC MAIL) and making a concept that was previously only latent available for the first time. As we'll see in the next two subsections, however, Carey has empirical grounds to reject modest accounts of bootstrapping for the episodes that interest her.

4.2. Unobvious composition

One way Modest Bootstrapping might work is through unobvious composition. By composing available concepts in new and unobvious ways, thinkers could increase their available expressive power. For example, through enough practice sorting things according to whether they are *burse* or not, you could make the latent concept BURSE available.

While BURSE is a toy example, various mathematical concepts, such as CONTINUUM, could also be learned in this way since they are constructed from familiar and widely available concepts, albeit none too easily. Moreover, although BURSE admits of a Boolean definition, other latent concepts might have different sorts of structures. For example, perhaps certain concepts have prototype structures. If so, some latent concepts might be constituted by sufficiently unobvious complexes of weighted features. Or perhaps certain concepts are defined by their place in a theory. If so, some latent concepts might be constructed from sufficiently unobvious Ramsey sentences (see Rey, 2014, pp. 114–116 and Section 5.4 below).

Piantadosi et al.'s (2012) account of how children learn the integer concepts exemplifies this unobvious-composition strategy. They propose a Bayesian statistical model whereby children build positive integer concepts from a variety of primitive representations and operations implemented in an innate lambda calculus, including set-theoretic and logical operations, operations over words in the counting routine, operations that test whether a set has one, two, or three members, and recursion. The model's primitives are sufficiently rich to explain how the successor function could be arrived at compositionally, but without needing to suppose that the successor function is available to children from the outset.

Piantadosi et al. (2012, p. 200) write that their approach "is inspired by the bootstrapping theory of Carey," but Carey unequivocally rejects it. She objects that Piantadosi et al. "merely assume—without evidence—that full general resources of lambda calculus and logic are available for the generation of hypotheses about what 'one', 'two', 'three', 'four', 'five'... through 'ten' mean" (2014, p. 151). Thus, Carey does not deny that new concepts *can* be learned through unobvious composition. She is just skeptical of theories that claim that they are so learned by being composed from conceptual resources whose existence is empirically unsupported. Moreover, Carey goes to great lengths to show that the resources that we *do* have empirical evidence to attribute to children at the time of the bootstrap to integer concepts—such as analog magnitude and object file representations—are too impoverished to generate definitions of the integers.[10]

It is instructive to compare Carey's skepticism towards the unobvious composition strategy for integer concepts with Fodor's general skepticism towards the strategy. Fodor is skeptical that

---

[9] These distinctions are partially inspired by Rey's (2014, p. 112) helpful distinction between *manifested* and *possessed* concepts, though our taxonomies are marked by differences as well as similarities. His *manifested* concepts are roughly my *available* concepts, but I prefer the term "available" because it stresses that being *deployed* is not what's at issue. I also interpret his category of *possessed* concepts as roughly corresponding to my category of *latent* concepts, but I take issue with the way Rey explicates possessed concepts. Early in Rey, 2014 (p. 112) he says, "I shall take it as a useful point of agreement and departure for the moment that 'possession' is the relation that Chomskyan linguists think neonates bear to 'Universal Grammar,' before they begin manifesting this grammatical competence." One problem with this analogy, however, is that elements in a Universal Grammar are subdoxastic, and it is in the nature of subdoxastic states that they never become available for conscious deployment. Yet on Rey's view, possessed concepts can become available for conscious deployment. A second problem is that there are multiple ways for a concept to be latent, and Universal Grammar really only exemplifies one of these ways. Later in his article (p. 125) Rey offers a different formulation. He says, "one might usefully think of the set of concepts that are innately possessed as being just that set that *can come to be manifested by learning and bootstrapping*." Yet as Carey (2014) points out, this formulation threatens to be empty since everyone who thinks that a concept can be learned will then have to agree that it is innately possessed—even empiricists. A further difference between our taxonomies is that Rey has no term for the category of concepts that I call "foreign."

[10] For further criticisms of Piantadosi et al.'s model, see Rips, Asmuth, and Bloomfield (2013).

unobvious composition could underpin concept learning because he takes it to be a general fact that most lexical concepts (i.e., concepts such as DOG, PAINT, and CHAIR that are expressed by single morphemes in natural languages) lack definitions (Fodor, 1981, 1998). Carey shares Fodor's concern for many concepts, but not for integer concepts, which clearly can be defined (for example, in terms of Hume's Principle and second-order logic). For integer concepts, Carey's concern is that they lack definitions *in terms of other concepts that are available to the learner*. Thus, bootstrapping of integer concepts has to be more than unobvious composition.

4.3. Selection

There is, however, another way that Modest Bootstrapping could work: by selecting concepts that are lying in wait and then making them available. Here is a cartoonish version of the idea that at least helps to paint the picture: inside the head of every child is a *latent concepts box* that contains every concept that can become available for the child. Thus, barring disability, a child's latent concept box will contain the concepts SUCCESSOR, DENSITY, SPECIAL RELATIVITY, HIPSTER, SOCIAL MEDIA, etc., as well as the concepts that no one is presently in a position to specify but that future humanity has the potential to uncover. To start with, however, the child cannot use these lying-in-wait concepts in general cognitive processes because those processes take place outside of the box and the concepts are stuck inside the box. In order for a concept to escape the box, the child needs to identify it as answering to her explanatory purposes. Only then can the concept be selected to join the rest of the child's available concepts and participate in general cognitive processes.

Although no one to my knowledge has actually posited a latent concepts box, Rey (2014) notes that the idea that new concepts are learned by being *selected* from a limited stock of innate concepts that are lying in wait has echoes in how a grammar is learned according to versions of Chomskyan linguistics whereby learning a grammar is a matter of selecting one of a limited number of possible grammars, all of which are innately represented. If concept learning worked similarly, that would make room for "the ecumenical view" that concepts are both learned and innate (Gross & Rey, 2012; Rey, 2014). Concepts could be learned by being selected from one's innate stock of lying-in-wait concepts.

Chomskyans are motivated to posit a universal grammar, in no small part, by how effortlessly children learn their native grammars. But while some concept learning is similarly effortless, consisting in what Carey and Bartlett (1978) call "fast mapping," the concept learning that Carey seeks to explain through bootstrapping is anything but. It takes children a year and a half to become cardinal principle-knowers after they memorize the count list. And that's just the beginning of their understanding of numbers. Many children never manage to make the conceptual leap to an understanding of fractions—a conceptual leap that marks the difference between scoring above and below the mean of 500 on the SAT (p. 346). A simple selection-based account thus fits poorly with the difficulty of many cases of concept learning.

It is possible, however, to combine the selection and unobvious composition accounts in a way that would explain why concept learning is so difficult. For example, one might suppose that the concepts needed to grasp Hume's principle and second-order logic are innate and lying in wait to be selected, and that concepts such as SEVEN are constructed therefrom. Since such a construction is far from trivial, one would expect children to find it difficult to learn the integers. Some such hybrid view that combines selection with unobvious composition seems to be what Rey (2014) ultimately has in mind for his ecumenical account of concept learning. And while Rey rejects Carey's account of bootstrapping for the reasons canvassed in Section 2, he proposes that a tamed version of bootstrapping, interpreted as a species of his ecumenical account, could overcome those worries.

Carey's (2014) principal objection to Rey's proposal is that it goes unsupported by the empirical evidence. For the conceptual discontinuities that interest her, there is simply no evidence that they are engendered by a combination of selection and unobvious composition. For example, Rey offers no specifics about how the integer concepts are generated. Are they constructed along Fregean lines with the concept ONE-TO-ONE CORRESPONDENCE lying in wait? Are they constructed along Peano's lines with the concept SUCCESSOR lying in wait? Rey doesn't say. Nor does he provide evidence that would favor one of these interpretations or any other. Rather, Rey's case for his ecumenical account seems to be based entirely on an argument from elimination: Radical Bootstrapping is incoherent; Modest Bootstrapping implemented through some combination of selection and unobvious composition is the only alternative; thus bootstrapping has to be a combination of selection and unobvious composition. The next two sections challenge this argument by targeting its first premise.

5. Computational constraints

5.1. Radical bootstrapping and the circularity challenge

Fodor and Rey are perfectly aware of the difference between an increase in available expressive power and an increase in total expressive power.[11] It is precisely because they take Carey to be committed to the latter, and thus to Radical Bootstrapping, that they are so skeptical of her ability to answer the circularity challenge. If bootstrapped concepts start off as foreign, they can't be constructed from the learner's other concepts. And if they can't be constructed from the learner's other concepts, there would seem to be nowhere for them to come from. Underlying this worry is an assumption that Fodor often makes explicit: the only way for concepts to be learned is through some form of hypothesis fixation in which to-be-learned concepts are built compositionally from other (available or latent) concepts (Fodor, 1975, 1980, 2008). Following Margolis and Laurence (2011), we can call this assumption the *building blocks model* of concept learning. Once the building blocks model is accepted, unobvious composition (perhaps supplemented by selection) seems to be the only option. It's Modest Bootstrapping or bust.

Like many of Fodor's critics, Carey rejects the building blocks model (pp. 513–514; 2014, pp. 138–139). The challenge confronting these critics, Carey included, is to characterize an alternative learning process by which foreign concepts can be made available. For Carey, of course, that alternative is supposed to be Radical Bootstrapping. The question is thus how Radical Bootstrapping works. And the key to answering this question, I will now argue, is the notion of a *computational constraint*.

5.2. Introducing computational constraints

When summarizing the processes involved in bootstrapping, Carey writes that they

> create explicit representations of knowledge previously embodied in *constraints on the computations* defined over symbols in one or more of the systems being integrated. (p. 418, emphasis added)

And when summarizing how children bootstrap concepts of the positive integers, Carey writes,

---

[11] See Rey's (2014) (p. 127) distinction between "functioning psychological expressive power" and "semantic expressive power" and Fodor, 1975 (pp. 84–6), 1980 (p. 151), and 1981 (pp. 270–1).

The child creates symbols that express information that previously existed only as *constraints on computations*. Numerical content does not come from nowhere, but the process does not consist of defining "seven" in terms of symbols available to infants. (p. 477, emphasis added)

Carey thus clearly maintains that bootstrapped concepts such as SEVEN are constructed, at least in part, from information that is implicitly coded in constraints on computations.

An analogy may help to motivate this idea. Just as any logic needs not only axioms but also rules of inference, any computational theory of mind that posits explicit representations also needs procedures that govern how those representations can be manipulated. Much of the power of a logical or mental system derives from its inference rules or procedures. But inference rules and procedures are also limited in a way that axioms and explicit representations are not. They cannot feature as premises or conclusions in reasoning, and so cannot be directly manipulated by other inference rules or procedures. They thus exist only as constraints on computations. Carey's basic insight is that the procedural constraints on computations that govern explicit representations can substantively and significantly contribute to the process of concept learning despite the fact that they are not themselves representations.

The importance of computational constraints is familiar from other areas of psychology, such as vision science. Because there are infinitely many possible ways of converting the two-dimensional array of light intensity values on the retina into a three-dimensional image, vision scientists posit computational constraints—for example, that light comes from above—to explain why the conversion proceeds as it actually does. Although these constraints are contingent, they generally yield veridical percepts in ecologically valid contexts. Carey likewise holds that there are contingent computational constraints that govern bootstrapping.

As it happens, this interpretation of bootstrapping allies it with several alternatives to the building blocks model that have arisen from the philosophical literature (Block, 1986; Laurence & Margolis 2002; Margolis 1998; Margolis & Laurence 2011; Strevens 2012; Weiskopf 2008). Emerging from these alternatives are two key lessons. The first is that it can be helpful to think about concept learning in terms of how a blank mental symbol acquires the *metasemantic* properties in virtue of which it inherits its representational content. The second is that concepts—whether available or latent—are not the only things that can help to explain how new metasemantic properties are acquired. At the very least, there are also computational constraints.

In the next two sub-sections, I'll review two philosophical accounts of concept learning that embody these lessons and inform Carey's own approach. The first, due to Block (1986), illustrates how computational constraints that are formulated over *external* symbols can help to increase one's total expressive power. The second, due to Margolis (1998) and Laurence and Margolis (2002), shows how computational constraints that are *internal* to the mind can do the same.

5.3. External computational constraints

A *semantic* theory for a symbol system assigns meanings to symbols. For example, a semantic theory for French might assign the meaning *dog* to the word "chien." A *metasemantic* theory for a symbol system explains what the ultimate facts are in virtue of which symbols have their meanings. One metasemantic theory, *conceptual role semantics*, takes its cue from the Wittgenstein (1953) idea that meaning is determined by use. Thus, the word "chien" is not intrinsically meaningful; its meaning is determined by the way French speakers use it. Proponents of concep-

tual role semantics apply this idea to mental symbols, maintaining that the content of a mental symbol is determined by its conceptual role—its dispositional causal relations to other symbols and the world. Following Block (1986), Carey (p. 514ff) endorses a dual-factor conceptual role theory in which conceptual role determines one factor of a concept's content (its so-called *narrow content*) all by itself, and determines the other factor of a concept's content (it's so-called *wide* or *referential content*) in conjunction with contextual factors. For Carey, conceptual role is thus the most significant metasemantic property that concepts have. So in thinking about how a concept is acquired one can ask how a blank mental symbol comes to have a new conceptual role—i.e., how it comes to be *used* in a new way by operations in the mind.

To explain how *external constraints* can endow a symbol with a new conceptual role, Carey rehearses Block's description of how he learned concepts from physics such as ENERGY and MOMENTUM.

When I took my first physics course, I was confronted with quite a bit of new terminology all at once: "energy", "momentum", "acceleration", "mass", and the like… I never learned any definitions of these new terms in terms I already knew. Rather, what I learned was how to *use* the new terminology—I learned certain relations among the new terms themselves (e.g., the relation between force and mass, neither of which can be defined in old terms), some relations between the new and the old terms, and, most importantly, how to generate the right numbers in answers to questions posed in the new terminology (Block, 1986, pp. 647–648; quoted in Carey 2009, p. 419)

The new terms ("energy," "momentum," etc.) served as meaningless external placeholders for Block. And the relations he memorized among those placeholders (e.g., "F = ma") along with the rules he learned for how the placeholders could be manipulated to generate "the right numbers in answers to questions posed in the new terminology" constrained the conceptual roles that the placeholders eventually acquired. These memorized relations and rules thus served as external constraints on the way these placeholders were used in his mind, and thus the computations they entered into. In other words, they served as external computational constraints. Carey maintains that the process Block describes is an essential component of bootstrapping, and in Section 6 we'll see how she applies it to some actual case studies.

One might object that Block's suggestion still falls prey to the circularity challenge since any external computational constraints need to be *interpreted* by the learner if they're to generate new concepts, and the learner can only interpret those constraints if she has the new concepts to start with. But the apparent plausibility of this objection trades on an equivocation.

On the one hand, if "interpreted" is taken to mean *defined in terms of the learner's other concepts*, then it's true that the learner can only interpret the constraints if she has the new concepts to start with (at least latently). But the whole point of Block's example is that external constraints needn't be so "interpreted" in order to generate new concepts. In his physics course Block never learned a definition of the form MASS = C AND D BUT NOT E (where *C*, *D*, and *E* are concepts that were available to Block at the outset). While Block did learn some relations among the new symbols and his old concepts, those relations were not a definition and did not suffice on their own to give the symbols their new meanings. Block also had to learn a bunch of relations among the symbols directly (e.g., "F = ma") in virtue of which they acquired fundamentally new conceptual roles, and, therefore, fundamentally new meanings. The process of concept learning Block describes doesn't fit the building blocks model.

On the other hand, if "interpreted" is taken in a looser sense to mean *endowed with a meaning*, then it's true that external constraints need to be interpreted by the learner, but it's false that the learner needs to possess the new concepts (even latently) prior to the interpretation. Since the new concepts aren't learned by being composed from old concepts à la the building blocks model, the new concepts needn't be expressible in terms of the old concepts.

Rey objects that Block's proposal only gives the *illusion* of circumventing the building blocks model and thus overcoming the circularity challenge. The problem, according to Rey, is that "what Block is presenting here is a description of how one might learn a slew of terms at once by way of a 'Ramsey sentence'," and a Ramsey sentence is "a straightforward logical construction out of concepts already understood" (Rey, 2014, pp. 114–115). Without getting bogged down in the technical details, a Ramsey sentence takes advantage of techniques in quantificational logic to enable multiple new terms to be simultaneously inter-defined with one another and the language's old terms (Lewis, 1970; Ramsey, 1931). Rey (2014, p. 116) grants that a Ramsey sentence may not be *explicitly* represented in the learner's brain, but holds that "the expressive power of the network of roles could be captured by such a sentence." Thus, Rey concludes that the method of concept learning that Block articulates cannot increase one's total expressive power.

It is one thing, however, to point out—as Rey correctly does—that Ramsey sentences have the expressive power to define each of a number of interrelated concepts in terms of preexisting concepts and one another, and quite another matter to suppose that learners have the conceptual resources to construct Ramsey sentences. Both assumptions are necessary if the newly learned concepts are to count as having been latent in the learners all along. Yet the conceptual resources needed to construct Ramsey sentences include some fairly sophisticated logical machinery, much of which wasn't articulated by logicians until the late 19th Century. It is thus surprising that Rey provides no evidence that the relevant logical concepts are available to, or latent in, young learners. By contrast, Carey provides a wealth of evidence that each input to the bootstrapping process as she characterizes it is available to young learners. Moreover, Carey expresses skepticism that sophisticated logical concepts are available to young learners since her "current guess is that innate logic is largely implicit, embodied in computations" (2014, p. 146). Thus, while it may be safe to assume that a college-aged *Ned Block* had the requisite logical concepts to construct Ramsey sentences, it would be question-begging to assume that all three- and four-year-olds have them (even latently) when they acquire integer concepts.

Carey's point thus stands that computational constraints among external symbols have the potential to generate new conceptual roles that in turn increase one's total expressive power. Of course, whether such constraints *actually* increase a thinker's total expressive power will depend on auxiliary empirical assumptions about the concepts the thinker starts with prior to the episode of bootstrapping, including assumptions about their logical concepts. But that's always the case, and amounts to little more than the observation that Carey's hypothesis is empirical and can thus be overturned by evidence that the conceptual starting point is richer than she hypothesizes.

5.4. Internal computational constraints

To see how *internal constraints* can support an increase in expressive power, it is helpful to review Margolis's (1998) account of how natural kind concepts are learned. Margolis explicates his account against the backdrop of Fodor's (1990) asymmetric-dependence theory of content, which provides a referential metasemantics for mental symbols. To a first approximation, Fodor's asymmetric-dependence theory holds that a symbol represents a property when the symbol and property stand in an appropriate law-like causal relation. To explain how a symbol and property enter into the appropriate causal relation, Margolis appeals to a learning mechanism (or *sustaining mechanism* as he calls it, since it sustains the appropriate causal relation) with two components: a *kind syndrome* and an *essentialist disposition*. The kind syndrome is roughly a perceptually based prototype, and the essentialist disposition treats instances as members of the same kind just in case they have the same essential property as exemplars of the syndrome. Together, the kind syndrome and essentialist disposition sustain the appropriate causal relation between a newly introduced symbol and the natural kind to which it refers. Thus, when a learner first comes across a tiger, her natural kind learning mechanism creates a symbol that gets associated with her prototype of a tiger, and classifies things as a tiger just in case they have the same essential property as prototypical tigers. As a result, the learner's newly created symbol bears the appropriate causal relation to tigers.

Crucially, this account does not posit a prefabricated TIGER concept that is merely selected. Nor does it claim that the concept TIGER is composed from other concepts—not even from the concept SAME NATURAL KIND AS. For while we can gloss the essentialist disposition as being implicitly governed by the constraint *same natural kind as*, the learner need not have any explicit representations with that content. The essentialist disposition is an internal constraint on computations that facilitates concept learning.

Carey (p. 518; 2014, p. 142) writes approvingly of Margolis's explanation of how concepts such as TIGER are learned, with one small amendment. She swaps the metasemantics of Fodor's asymmetric-dependence theory for the dual-factor conceptual role theory she prefers. In her hands, the computational constraints of the essentialist disposition help to explain how the concept TIGER gets its distinctive conceptual role (rather than explaining how it enters into the appropriate law-like causal relation). But the upshot for the circularity challenge is the same: internal computational constraints contribute to a blank symbol's acquiring new metasemantic properties in virtue of which the symbol inherits a new content that increases the thinker's total expressive power.

Not only does Carey endorse Margolis's natural kind learning mechanism, she maintains that similar learning mechanisms can be found throughout the animal kingdom. For example, clever experiments in a planetarium show that the indigo bunting, a small bird that navigates by the stars, learns the identity of the North Star by locating it at the sky's center of rotation (pp. 15–16; 2014, pp. 140–141). Buntings have a domain-specific learning mechanism that takes perceptions of the night sky as inputs and yields representations of the North Star as outputs. This mechanism is governed by a computational constraint that implicitly encodes the content *center of rotation*, enabling the representation NORTH STAR to be learned without being defined compositionally from other concepts. As we'll see, Carey thinks that there are also internal constraints that help to underpin Radical Bootstrapping in humans.

Margolis's account of concept learning has not gone unchallenged. Fodor (2008, p. 144) objects:

'You can learn (not just acquire) A' and 'Learning A is sufficient for acquiring B' just doesn't imply 'You can learn B'. For, the following would seem to be a live option: If you acquire a concept by learning a theory, then something is learned (namely, the theory) and something is (merely) acquired (namely, the concept); but what is learned isn't (merely) acquired and what is (merely) acquired isn't learned.

In other words, Fodor is prepared to grant that kind syndromes can be learned, and that (given the existence of an essentialist disposition) learning a kind syndrome is sufficient for acquiring a concept, but he denies that the concept is thereby learned. Presumably Fodor would say something similar about bootstrapping insofar as it too leans on internal computational constraints to explain how concepts are acquired.

Yet Fodor's objection here threatens to beg the question against Margolis. Fodor challenged learning theorists to come up with an alternative to the building blocks model; Margolis articulated an alternative; and now Fodor insists, rather flat-footedly, that the alternative doesn't count as genuine learning, only mere acquisition. The obvious question is *why* it shouldn't count. What does genuine learning require that this alternative lacks?

As Fodor often notes, not all cases of concept acquisition count as concept learning. Perhaps Newton acquired the concept GRAVITY because an apple fell on his head, causing his neurons to jiggle in just the right way; and perhaps some futuristic neurosurgeon could implant new concepts in your brain. Neither process would count as concept learning because (Fodor hypothesizes) concept learning is necessarily a *rational* process, which requires that it be *intentional*; its inputs must bear a sensible semantic relation to its outputs. Fodor's ultimate worry about Margolis's proposal, I take it, is that it fails to meet this condition. It's thus more like a bump on the head or futuristic neurosurgery than genuine learning.

Margolis and Laurence (2011, p. 529) grant that the inputs to a learning process need to be semantically related to its outputs. Carey would likely embrace this condition as well given her characterization of "learning processes as those that build representations of the world on the basis of computations on input that is itself representational" (p. 12). So it seems to be common ground that Fodor's condition on learning is well motivated; the inputs to a learning process must bear a sensible semantic relation to its outputs. But what Fodor seems to miss is that this condition is compatible with the learning mechanisms that Margolis and Carey isolate. The reason is that the internal constraints themselves admit of intentional characterization in terms of implicit contents. The essentialist disposition implicitly codes the content *same natural kind as*, so when taken together with the explicit content of the kind syndrome, there is a perfectly intelligible intentional explanation of where the outputted concept comes from. Likewise, the constraints governing the indigo bunting's learning mechanism implicitly code the content *center of rotation*, so when taken together with the explicit content of the bunting's perceptual representations of the night sky, they lead in a semantically intelligible way to a representation of the North Star.

In reflecting on the idea that internal computational constraints can contribute to concept learning it is worth recalling, and dwelling upon, the analogy to vision science. Internal constraints, such as Margolis's essentialist disposition, promise to behave in many ways like constraints in vision science, such as the assumption that light comes from above. Just as the assumption that light comes from above has figured prominently in explanations of various visual phenomena, the essentialist disposition promises to figure prominently in explanations of various instances of concept learning. Just as the assumption that light comes from above is characterized intentionally, in terms of the directionality of light in the world, the essentialist disposition is characterized intentionally, in terms of the property of being a natural kind. Just as the assumption that light comes from above is implicitly coded as a constraint that governs transformations among explicit representations without itself being explicitly represented, the essentialist disposition is implicitly coded as a constraint that governs transformations from kind syndromes to natural kind concepts without itself being explicitly represented. Finally, there is a well-entrenched method

for discovering the computational constraints that govern both vision and concept learning: researchers create ecologically invalid conditions in the laboratory and see how the system breaks down.

Rey (2014, p. 122) agrees, "*pace* Fodor, that it is not *fortuitous* that the sustaining mechanism establishes a certain nomic dependence." In other words, Rey accepts that the operation of Margolis's natural kind mechanism can be characterized as rational and intentional. But Rey raises a different objection to Margolis's natural kind learning mechanism. He complains that Margolis fails to explain why the mechanism would lock onto a kind such as *zebra* rather than some deviant category such as "*plains zebra, Grévy's zebra, mountain zebra, African equids, horses, mammals, animals, living things, terrestrials*…" (emphasis in original). But here Rey is just raising the deviant-interpretation challenge again, which we've agreed to set aside. As we've already noted, the deviant-interpretation challenge is independent of the circularity challenge and afflicts all theories of learning, not just theories of concept learning.[12]

I conclude that internal computational constraints are like external computational constraints insofar as they can participate in learning processes that increase a thinker's total expressive power without circularity.

6. Radical bootstrapping

6.1. Radical bootstrapping in outline

Sometimes new concepts are the direct output of a dedicated mechanism (such as Margolis's natural kind learning mechanism) that was selected through evolution to produce concepts of a particular type. Carey calls this the "relatively easy route to new conceptual primitives" (2014, p. 142). By contrast, Carey calls bootstrapping the "relatively hard route to new conceptual primitives" (2014, p. 145). There is no dedicated bootstrapping mechanism that spits out concepts of a particular type. Rather, bootstrapped concepts are outputs of a more complex, three-step process:

(a) The construction of a set of placeholders.
(b) The use of computational constraints to partially interpret those placeholders.
(c) The completed interpretation of those placeholders through modeling processes.

Although commentators often ignore the middle step, we can now see that it is essential. Without step (b), it really would be a mystery how the placeholders could acquire contents that increase the thinker's total expressive power. Modeling processes are not sufficient on their own. With step (b), however, genuinely new explicit content can emerge. Computational constraints—whether external, internal, or both—control the way that placeholders are related to one another and to the thinker's other concepts, which in turn endows them with new conceptual roles and thus new contents.

6.2. Radical bootstrapping of integer concepts

Carey's most detailed discussion of bootstrapping concerns children's acquisition of integer concepts. To start with, children

---

[12] As Carey (2014, pp. 142–145) observes, it's not clear in any case that Margolis is indebted for an explanation of why the natural kind mechanism operates in the way that it does rather than in some other way. As a matter of contingent fact, it's dedicated to generating basic-level kinds (kinds at the grain of *zebra, tiger, dog,* etc.). It might have been governed by different constraints—just as the visual system might have been governed by the assumption that light comes from below. But that's just to say that the constraints are contingent.

memorize the count list, which serves as a meaningless placeholder system. Computational constraints, both internal and external, are then used to partially interpret those placeholders.

To see how internal computational constraints help to interpret the placeholders, consider Carey's summary of how enriched parallel individuation endows the first few words in the count list with meaning.

> The meaning of the word "one" could be subserved by a mental model of a set of a single individual {i}, along with a *procedure* that determines that the word "one" can be applied to any set that can be put in 1–1 correspondence with this model. Similarly, two is mapped onto a longterm memory model of a set of two individuals {j, k}, along with a *procedure* that determines that the word "two" can be applied to any set that can be put in 1–1 correspondence with this model. And so on for "three" and "four." (p. 476, emphasis added)

Here the words "one," "two," "three," and "four" acquire new conceptual roles by being associated with procedures that exist "only as constraints on computations" (p. 477). These procedures ensure that the child holds that there is "one" F just in case the Fs can be put in one-to-one correspondence with the member of the set {i} in long-term memory, that there are "two" Fs just in case the Fs can be put in one-to-one correspondence with the members of the set {j, k} in long-term memory, etc. In establishing these new conceptual roles, the child's object file systems are doing double duty: they are representing the objects that belong to two sets; and they are determining that the objects from those sets stand in one-to-one correspondence. While the object files explicitly represent objects, they only implicitly represent that the objects are in one-to-one correspondence. There is a procedure by which it is determined whether two sets stand in one-to-one correspondence. But there is no explicit representation ONE-TO-ONE CORRESPONDENCE in the system. Thus, the child does not compositionally define the concepts ONE, TWO, THREE and FOUR from the concepts object and ONE-TO-ONE CORRESPONDENCE. But we can nevertheless provide an intentional explanation of how the child imbues the first few number words with their contents by appealing, in part, to the content of the procedure implicitly coded in the object file system that places objects in one-to-one correspondence.

External constraints also contribute to the bootstrapping process for integer concepts. On the way to interpreting their count list, children learn to play a game that consists of pointing to each member of a collection as they rehearse the ordered count list, such that the last word uttered corresponds to the cardinality of the collection. Of course, the children do not conceptualize what they are doing as representing the cardinality of a collection. As far as they're concerned, they're just playing a game, akin to "eeny, meeny, miny, mo." But from our perspective as theorists we can see that they are endowing the words in their count list with new conceptual roles. Thus, "three" comes to have the conceptual role of being uttered in the presence of collections of three objects; "four" comes to have the conceptual role of being uttered in the presence of collections with four objects; etc.

At this point, the child has partially interpreted her placeholder terms by virtue of having learned two procedures—one that draws on enriched parallel individuation and a second that draws on the counting game. But this interpretation is only partial. There is nothing that integrates these two procedures, and so it remains an open possibility for the child that the terms "one," "two," "three," and "four" employed in each process are systematically ambiguous. There is also no appreciation that the words in the count list are ordered by the successor relation, such that the difference between the referents of any two adjacent count words is exactly one. The modeling processes of analogy and induction overcome these limitations and enrich the interpretation.

Using the procedural knowledge from their enriched parallel individuation system, the child notices that when a collection with "one" F is combined with another collection with "one" F, the result is a collection with "two" Fs; that when a collection with "two" Fs is combined with another collection with "one" F, the result is a collection with "three" Fs; and that when a collection with "three" Fs is combined with a collection of "one" F, the result is a collection with "four" Fs. The child then draws an analogy: just as the terms "one," "two," "three," "four" that are associated with enriched parallel individuation refer to quantities that are separated by the addition of "one" individual, perhaps those same terms when they are associated with the counting game also refer to quantities that are separated by the addition of "one" individual. Using induction, the child then extrapolates: perhaps *every* term in the count list refers to a quantity that is "one" more than the quantity designated by the preceding term. In this way, the modeling processes of analogy and induction integrate the separate pieces of procedural knowledge that had been associated separately with enriched parallel individuation and the counting game, engendering new conceptual roles for the count list symbols that embody an ordering by the successor relation.

Recall Rey's version of the circularity challenge from Section 2.1, which charges Carey with smuggling the concept SUCCESSOR in through the back door to support the crucial analogy and induction. The problem with Rey's objection should now be clear: it overlooks the contribution afforded by computational constraints. By the time the analogy and the induction occur, the placeholders have already been partially interpreted by way of those constraints. In particular, the child has already noticed that the quantities associated with "one" to "four" by enriched parallel individuation differ by "one" individual. The child thus has all the resources she needs to analogize that the same difference characterizes the phonetically identical terms in her memorized count list, and then to induce that it characterizes the remainder of the terms in the count list. Doing these things establishes a new conceptual role for the terms in the count list that embodies the successor relation without explicitly presupposing it or defining it compositionally from simpler concepts.

Of course, this explanation is not without its limits. One might still ask, as Rey and Rips et al. do, why children induce that the remainder of their count list is ordered by the successor relation rather than in some other way. But that's just to point out that Carey, like everyone else who appeals to induction as an element in an account of learning, has not supplied an answer to the deviant-interpretation challenge.

6.3. Radical bootstrapping beyond integer concepts

Although the bootstrapping of integer concepts is Carey's most detailed case study, she also discusses the bootstrapping of rational number concepts such as $^1/_3$ and of physical concepts such as MATTER. Again, Carey maintains that placeholders, computational constraints, and modeling processes play an essential role (p. 418).

To grasp rational number concepts, children must learn how fractions and decimals represent, that division (which relates the numerator and denominator of a fraction) is distinct from subtraction, that not all numbers are ordered by the successor relation (e.g., there is no successor to $^1/_3$), and that in between any two integers (e.g., 1 and 2) there are infinitely many rational numbers. But this conceptual transition is exceedingly difficult and takes children many years to complete. Thus, many 11- and 12-year-olds report that $^1/_4$ is bigger than $^1/_3$ because 4 is bigger than 3, that 1.032 is bigger than 1.22 because 32 is bigger than 22, that there

are no numbers (or only a few numbers) between 0 and 1, and that you soon get to 0 if you keep dividing by 2 (pp. 344–359).

To acquire physical concepts such as MATTER, WEIGHT, and DENSITY, children must learn to differentiate weight (an extensive magnitude) from density (an intensitve magnitude), air from nothingness, and materiality from physical reality. But for a long period of time children fail to systematically grasp such differentiations. Thus, they opine that a small piece of Styrofoam weighs zero grams; they fail to conserve matter and weight; they claim that dreams are made of air; they insist that shadows exist in the dark but we just can't see them; and they judge a large aluminum object to be made of heavier stuff than a small steel object (pp. 379–411).

Carey argues that these two transitions are facilitated by mutually supporting bootstrapping mechanisms. Each transition begins with placeholder structures. For the transition to rational number concepts, the notation of fractions and decimals serve as placeholders. For the transition to physical concepts, words such as "matter," "weight," and "density" serve as placeholders. Some relations among these placeholders and one's other concepts are learned directly, as when children are taught the equation D = M/V or are told, "The larger the denominator the smaller the fraction." These learned relations place external constraints on how the placeholders are ultimately interpreted. They forge conceptual roles for the placeholders that help to determine their meanings in much the same way that learning to play the counting game does for integer terms.[13]

Finally, children deploy modeling processes to complete the interpretation of their placeholders. For example, they draw analogies between their partially interpreted placeholder terms and other domains that they have already conceptualized. Thus, one curriculum intervention discussed by Carey that has been especially successful at helping children transition to an understanding of matter and density makes use of analogies to other extensive and intensive quantities that are better understood, such as dots per box and the sweetness of a solution. By analogizing to these better-understood domains, children are able to construct new conceptual roles for their placeholder terms, thereby endowing them with further content.

As with the case of integer concepts, there is no worry about circularity for these episodes of bootstrapping because the new concepts aren't defined from old ones along the lines of the building blocks model. Rather, a set of blank symbols (placeholders) are introduced that inherit new contents at least in part from computational constraints that help to determine their conceptual roles—their dispositions to causally relate to one another, to other symbols, and to the world.

7. Conclusion

Carey's account of bootstrapping confronts the circularity challenge: how can bootstrapping alter a thinker's expressive power without presupposing that the thinker had concepts with the new expressive power all along? This paper has explored two interpretations of Carey's account that promise to answer this challenge.

The first is Modest Bootstrapping, which throws in the towel on the possibility of explaining how thinkers could alter their total expressive power, but does seek to explain how they can alter their available expressive power. Towards this end, proponents of Modest Bootstrapping posit latent concepts from which newly available

concepts are composed and/or selected. The main problem with this approach is the dearth of empirical evidence that the posited latent concepts actually exist.

The second interpretation is Radical Bootstrapping, which tackles the circularity challenge head on by explaining how foreign concepts are made available through an extended, three-step process. First, a network of placeholders is constructed. Then computational constraints—both internal and external to the mind—circumscribe conceptual roles for those placeholder symbols, thereby molding their contents. Finally, modeling processes such as induction enable the learner to complete the interpretation of the placeholders.

At this point, the deviant-interpretation challenge emerges: why do children induce *this* way rather than *that* way? And I have not explained how Carey—or anyone else—can address it. To this extent, my defense of bootstrapping has been partial. When it comes to induction, Carey's account of bootstrapping is no better off—but also no worse off—than every other account of ampliative learning.

References

Block, N. (1986). Advertisement for a semantics for psychology. *Midwest Studies in Philosophy, X,* 615–678.
Carey, S. (2004). Bootstrapping and the origin of concepts. *Daedalus, 133,* 59–68.
Carey, S. (2009a). *The origin of concepts.* New York: Oxford University Press.
Carey, S. (2009b). Where our number concepts come from. *The Journal of Philosophy, 106,* 220.
Carey, S. (2011a). Précis of *The Origin of Concepts. Behavioral and Brain Sciences, 34,* 113–124.
Carey, S. (2011b). Author's response: Concept innateness, concept continuity, and bootstrapping. *Behavioral and Brain Sciences, 34,* 152–162.
Carey, S. (2014). On learning new primitives in the language of thought: Reply to Rey. *Mind & Language, 29,* 133–166.
Carey, S., & Bartlett, E. (1978). Acquiring a single new word. *Proceedings of the Stanford Child Language Conference, 15,* 17–29.
Davidson, K., Eng, K., & Barner, D. (2012). Does learning to count involve a semantic induction? *Cognition, 123,* 162–173.
Fodor, J. A. (1975). *The language of thought.* Cambridge: Harvard University Press.
Fodor, J. A. (1981). The present status of the innateness controversy. In *In his representations: Philosophical essays on the foundations of cognitive science* (pp. 257–316). Cambridge: MIT Press.
Fodor, J. A. (1990). *A theory of content and other essays.* Cambridge, MA: MIT Press.
Fodor, J. A. (1998). *Concepts: Where cognitive science went wrong.* Oxford: Oxford University Press.
Fodor, J. A. (2008). *LOT 2: The language of thought revisited.* New York: Oxford University Press.
Fodor, J.A. (2010). Woof, woof. *Times Literary Supplement,* October 8: 7–8.
Fodor, J. A. (1980). On the impossibility of acquiring "more powerful" structures: Fixation of belief and concept acquisition. In M. Piatelli-Palmerini (Ed.), *Language and learning* (pp. 142–162). Cambridge: Harvard University Press.
Gelman, R. (2011). The case for continuity. *Behavioral and Brain Sciences, 34,* 127–128.
Goodman, N. (1955). *Fact, fiction and forecast.* Cambridge, MA: Harvard University Press.
Gross, S., & Rey, G. (2012). Innateness. In E. Margolis & S. Stich (Eds.), *Oxford handbook on cognitive science* (pp. 318–360). Oxford: Oxford University Press.
Kripke, S. (1982). *Wittgenstein on rules and private language.* Cambridge, MA: Harvard University Press.
Laurence, S., & Margolis, E. (2002). Radical concept nativism. *Cognition, 86,* 25–55.
Lewis, D. (1970). How to define theoretical terms. *Journal of Philosophy, 67,* 427–446.

---

[13] Carey says less about how *internal* computational constraints might contribute to these episodes of bootstrapping. It may be that they aren't involved at all (there is no reason to think that every episode of bootstrapping requires contributions from both external and internal constraints). Or it may be that further research is needed to reveal how they contribute.

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

Margolis, E., & Laurence, S. (2008). How to learn the natural numbers: Inductive inference and the acquisition of number concepts. *Cognition, 106*, 924–939.

Margolis, E., & Laurence, S. (2011). Beyond the building blocks model. *Behavioral and Brain Sciences, 34*, 139–140.

Pepperberg, I. M., & Carey, S. (2012). Grey parrot number acquisition: The inference of cardinal value from ordinal position on the numeral list. *Cognition, 125*, 219–232.

Piantadosi, S., Tenenbaum, J., & Goodman, N. (2012). Bootstrapping in a language of thought: A formal model of numerical concept learning. *Cognition, 123*, 199–217.

Ramsey, F. P. (1931). In R. B. Braithwaite (Ed.), *The foundations of mathematics and other logical essays*. London: Routledge & Kegan Paul.

Rey, G. (2014). Innate *and* learned: Carey, mad dog nativism, and the poverty of stimuli and analogies (yet again). *Mind & Language, 29*, 109–132.

Rips, L. J., Asmuth, J., & Bloomfield, A. (2006). Giving the boot to the bootstrap: How not to learn the natural numbers. *Cognition, 101*, B51–B60.

Rips, L. J., Asmuth, J., & Bloomfield, A. (2008). Do children learn the integers by induction? *Cognition, 106*, 940–951.

Rips, L. J., Asmuth, J., & Bloomfield, A. (2013). Can statistical learning bootstrap the integers? *Cognition, 128*, 320–330.

Rips, L. J., & Hespos, S. J. (2011). Rebooting the bootstrap argument: Two puzzles for bootstrap theories of concept development. *Behavioral and Brain Sciences, 34*, 145–146.

Shea, N. (2009). New concepts can be learned. *Biology and Philosophy, 26*, 129–139.

Shea, N. (2011). Acquiring a new concept is not explicable-by-content. *Behavioral and Brain Sciences, 34*, 148–149.

Spelke, E. S. (2011). Quinean bootstrapping or Fodorian combination? Core and constructed knowledge of number. *Behavioral and Brain Sciences, 34*, 149–150.

Strevens, M. (2012). Theoretical terms without analytic truths. *Philosophical Studies, 160*, 167–190.

Weiskopf, D. (2008). The origins of concepts. *Philosophical Studies, 140*, 359–384.

Wittgenstein, L. (1953). *Philosophical investigations*. Oxford: Blackwell.