LANGUAGE LEARNING AND DEVELOPMENT
https://doi.org/10.1080/15475441.2018.1499517

Why We Should Abandon the Semantic Subset Principle

Julien Musolino[a], Kelsey Laity d'Agostino[b], and Steve Piantadosi[c]

[a]Department of Psychology and Center for Cognitive Science, Rutgers University, Piscataway, NJ, USA; [b]Department of Philosophy, Rutgers University, New Brunswick, USA; [c]Department of Brain and Cognitive Sciences, University of Rochester, Rochester, USA

ABSTRACT

In a recent article published in this journal, Moscati and Crain (M&C) showcase the explanatory power of a learnability constraint called the Semantic Subset Principle (SSP) (Crain et al. 1994). If correct, M&C's argument would represent a compelling demonstration of the operation of an innate, domain specific, learning principle. However, in trying to make the case for the SSP, M&C fail to clearly define their hypothesis, omit discussion of key facts, and contradict the theory itself. Moreover, their presentation does not address the core arguments and alternatives against their account available in the literature and misrepresents the position of their critics. Once these shortcomings are understood, a very different conclusion emerges: its historical significance notwithstanding, the SSP represents a flawed and obsolete account that should be abandoned. We discuss the implications of the demise of the SSP and describe more powerful and plausible alternatives that provide a better foundation for ongoing work in language acquisition.

Introduction

A common assumption within generative models of language acquisition is that learners are guided by a built-in constraint called the Subset Principle (SP) (Baker, 1979; Pinker, 1979; Dell, 1981; Berwick, 1985; Manzini & Wexler, 1987; among others). In essence, SP compels children to be conservative in the way they generalize, and ensures that they always select the smallest possible language. While much of the early work on SP focused on the acquisition of syntax, subsequent research on the acquisition of semantics led investigators to recognize that there may be semantic subset problems as well. In order to solve these problems, a semantic version of the subset principle was introduced, called the Semantic Subset Principle (SSP) (Crain, 2012; Crain, Ni, & Conway, 1994; Crain & Thornton, 1998).

The general idea underlying the logic of the SSP is that the sentences that learners encounter in their language can sometimes have more than one interpretation. Moreover, these interpretations can create subset/superset patterns, where one interpretation is true in a subset of the circumstances that make the other(s) true. If so, and unless children initially hypothesize that only the subset interpretation is possible, acquisition would fail because learners would never observe contradictory evidence and therefore, under an error-based learning scenario, they would never change their hypothesis to the correct one for their target language. Both this problem and the approach to solving it have roots in the assumption that language learning is driven by positive evidence, and that learners change their hypotheses about grammar or meaning only when faced with contradictory evidence. If true, the SSP would represent a compelling demonstration of the operation of an innate,

CONTACT Julien Musolino jumusolino@gmail.com Department of Psychology, Rutgers University, 152 Frelinghuysen Road, Piscataway, NJ 08854, USA

domain-specific principle guiding language development. This conclusion, in turn, would provide strong support for a certain class of nativist accounts of language acquisition (Crain, 2012).

Since it was introduced, the SSP has been applied to a broad range of phenomena, including the acquisition of the focus operator *only*, the acquisition of scope relations, polarity phenomena, and modals, to name just a few. Nevertheless, the SSP has not been immune to criticism. Roughly a decade after it was introduced, the SSP was systematically critiqued in two independent articles reaching similar conclusions from different but converging lines of reasoning (Gualmini & Schwarz, 2009; Musolino, 2006). In their reply, proponents of the SSP argued that their critics were misguided, and they offered a new set of empirical findings as well as a review of the literature that they argued vindicates the SSP (Crain, 2012; Moscati & Crain, 2014).

In this article, we show that these recent attempts to defend the SSP fail to deliver on their promise. More importantly, we demonstrate that the SSP cannot be salvaged because, its historical significance notwithstanding, it represents an obsolete scientific theory which should be abandoned by practitioners in the field of language acquisition and must be superseded by alternatives, some of which are already available.

In order to reach this general conclusion, we begin by showing that while only one fully explicit definition of the SSP can be found in the literature, the principle has been implemented in at least five different ways. The SSP therefore lacks the precision excepted of scientific hypotheses and, as a consequence, leads to a proliferation of mutually incompatible interpretations (Section The many faces of the SSP). We then show that the SSP suffers from definitional, empirical, and conceptual problems. Specifically, we show that the only explicit definition of the SSP is fatally flawed because the principle is defined extensionally. Moreover, key predictions of the SSP have been empirically falsified. We then consider two central domains where subset problems arise, the acquisition of the lexicon and the acquisition of sentential meaning, and show that the SSP isn't conceptually necessary to solve these problems (Section Evaluating the evidence).

In section 4, we consider the most recent attempts to defend the SSP found in the literature (Crain, 2012; Moscati & Crain, 2014). We show that these attempts fail because they are self-contradictory, they misrepresent the criticism that has been leveled against the SSP, and they do not address the real definitional, empirical, and conceptual problems faced by the SSP. Finally, we consider some of the implications of the demise of the SSP for theories of the acquisition of semantic knowledge.

The many faces of the SSP

Before introducing the SSP, let us consider the specific rationale invoked by Crain and Thornton (1998) to justify postulating such a principle. These authors observe the following.

> Sometimes, more than one interpretation of a sentence is made available by Universal Grammar. To further complicate matters, these alternatives may form a subset/superset relation; that is, the circumstances that make the sentence true on one interpretation may be a proper subset of the circumstances that make it true on another interpretation. In such cases, a semantic subset problem arises if the target language includes the subset reading, but not the superset reading. To avoid semantic subset problems, the interpretive options for sentences must be ordered in the LAD (Language Acquisition Device) by a principle instructing learners to initially choose the representation that is true in the smallest set of circumstances. This is called the Semantic Subset Principle (Crain, 1992, 1993; Crain et al., 1994; Crain&Philip, 1993). (pp. 117–118)

In order to avoid "semantic subset problems", Crain and Thornton (1998), based on previous work by Crain and collaborators (Crain et al., 1994), propose the Semantic Subset Principle that they define as follows.

> Suppose that the interpretative component of UG makes two interpretations, A and B, available for a sentence, S. If so, then see if S is true in a narrower range of circumstances on interpretation A than on interpretation B. If so, then A will be hypothesized before B in the course of language development (p. 118)

The first published application of the SSP involves children's interpretation of the focus operator *only* in ambiguous sentences like (1).

(1) The dinosaur is only painting a house.
a. The only thing the dinosaur is doing is painting a house.
b. The only thing the dinosaur is painting is a house.

On one interpretation, (1) can be paraphrased as meaning that the only thing that the dinosaur is doing is painting a house, (1a). Alternatively, (1) can also be paraphrased as meaning that the only thing that the dinosaur is painting is a house. Notice now that interpretations (1a) and (1b) fall into the kind of subset-superset relationship described by Crain and Thornton (1998) since (1a) entails (1b). To be sure, if it is true that the only thing that the dinosaur is doing is painting a house, (1a), then it must also necessarily be true that the only thing that the dinosaur is painting is a house, (1b). The reverse does not hold since the dinosaur could be flying a kite and painting a house, which would make (1b) true, but would falsify (1a). Thus, (1a) is true in a subset of the circumstances that make (1b) true.

According to Crain et al. (1994), the logic of the SSP therefore leads to the prediction that children should initially interpret sentences like (1) in English on reading (1a) only—thus unambiguously—and then later on in the course of language development add reading (1b). Crain et al. (1994) tested this prediction experimentally and found that some preschoolers do indeed appear to be restricted to interpretation (1a) of sentences like (1). Here then, the SSP is applied by appealing to ambiguity within a language; in this case English. Let us call this implementation 1.

In subsequent work, Musolino, Crain, and Thornton (2000) applied the logic of the SSP to sentences like (2).

(2) Every horse didn't jump over the fence.
a. None of the horses jumped over the fence.
b. Not all the horses jumped over the fence.

Here too, we have an ambiguous sentence whose readings create a subset-superset relationship. On reading (2a), the example in (2) can be paraphrased as meaning that none of the horses jumped over the fence while on reading (2b) it can be paraphrased as meaning that not all the horses jumped. As before, one of the two readings, namely (2a), asymmetrically entails the other, (2b). Indeed, if it is true that none of the horses jumped over the fence, it necessarily follows that not all the horses jumped; but not vice-versa.

Musolino et al. (2000) point out that in a language like Chinese, the sentence in (2) is not ambiguous and can only be interpreted on the "none" reading, as shown in (3).

(3) Mei-pi ma dou mei tiao-guo langan.
Every horse all not jump-over fence
'Every horse didn't jump over the fence'
a. None of the horses jumped over the fence.

The problem then, according to Musolino et al. (2000), is that if learners were to initially assume that sentences like (2) and (3) are ambiguous, but were to find themselves in a situation where the target language is Chinese, they would reach the wrong conclusion since sentences like (3) are not ambiguous in Chinese. To make things worse, children would face a potential learnability problem since they would need to jettison the (incorrect) "not all" reading of sentences like (3) in the course of language development. However, since they would always hear sentences like (3) in situations that verify the truth conditions of both the "none" and the "not all" reading, since "none" entails "not all", children would never have any reason to abandon the "not all" reading.

The way out of this potential learnability trap, according to Musolino et al. (2000), is for the SSP to guide children's initial conjectures and enjoin them to assume that all languages, including English, behave like Chinese with respect to sentences like (2) and (3). In other words, children would begin with the assumption that sentences like (2) and (3) can only receive one interpretation, namely the "none", or subset, reading. If it turns out that the target language is Chinese, then nothing further needs to happen. If, on the other hand, the target language is English, then sooner or later, children will encounter sentences like (2) uttered in a context in which not all the horses jumped over the fence. This, in turn, will falsify children's initial "none" interpretation and will compel them to add the "not all" reading to the set of possible interpretations for sentences like (2).

Musolino et al. (2000) tested the prediction just described and found that, as expected, English-speaking preschoolers appear to be restricted to the "none" reading of ambiguous sentences like (2) in English. Notice that in this case, the SSP is invoked by appeal to a cross-linguistic comparison between languages in which a particular sentence is either ambiguous (in this case English) or not (Chinese). This is different from the case of *only* where the sole consideration was whether the relevant sentences were ambiguous within a given language (in that case English). We call the between-language comparison implementation 2.

In another set of studies, Moscati and Crain (2014) apply the logic of the SSP to the case of negative sentences containing modal expressions (e.g., *might, can*). Using examples like (4) and (5), Moscati and Crain point out that in English the modals *might* and *can* interact differently with negation. In (4), *might* takes scope over negation to yield a reading that can be paraphrased as "it is possible that John will not come" (possible > not). By contrast, negation in (5) takes scope over the modal *can* to yield an interpretation that can be paraphrased as "it is not possible that John will come" (not > possible).

(4) John might not come (possible not)
(5) John cannot come (not possible)

In order to account for the difference in interpretation between (4) and (5), Moscati and Crain (2014) postulate the existence of a lexical parameter, R, which interacts with negation to yield different scope patterns. When the value of the parameter is set to –R, as in (4), the modal is interpreted in its surface position, yielding the possible > not scope pattern. However, when the value of the parameter is set to + R, as in (5), the modal is forced to undergo covert movement, yielding the not > possible scope relation.

Furthermore, Moscati and Crain point out that the interpretation of (4) and (5) create the usual subset-superset pattern. That is, if it is not possible for John to come, (5), then it necessarily follows that it is possible that John might not come, (4), but not vice-versa. In other words, (5) asymmetrically entails (4). Since (5), the subset reading, corresponds to the + R value of the parameter whereas (4) corresponds to the –R value, Moscati and Crain invoke the logic of the SSP to claim that children will initially assume that all modals receive the + R value of the parameter. The rationale offered by Moscati and Crain should by now be familiar. If children were to initially assume that all modals are associated with the –R value, they would incorrectly interpret (5) to mean (4). For adults, (5) would only be true in situations in which John doesn't come. For children, however, the same sentence would be true both in situations in which John doesn't come and in situations in which he comes thereby giving rise to the usual learnability trap.

On the other hand, if children were to initially assume that all modals take the value + R, they would interpret (4) to mean (5), namely that John cannot come. However, upon hearing (4) in a context in which John does come, thereby falsifying their initial conjecture, children would be able to learn, according to Moscati and Crain, that sentences like (4) are associated with the –R value of the parameter. Here then, we have an implementation of the SSP that does not rely on the presence of ambiguous sentences either within or across languages. Instead, we have a pair of unambiguous sentences, (4) and (5), whose individual interpretations, when compared, fall in a subset-superset

relationship. This relationship is mediated by a proposed lexical parameter which takes different values in the sentences under consideration. We call this implementation 3.

Instead of testing children's interpretation of sentences like (4) and (5) in English, Moscati and Crain (2014) turned to Italian to find empirical support for their SSP-based claim. Italian, unlike English, relies on word order to convey the differences in interpretation discussed above. When the modal *potere* precedes negation as in (6), we get the possible > not reading, and when it follows negation, as in (7), we get the not > possible reading.

(6) Gianni può non venire (possible not)
'Gianni might not come'
(7) Gianni non può venire (not possible)
'Gianni cannot come'

Thus, Italian differs from English in not exhibiting polarity restrictions on the modal *podere* and there is therefore no need to invoke parameter R to account for the facts in (6) and (7). Moscati and Crain (2014) acknowledge that much and yet, they argue that the logic of the SSP nevertheless applies to Italian: "Although modal expressions do not exhibit polarity restrictions for adult speakers of Italian, the SSP predicts that Italian-speaking children nevertheless initially assign polarity restrictions to modal expressions. In particular, the prediction is that children assign the value [+ R] to the modal expression *potere*" (p. 349). Here then, we have two unambiguous sentences whose individual interpretations create a subset-superset configuration and no lexical parameter mediating the relevant scope relations. This is implementation 4.

Work by Notley, Zhou, Jensen, and Crain (2012), Notley, Thornton, and Crain (2012), and Notley, Zhou, and Crain (2015) provides yet another implementation of the learnability considerations underlying the logic of the SSP. Notley et al. (2012) propose to soften the "more categorical terminology of the Semantic Subset Principle" and in doing so, they introduce The Semantic Subset Maxim (SSM). These authors explain that they "have chosen the word 'maxim' to replace principle' since what is represented is a default preference for the subset reading, not the absolute presence or absence of an interpretation" (p. 491). They then define the SSM as follows:

"Within any given language, a sentence S containing two logical operators will always have two available interpretations, A and B. The child's task is to determine which of these interpretations in preferred in the local language. Faced with this ambiguity, the Semantic Subset Maxim enjoins children to initially favour the scope relationship that makes the sentence true in the narrowest range of circumstances, the subset reading" (p. 490–91).

The role of the SSM, as Notley et al. explain, is to encourage "children to proceed in a certain way to allow them to converge on the correct adult preferences as rapidly and effortlessly as possible" (p. 491). With the introduction of the SSM then, Notley et al. retreat from the stronger claim that children initially lack one of the readings of scopally ambiguous sentences, as the SSP would predict, and propose instead that both readings are available but that children have different *preferences*

Table 1. The many faces of the SSP.

Implementation	Phenomenon	Application	Difference
Type 1- Crain et al. (1994)	Sentences containing *only* in English	Ambiguity within a single language	Children and adults have different grammars
Type 2 - Musolino et al. (2000)	Scope relations in English and Chinese	Ambiguity between two languages	Children and adults have different grammars
Type 3 - Moscati and Crain (2014)	Modal expressions in English	Unambiguous sentences within a language with polarity restrictions	Children and adults have different grammars
Type 4 - Moscati and Crain (2014)	Modal expressions in Italian	Unambiguous sentences within a language without polarity restrictions	Children and adults have different grammars
Type 5 – Notley et al. (2012)	Disjunction in English and Mandarin	Presence of two logical operators in a sentence	Children and adults have the same grammar but different preferences

compared to adults. We call this implementation 5. Table 1 summarizes the different implementations discussed in this section.

So far, we've observed that while only one explicit definition of the SSP can be found in the literature, the principle itself has been implemented in at least five different ways. In the next section, we explain why this is a major problem for the approach developed by Crain and his collaborators. In a nutshell, the only explicit definition of the SSP is a non-starter because it is too vague and riddled with contradictions. Consequently, the definition itself is only loosely related to the way the principle is actually applied, leading to the proliferation of interpretations and implementations that we described above. Moreover, these different implementations have different, and often contradictory empirical consequences, which means that virtually any set of data can be explained by the SSP. Indeed, these different implementations allow proponents of the SSP to choose, based on the circumstances, when the SSP applies and when it does not, allowing them to steer clear of potentially inconvenient results.

Evaluating the evidence

Definitional problems

In the previous section, we observed that while only one explicit definition of the SSP can be found in the literature, proponents of the principle have implemented it in as many as five different ways. Why should this be? To address this question, taking a closer look at the definition of the SSP, the one given by Crain and Thornton (1998), repeated below (see also Crain et al., 1994), can be informative.

> Suppose that the interpretative component of UG makes two interpretations, A and B, available for a sentence, S. If so, then see if S is true in a narrower range of circumstances on interpretation A than on interpretation B. If so, then A will be hypothesized before B in the course of language development (p. 118)

As already pointed out by Musolino (2006), this definition is problematic because the principle is defined extensionally, that is, in terms of sentences. In other words, "sentence S" can only be a sentence of a particular language, L, since Universal Grammar does not contain a repository of sentences. The definition then amounts to stating that in order to find out whether the SSP applies, a learner must *already* know that S is ambiguous in the language she is trying to learn, L, since ambiguity is one of the necessary preconditions for the application of the SSP. But if the learner already knows that S is ambiguous in L, then she has converged on the adult grammar and the learning is done. Having begun with the assumption that S is ambiguous, and therefore allows reading A and B, why would the learner then need to assume that S is *unambiguous*, and allows only the subset reading, A, to finally conclude, later in the course of language acquisition, that the sentence is ambiguous after all and allows both readings A and B?[1]

It should be plain to see, then, that Crain and Thornton's most explicit definition is a non-starter. A flawed definition, in turn, gives investigators latitude to interpret the principle which may explain why we can find as many as five different implementations in the literature, as discussed in the previous section. It should also be pointed out that these different implementations do not represent successive refinements of the principle whereby an earlier version is replaced by an improved one in light of new data or conceptual difficulties raised by other researchers. Indeed, when Moscati and Crain (2014) defend the SSP against its critics (Gualmini & Schwarz, 2009; Musolino, 2006), a point

[1]Notice that what we describe here is an explicit formulation of the logic underlying Crain et al.'s (1994) application of the SSP to sentences containing the focus operator *only*. These sentences are ambiguous *in English*, and that's the pre-condition for the application of the SSP. In other words, we have a learning principle that presupposes, as a prerequisite for its application, knowledge of the very facts that it is supposed to help children acquire. One could of course always object that this is not how the SSP should be interpreted, but if so, one would need to give a different explicit definition of the principle (for further discussion along those lines see Musolino, 2006). All we are saying here is that the explicit definition provided by Crain and Thornton (1998) is a non-starter.

to which we return later, they never offer a precise definition of the principle, and they use the term SSP to refer to different implementations of the idea, including the original version by Crain et al. (1994), their own version involving lexical parameters, as well as the Semantic Subset Maxim.

Likewise, even after the Semantic Subset Maxim was introduced, ostensibly as an improvement over the Semantic Subset Principle (Notley et al., 2012; Notley, Thornton, and Crain, 2012), the SSP continued to be used as an explanatory principle (Crain, 2012; Moscati & Crain, 2014). Even several years after the SSM made its appearance in the literature, Moscati and Crain (2014) made no mention of it in their article and instead defended the SSP. All this shows that the different implementations of the SSP do not represent successive improvements but rather that they simply co-exist in the literature allowing proponents to freely rely on different implementations as the circumstances warrant.

The latitude surrounding the implementation of the SSP, in turn, has another important consequence: it renders the principle unnecessarily vague and its application open to subjective interpretation (at least for someone intent on defending it).

To see this, consider the example in (8). How, according to the SSP, should young children interpret that sentence?

(8) Some girls didn't ride on the merry-go-round.[2]
a. There are some girls that didn't ride on the merry-go-round (some > not)
b. *It is not the case that some girls rode on the merry-go-round (not > some)

If we adopt the original proposal of Crain et al. (1994), what we called implementation 1, then the SSP simply doesn't apply because (8) isn't ambiguous. If, however, we use the implementation advocated by Crain (2012) and Moscati and Crain (2014), then the SSP does apply because of the polarity restriction on *some*. As these authors explain:

> To avoid subset problems, however, the SSP requires children to initially adopt the [–PPI] value of the lexical parameter, so the lexical item some is not raised but rather interpreted in its surface syntactic position. Because children assign the value [–PPI] to some, its interpretation with respect to negation is determined by surface word order. For children, then, when some appears in subject position, it is interpreted as taking scope over negation (some > not), whereas when some appears in object position, negation takes scope over some (not > some) (p. 27)

So, if we follow Moscati and Crain (2014)—what we called implementation 3—the SSP applies and predicts that children should initially assign sentence (8) the (some > not) interpretation. However, if we applied the Semantic Subset Maxim, implementation 5, we would get yet a different result. Recall that according to the SSM, "Within any given language, a sentence S containing two logical operators will *always have two available interpretations, A and B* [our emphasis]... Faced with this ambiguity, the Semantic Subset Maxim enjoins children to initially favour the scope relationship that makes the sentence true in the narrowest range of circumstances, the subset reading". If so, the SSM would predict that (8) is ambiguous, and that children should display a preference for the (not > some) reading, the subset one in this case.

In summary, as this simple example illustrates, depending on which implementation of the SSP one chooses to adopt, the principle can either (a) not apply at all, (b) apply and predict result A, or (c) apply and predict the opposite of A. At minimum then, proponents of the SSP owe their critics a precise definition of their principle as well as a clear and unambiguous algorithm for its application.

Empirical problems

Recall that regardless of the particular implementation that one chooses to adopt, the SSP predicts that children initially lack one of the readings available to adults, namely the superset reading. To the

[2]The standard interpretation of sentences like (8) is that the subject must take wide scope over negation (see Lasnik, 1972; Musolino, 1998; references cited therein).

extend that proponents of the SSP claim that a particular non-adult interpretive pattern reflects the operation of the SSP, a demonstration that children do not in fact lack knowledge of the superset reading would falsify such claims.

The first case where falsification of SSP-based claims has been documented involves sentences like (9).

(9) Every horse didn't jump over the fence.
  (a) None of the horses jumped over the fence (every > not)
  (b) Not every horses jumped over the fence (not > every)

The initial observation, due to Musolino (1998), is that preschoolers, unlike adults, appear to be restricted to the subset reading of sentences like (9), namely (9a). This, in turn, has been interpreted by Musolino et al. (2000) as reflecting the operation of the SSP (see section The many faces of the SSP). In other words, the claim is that children initially lack the not > every, superset reading, in accordance with the logic of the SSP. However, it was later demonstrated that preschoolers can indeed access the not > every, superset reading of sentences like (9) under certain contextual manipulations (Musolino, 2011; Musolino & Lidz, 2006).

Moscati and Crain (2014) acknowledge that much when they write that:

> There are findings from studies of children's interpretations of negative sentences with quantificational expressions that are potentially troublesome for the SSP. In particular, several studies have found that children who exhibit a clear preference for the subset reading of certain sentences will also accept these sentences in circumstances in which only the weaker, superset reading is true (Gualmini et al., 2008; Musolino & Lidz, 2006) (p. 371)

The way out of such problems, according to M&C, is to argue that "according to the version of the SSP we have adopted in the present paper, this can only happen if children have reset the lexical parameter to the superset value" (p. 371). The problem with this line of reasoning is that it is contradicted by what M&C tell us elsewhere in their analysis. In order to deflect a potential criticism raised by Musolino (2006), M&C argue that "in the case of universal quantifiers, there is no lexical parameter to be set". However, the example they cite from Musolino and Lidz (2006) does involve universal quantifiers! M&C's argument therefore falls apart and the experimental evidence presented by Musolino and Lidz (2006) does falsify earlier, SSP-based claims regarding children's interpretation of sentences like (9).

What's more, Conroy (2008) has shown that children's preferences for the subset, every > not reading of sentences like (9) can be induced in adults. In other words, adults can be "turned into children". Moreover, the same contextual manipulations that lead children to behave more like adults have the same effects on adults who have been "turned into children". Such findings torpedo any account of parameter setting or re-setting and the kind of learnability claims that underlie the logic of the SSP. Indeed, it would be strange to claim that the reason adults can be made to behave like children is that they have reverted to the child setting of the relevant parameter. Likewise, adults' experimentally induced preferences for the subset reading of sentences like (9) cannot have anything to do with "learnability" since adults, by definition, have converged on their target grammar.

Another kind of empirical problem faced by the SSP has to do with the learning scenario through which children are believed to reach adult-like knowledge of their language. On this scenario, children are initially restricted to the subset reading of (potentially) ambiguous sentences and then later on, add the superset reading upon exposure to sentences used by adults on that particular reading. To take a concrete example, the claim is that children are initially restricted to the "none" reading of sentences like (9). However, the "none" reading will be falsified when children hear sentences like (9) used on the "not all" reading. This, in turn, should compel children to change their grammar and add the superset reading.

This important prediction could easily be tested experimentally. As a first step, one would need to select a group of children who interpret sentences like (9) on the every > not, subset reading. The next step would be to expose these children to sentences like (9) used to describe situations in which the subset reading is false, but the superset reading is true. This, according to the logic of the SSP, should compel children to change their grammar. Moreover, the number of instances on which children hear sentences like (9) used to describe a situation in which the superset reading is true could be controlled so that one could determine the amount of evidence needed for the switch to take place. This could of course be repeated with sentences containing different quantificational expressions. Such evidence, in turn, would provide support for the kind of learning scenario predicted by the logic of the SSP. The fact that such evidence isn't available, combined with the other problems discussed in this article, further weakens the case for the SSP.

In fact, there are reasons to doubt the plausibility of the learning scenario just described. First, it has been pointed out that sentences like (9) are almost always used by adult speakers of English on the not > every, superset reading (see Musolino & Lidz, 2006 for relevant evidence). The same holds for sentences containing *some* in the scope of sentential negation such as (10).

(10) John didn't understand something.

The standard account is that *some* is a positive polarity item and that (10) must be interpreted as meaning that there is something such that John didn't understand it. In other words, *some* must be interpreted outside the scope of negation, some > not, which corresponds to the superset reading. This means that whenever adults use sentences like (10), they do so with the intent to convey the superset interpretation.

In light of such observations, it is puzzling indeed that children as old as five years of age in typical experiments designed to test the predictions of the SSP still haven't managed to converge on the adult grammar. If adults almost always use sentences like (9) or (10) on the superset reading and children must set only a single parameter to interpret these sentences, one would not expect them to take more than five years to learn the relevant facts. Unless, of course, the evidence is so rare that children do not reliably come across it within a five-year window; this in turn would predict large variability in acquisition ages due to the variability of waiting time distributions (see, e.g., the model in Mollica & Piantadosi, 2017). But, generally, if data is rare, one wonders how all children manage to successfully converge on their target language.

Conceptual problems

The flawed definition and inconsistent application of the SSP, as well as the contravening empirical evidence, pose significant problems for the principle's supporters. In his 2012 book, Crain acknowledges that when he writes:

> There is controversy surrounding the empirical adequacy of the Semantic Subset Principle. We think the controversy misses the mark. As pointed out by Williams (1983), one might just as well talk about the Semantic Subset Problem (p. 184).

Even if there are problems with the SSP as it stands, the idea goes, the principle is logically necessary for language learners. Given that children have to avoid the learnability pitfall that would arise from initially adopting superset interpretations, they must rely on something like the SSP to learn adult interpretations from only positive evidence. In this section, we challenge the need for the SSP on conceptual grounds. First, we point out that although there are lexical subset problems, children are able to master word learning without relying on the SSP. Second, when it comes to acquiring sentential meaning, learners can rely on multiple sources of evidence to solve potential subset problems, including indirect positive evidence and their knowledge of the compositional nature of semantics. Third, there now exist

fully implemented and provably learnable probabilistic models which can solve subset problems in any area of language acquisition and therefore render the SSP obsolete (Piantadosi, 2011).

There is one area of language acquisition where potential semantic subset problems are ubiquitous: the acquisition of the lexicon. Consider the problem that a learner faces when trying to acquire the meaning of the word *dog*. A Chihuahua is a dog, a dog is a mammal, a mammal is an animal, an animal is a living thing, and so on. Upon encountering the word *dog*, learners then face a semantic subset problem. If they were to assume that *dog* means *animal*, no amount of positive evidence would be sufficient to falsify that assumption. To be sure, every time an adult would say "Oh, look at the dog!", that sentence would be true on the child's interpretation as well, "Oh, look at the animal!". On the other hand, if children followed the SSP, and initially assumed, say, that *dog* means *Chihuahua*, then that conjecture would be easily falsified when the child hears an adult exclaim "Oh, look at the dog!" in the presence of a German Sheppard.

This puzzle hasn't escaped the attention of researchers working on the acquisition of the lexicon. Reviewing the problems faced by word learners, Gleitman (1990) points out that "words that stand in a subset relation pose another serious problem for an unaided observation-based learning procedure" (p. 18). However, in light of the extensive literature on word learning, Gleitman concludes that "the idea that the child always begins with the most inclusive interpretation consistent with the data is falsified by the empirical facts" (p. 18). More recently, Xu and Tenenbaum (2007) conducted a series of word learning experiments that falsify the predictions of the SSP (for a similar conclusion see Piantadosi, 2011). In other words, lexical acquisition might be replete with semantic subset problems, but the SSP isn't the right solution to such problems.

Compellingly, children's behavior in word learning, even when faced by a subset problem, can be well described with a Bayesian learning model. A tutorial paper on simplicity, Bayesian inference, and the size principle can be found in Perfors, Tenenbaum, Griffiths, and Xu (2011). There are two key components of this kind of model: a prior over word meanings specifying how likely any particular meaning is before learning data is observed. Typically, such priors can be viewed as formalizing a notion of simplicity: learners are assumed to prefer "simple" hypotheses and only move to more complex ones when the data justifies it, a cognitive form of Occam's razor. The second ingredient to Bayesian models is a likelihood function that says how likely the observed data would be if the hypothesis were true. A standard likelihood is to use the "size principle" (Tenenbaum, 1999) which says that the probability of observing a data point depends on the size of the hypothesis. For instance, if a learner sees a Dalmatian called a "dog", they would be able to compute that if "dog" meant Dalmatian, the probability of seeing a Dalmatian would be 1; if "dog" meant dog, the probability of seeing a Dalmatian would be something like 1/(#kinds of dogs); if "dog" meant animal, then the probability of seeing a Dalmatian would be something like 1/(#kinds of animals). Thus, learners in this kind of model have a soft preference for more specific hypotheses, but that preference is driven by rational statistical considerations of thinking about the probability of generating the data they observe. As Xu & Tenenbaum show, the gradience in children's inferences in learning are well described by this kind of model.

Note that although the size principle likelihood captures some of the same intuition as the SSP–namely a preference for narrower meanings–there are several key differences. First, the size principle is not a hard constraint like the SSP. Instead, it is one component of a Bayesian model and learners will be expected to trade that component off against the prior, which may push preferences in another direction. In other words, it is a weaker probabilistic tendency, rather than a strict rule guiding acquisition. Notably, because of this tradeoff, children will sometimes narrow inferences when presented with confirming evidence, showing that the error-driven nature of the SSP accounts is not empirically justified. For instance, a child who sees one instance of a Dalmatian called a "dax" may think Dalmatian means "dog" if a basic level category is given a higher prior. However, if they see three or four Dalmatians called a "dax", they will narrow their inferences and believe "dax" means Dalmatian. This shows that the predictions of Bayesian or size principle learners aren't so much about the timeline of acquisition (e.g., what is learned first) as they are about how data

interacts with learners' hypotheses. Learners, like good statisticians, change hypotheses when the data support it, not just when the data contradicts their current hypothesis. Second, the size principle is at least potentially domain general in that it occurs in nearly all Bayesian cognitive models in and outside of language. This means that the philosophy behind it is entirely different: rather than thinking about what innate principle could solve each specific linguistic problem, Bayesian theorists tend to consider the richness of human inference and learning, even outside of language, and seek mechanisms that can deal broadly with similar problems that crop up in many domains. Thirdly, the size principle is not a principle in the sense of most language acquisition. This is how the size principle came to be applied to word learning, generalization, and concept acquisition. More broadly, the size principle is a consequence of reasoning statistically—this in part justifies its domain generality. Any learner who thinks about the process that generates the data they see and reasons with well-formed probabilities will inevitably come to the conclusion that broader hypotheses assign the data less likelihood. Any hypothesis that predicts more different kinds of data *must* assign data a lower likelihood (and thus disfavor them), a form of Occam's razor that motivates even Bayesian philosophy of science (see, e.g., Godfrey-Smith, 2009). Thus, for word learning, not only is the SSP not necessary, but the subset problem can be solved by appealing to much more general, and softer, ideals of statistical inference.

In addition to learning what the words of their language mean, children end up being able to compute the meaning of a potentially unlimited number of sentences. According to proponents of the SSP, the acquisition of sentential meaning involves semantic subset problems too and so we need the SSP to solve these problems (see section 2). However, if the SSP isn't necessary for lexical acquisition, in spite of potential semantic subset problems, one might wonder why learners would need anything like it to acquire sentential meaning. After all, the meaning of sentences, including ambiguous ones, is completely predictable from the meaning of the words they contain and the way those words are arranged. That is, semantics is compositional. Once children have learned the meaning of the relevant words and the grammar of their language, compositionality gives them the meaning of any novel sentence for free (abstracting away from idiomatic expressions).

One might object that having to learn the grammar of one's language creates another source of semantic subset problems, compositionality notwithstanding. However, even if grammatical acquisition involves having to solve potential semantic subset problems, learners have other tools at their disposal—including rational statistical inference—as in the case of lexical acquisition, thereby rendering the SSP conceptually unnecessary.

Another important tool that children have at their disposal to learn the meaning of the sentences of their language is indirect positive evidence. To see how this works, consider again the case of sentences like *Every horse didn't jump over the fence*, which are ambiguous, and their Chinese counterpart which can only receive a every > not reading. Musolino, Crain, and Thornton postulated the existence of a parameter to account for that difference between English and Chinese. In addition to being ad hoc, this solution is also unnecessary. As Musolino (2006) demonstrates, learners of English can deduce the ambiguity of sentences like *Every horse didn't jump over the fence* once they have learned other properties of English grammar for which there is ample positive evidence. Likewise, children learning Chinese can deduce that the same sentences are unambiguous in their language based on properties of the grammar of Chinese that they have to learn independently and for which there is also positive evidence (see Musolino, 2006 for details).

Reviewing the logic of subset problems, Piantadosi (2011) concludes that there are compelling problems with the Semantic Subset Principle. In the context of quantifier learning (e.g., "every" vs "some" above) Piantadosi proposes using the same tools as Xu and Tenenbaum (2007) in order to solve subset problems, applicable in principle to any domain of language acquisition. Piantadosi's suggestion is based on a fully implemented and provably learnable probabilistic model in which subset problems are solved by any learner who engages in sound statistical reasoning. Indeed, essentially the same model has been applied to explain concept learning in other—even novel—

domains (Piantadosi, Tenenbaum, & Goodman, 2016), supporting the domain-generality of statistical inference schemes that handle the subset problem.

In summary, the facts surrounding lexical acquisition, combined with the availability of indirect positive evidence, the compositional nature of semantics, and general probabilistic solutions to subset problems, powerfully negate the conceptual necessity and the empirical adequacy of the SSP.

Recent attempts to defend the SSP

Moscati and Crain (2014) offer a defense of the SSP in which they reaffirm the empirical adequacy and the conceptual necessity of the principle. In doing so, they focus on two empirical counterexamples raised by Musolino (2006), and claim that these two cases in fact pose no real threat to the SSP. M&C then proceed to compare the empirical coverage of the SSP to that of a competing account based on surface syntactic relations. They conclude that the extant literature vindicates the SSP in that it "constitute[s] *direct* counterevidence to an account based on surface syntactic relations, and as evidence supporting the SSP" (p. 373).

The counterexamples that M&C single out are given in (11) and (12).

(11) Some girls won't ride on the merry-go-round.

(12) The Smurf didn't buy every orange.

According to M&C, Musolino (2006) contends that a learner following the SSP should initially assign (11) the subset, not > some reading, and (12) the subset every > not reading, contrary to fact (Musolino, 1998, 2006, 2011). However, M&C disagree. For them, "this prediction, attributed to the SSP, does not follow from the formulation of the SSP presented in this article" (p. 27). The version of the SSP adopted by M&C involves lexical parameters. In the case of sentences involving *some* and negation, like (11), the claim is that children do not initially treat *some* as a positive polarity item (PPI) (see section 3.1). Applied to example (11), this means that children are expected to assign the some > not reading, and not the not > some reading, as Musolino (2006) claims. In the case of (12), M&C argue that there are no lexical parameters to be set when the universal quantifier (*every*) is involved, and so the SSP simply doesn't apply.

M&C's line of defense faces a number of serious problems. The first is that their description of the counterexamples in (11) and (12), ascribed to Musolino (2006), relies on a distorted version of the original argument. Indeed, Musolino (2006) begins by pointing out, as we have seen in section 2, that the definition of the SSP is a non-starter because the principle is defined extensionally. He then asks what would happen if the SSP were re-defined intensionally, and offers two possible reformulations that preserve the logic of the SSP. The relevant one is given below.

Option 1: Suppose that UG allows two categories α and β to interact in such as way as to give rise to two interpretations, A and B when these categories are syntactically combined. Suppose further that A asymmetrically entails B. If so, then a learner should initially hypothesize that when syntactically combined, α and β yield only one interpretation, namely A.

Under this definition, the SSP indeed predicts that children should interpret sentences like (11) and (12) on the not > some and every > not readings, respectively, contrary to fact.

The argument presented by Musolino (2006) then has the following structure.

(a) The definition of the SSP is incoherent because it is defined extensionally.
(b) If we reformulate the SSP intensionally while preserving its underlying logic, we have two options.
(c) If we examine option 1 (or option 2), the SSP makes the wrong empirical predictions.

By contrast, the argument that M&C ascribe to Musolino (2006) looks like this:

(a) The SSP makes the wrong predictions for sentences like (11) and (12).

Needless to say, this is hardly a fair rendition of the original argument.

The second problem with M&C's defense of the SSP comes from the elbow room available to them in virtue of the five different implementations of the principle found in the literature. All M&C have to do in order to counter potential empirical objections is to select an implementation of the SSP that is compatible with the counterexamples presented by their critics. In the case at hand, that version is the one involving lexical parameters, what we called implementation 3. Again, it is important to emphasize that M&C did not select implementation 3 after having carefully reformulated the SSP or explained that all other implementations were early mistakes that they now disavow. Indeed, M&C do not even offer a definition of the SSP, and when they introduce the principle, they write about the SSP without further qualification. To quote M&C, "The proposal that children initially adopt the subset value of such parameters is called the Semantic Subset Principle (SSP), which is a variant of the Subset Principle introduced by Berwick (1985)." (p. 4). Never mind the fact that mention of parameters does not appear in the only explicit definition of the SSP available in the literature (see section 2).

The third problem is that in trying to defend the SSP against its critics, M&C end up contradicting themselves. Recall that the defense against examples like (12) invoked by M&C is that in the case of universal quantifiers, there are no polarity restrictions, and thus no lexical parameter to set. Therefore, the SSP simply doesn't apply in such cases. However, M&C use different standards in the case of modals. In their analysis, they describe how English *might* and *can* are subject to polarity restrictions but that this is not the case in Italian. And yet, they allow the SSP to apply to Italian modals, even though there are no polarity restrictions in that language (see section 2). Following this logic, what would prevent someone from claiming that even though universal quantifiers in English do not exhibit polarity restrictions, the SSP predicts that English-speaking children nevertheless initially assign universal quantifiers polarity restrictions? Again, notice that none of this follows from any definition of the SSP given by M&C. Instead, these pronouncements are unmotivated auxiliary hypotheses invoked solely to handle experimental results and possible counter-examples.

The fourth problem faced by M&C's defense of the SSP involves their claim that the available evidence regarding the way children interpret scope-bearing expressions and negation "constitute *direct* counterevidence to an account based on surface syntactic relations, and as evidence supporting the SSP" (p. 29). At issue here are predictions that M&C ascribe to "an account based on surface syntactic relations" for which they cite Musolino (2006) and Musolino and Lidz (2006). The argument is simple. If you believed that children assign scope on the basis of the surface syntactic position of the elements involved, you'd be wrong. This is because, M&C argue, there are far fewer examples of this kind compared to cases where children display a preference for strong readings, sometimes in direct contradiction to what one would expect based on surface syntax.

The problem here is that the account based on surface syntax that M&C ascribe to Musolino and his collaborators simply doesn't exit. Anticipating this misinterpretation of his ideas, Musolino (1998) included a section called "There is no Isomorphism principle" (p. 184) and another one called "Isomorphism is just an observation – not a principle" (p. 186). To quote Musolino (1998):

"My concern is the following: I can easily imagine someone reading my work and thinking: "I see, Musolino has this principle of isomorphism–not a new idea by the way–and he predicts that children will interpret QNPs where they occurs in the surface string". The trouble with this thinking is that it is exactly what I DO NOT claim or even think. First, there is no Isomorphism Principle, at least as far as I am concerned".

The idea should be clear enough. There are good reasons Musolino (1998) called his main empirical finding the *observation* of isomorphism and not the *principle* of isomorphism. Principles make predictions, observations don't. This crucial qualification was again reiterated in Musolino (2011). Thus, describing Musolino's work as "an account based on surface syntactic relations", and ascribing predictions to such an account, as M&C do, is a misrepresentation of the ideas developed by Musolino and his collaborators.

The fifth problem with M&C's defense is that they do not mention, let alone address, the key difficulties raised by Musolino (2006). Musolino's main objections to the SSP are that (a) the definition

of the principle is flawed, (b) under any coherent reformulation the SSP makes the wrong empirical predictions, and (c) the SSP isn't conceptually necessary because children can rely on indirect positive evidence as well as on the compositional nature of semantics to acquire sentential meaning.

The sixth problem with M&C's defense of the SSP involves their use of lexical parameters. The notion of a parameter has a long history in the generative tradition and some investigators have recently argued that it has become obsolete (Boeckx, 2014). Unfortunately, the way M&C use what they call "lexical parameters" doesn't make contact with that rich literature. Consequently, their use of the notion of a lexical parameter is unmotivated theoretically and highly idiosyncratic. Specifically, M&C seem to equate the notion of a lexical parameter with a simple difference between two languages. For example, modals in English work differently from modals in Italian, so they propose a parameter. But parameters are more than mere differences between languages (see Boeckx, 2014 for a review). Moreover, while M&C insist that their formulation of the SSP relies on lexical parameters, their notion of a lexical parameter seems to be only loosely related to the predictions of the SSP. As we discussed earlier, M&C explain that there is a lexical parameter in English that regulates the behavior of modals like *can* and *might* but that there is no such parameter in Italian. And yet, they claim that the SSP applies to Italian. In sum, M&C's use of lexical parameters does not follow from well-motivated linguistic analysis and bears the hallmarks of an ad-hoc stipulation designed to account for their experimental findings while insulating the SSP from criticism.

Concluding remarks

Our central conclusion in this article is that theories of the acquisition of semantic knowledge based on the SSP represent obsolete theories that must be abandoned. Indeed, a review of the relevant facts points to the conclusion that the SSP faces an insurmountable number of interrelated difficulties including definitional, empirical, and conceptual problems. In the remainder of this section, we offer some brief remarks about the implications of the demise of the SSP.

Scientific theories, whether successful or not, are at least in part intended to explain certain observations in their domain of application. In the case of the SSP, the central observation is that preschool children often display a preference for strong readings of (potentially) ambiguous sentences. We have shown that the SSP isn't the right account of such an observation. Thus, the account must be abandoned, but the observation remains. Absent the SSP, why should children display such preferences? Or perhaps better even, why do children, and sometimes adults, display a preference for strong readings? A related question concerns the uniformity of the phenomenon. To what extent do observationally related cases reflect the operation of a single causal mechanism? We leave these questions open for future research.

Another important implication of the demise of the SSP is that this conclusion weakens a certain class of nativist accounts of language acquisition (e.g., Crain, 2012). For proponents of such accounts, the fact that children's behavior is apparently constrained by the SSP leads to the conclusion that children's preferences for certain readings can be taken as evidence for the operation of innately specified learnability constraints. Remove the SSP from the equation, however, and the inference from behavior to underlying causal mechanism and thus to innateness is no longer valid.

This is of course not to say that the entire case for nativism, or logical nativism, collapses. All this means is that nativists have to be more nuanced and more careful in making the case for their position. Indeed, any nativist claim will have to be integrated with the most current and state of the art schemes for statistical learning. In this case, if there is a domain general mechanism of statistical inference that solves subset problems, that must be taken into account by nativist theorizing.

References

Baker, C. L. (1979). Syntactic theory and the projection problem. *Linguistic Inquiry*, *10*, 533–581.
Berwick, R. (1985). *The Acquisition of Syntactic Knowledge*. Cambridge, MA: MIT Press.
Boeckx, C. (2014). *What Principles and Parameters got wrong*. Oxford University Press: Oxford.

Conroy, A. (2008). *The Role of Verification Strategies in Semantic Ambiguity Resolution in Children and Adults*. (Unpublished Doctoral Dissertation). University of Maryland, College Park, MD.

Crain, S. (1992). The Semantic Subset Principle in the Acquisition of Quantification. Paper presented at the Workshop on the acquisition of WH-extraction and Related Work on the Acquisition of Quantification, University of Massachusetts, Amherst, MA.

Crain, S. (1993). Continuity and Modularity. In E. Clark (Ed.), Proceedings of the 25th Annual Stanford Child Language Research Forum, Chicago, IL: University of Chicago Press.

Crain, S. (2012). *The Emergence of Meaning*. New York, NY: Cambridge University Press.

Crain, S., & Philip, W. (1993). Global Semantic Dependencies in Child Language. Paper presented at the 16th GLOW colloquium, Lund, Sweden.

Crain, S., Ni, W., & Conway, L. (1994). Learning, parsing and modularity. In C. Clifton Jr., L. Frazier, & K. Rayner (Eds.), *Perspectives on Sentence Processing* (pp. 443–467). Hillsdale, NJ: Lawrence Erlbaum Associates.

Crain, S., & Thornton, R. (1998). *Investigations in Universal Grammar: A guide to Experiments on the Acquisition of Syntax and Semantics*. Cambridge, MA: MIT Press.

Dell, F. C. (1981). On the learnability of optional phonological rules. *Linguistic Inquiry*, *12*, 31–37.

Gleitman, L. (1990). The structural sources of verb meanings. *Language Acquisition*, *1*(1), 3–55. doi:10.1207/s15327817la0101_2

Godfrey-Smith, P. (2009). *Theory and reality: An introduction to the philosophy of science*. Chicago: University of Chicago Press.

Gualmini, A., & Schwarz, B. (2009). Solving learnability problems in the acquisition of semantics. *Journal of Semantics*, *26*(2), 185–215. doi:10.1093/jos/ffp002

Gualmini, A., Hulsey, S., Hacquard, V., & Fox, D. (2008). The question-answer requirement for scope assignment. *Natural Language Semantics*, 16, 205–237.

Lasnik, H. (1972) *Analyses of negation in English* (Doctoral dissertation). Cambridge, USA: MIT.

Manzini, M. R., & Wexler, K. (1987). Parameters, binding theory, and learnability. *Linguistic Inquiry*, *18*, 413–444.

Mollica, F., & Piantadosi, S. T. (2017). How Data Drive Early Word Learning: A Cross-Linguistic Waiting Time Analysis. In *Open Mind*. The MIT Press, Cambridge, USA.

Moscati, V., & Crain, S. (2014). When negation and epistemic modality combine: The role of information strength in child language. *Language Learning & Development*, *10*(4), 345–380. doi:10.1080/15475441.2014.880640

Musolino, J. (1998). *Universal Grammar and the Acquisition of Semantic Knowledge: An Experimental Investigation of Quantifier–Negation Interactions in English* (Unpublished doctoral dissertation). University of Maryland at College Park.

Musolino, J. (2006). On the semantics of the subset principle. *Language Learning & Development*, *2*(3), 195–218. doi:10.1207/s15473341lld0203_3

Musolino, J. (2011). Studying language acquisition through the prism of isomorphism. In J. De Villiers & T. Roeper (Eds.), *Handbook of Generative Approaches to Language Acquisition* (pp. 319–349). New York, NY: Springer.

Musolino, J., Crain, S., & Thornton, R. (2000). Navigating negative quantificational space. *Linguistics*, *38*(1), 1–32. doi:10.1515/ling.38.1.1

Musolino, J., & Lidz, J. (2006). Why children aren't universally successful with quantification. *Linguistics*, *44*(4), 817–852. doi:10.1515/LING.2006.026

Notley, A., Zhou, P., & Crain, S. (2015). Children's interpretation of conjunction in the scope of negation in English and Mandarin: New evidence for the semantic subset maxim. *Applied Psycholinguistics*, *37*(4), 867–900. doi:10.1017/S0142716415000296

Notley, A., Zhou, P., Jensen, B., & Crain, S. (2012). Children's interpretation of disjunction in the scope of 'before': A comparison of English and Mandarin. *Journal of Child Language*, *39*(3), 482–522. doi:10.1017/S0305000911000092

Perfors, A., Tenenbaum, J. B., Griffiths, T. L., & Xu, F. (2011). A tutorial introduction to Bayesian models of cognitive development. *Cognition*, *120*(3), 302–321. doi:10.1016/j.cognition.2010.11.015

Piantadosi, S. T. (2011). *Learning and the Language of Thought* (Unpublished doctoral dissertation). Massachusetts Institute of Technology, Cambridge, MA.

Piantadosi, S. T., Tenenbaum, J. B., & Goodman, N. D. (2016). The logical primitives of thought: Empirical foundations for compositional cognitive models. *Psychological review*, *123*(4), 392. doi:10.1037/a0039980

Pinker, S. (1979). Formal models of language learning. *Cognition*, *7*, 217–283. doi:10.1016/0010-0277(79)90001-5

Tenenbaum, J. (1999). *A Bayesian Framework for Concept Learning* (Unpublished doctoral dissertation). Massachusetts Institute of Technology, Cambridge, MA.

Williams, E. (1983). Semantic vs. syntactic categories. *Linguistics and Philosophy*, 6:3, 423–446.

Xu, F., & Tenenbaum, J. B. (2007). Word learning as Bayesian inference. *Psychological Review*, *114*(2), 245–272. doi:10.1037/0033-295X.114.2.245