No integration without structured representations

Reply to Pater (2018)

Iris Berent

Northeastern University

Gary Marcus

New-York University

Address for correspondence:

Iris Berent

Department of Psychology

Northeastern University

i.berent@neu.edu

125 Nightingale

360 Huntington Ave

Boston MA 02115

Phone (617) 373 4033

Fax: (617) 373-8714

ABSTRACT[1]

Pater's (2018) expansive review is a significant contribution towards bridging the disconnect of generative linguistics with connectionism, and as such, it is an important service to the field. But Pater's efforts for inclusion and reconciliation obscure crucial substantive disagreements on foundational matters. Most connectionist models are antithetical to the algebraic hypothesis that has guided generative linguistics from its inception. They eschew the notions that mental representations have formal constituent structure and that mental operations are structure-sensitive. These representational commitments critically limit the scope of learning and productivity in connectionist models. Moving forward, we see only two options: either those connectionist models are right, and generative linguistics must be radically revised, or they must be replaced by alternatives that are compatible with the algebraic hypothesis. There can be no integration without structured representations.

Key words: Algebraic rules, structured representation, connectionism, associationism, the computational theory of mind.

[1] This research was supported by NSF grants 1528411 and 1733984 (PI: IB)

1. INTRODUCTION. The rise of connectionism in the mid 1980's (Rumelhart et al. 1986) has sparked a debate that has been raging for three decades, continuing to this day (e.g., Fodor & Pylyshyn 1988; Pinker & Prince 1988; Elman et al. 1996; Marcus 2001; Berent 2013; Frank et al. 2013; Marcus 2018). It is difficult to understate the urgency of those exchanges; indeed, the stakes could not be higher. Connectionism has challenged the very foundation of cognitive science--what are mental representations, how do they support productivity, and how is knowledge acquired.

Language has been right at the heart of those discussions. Yet, surprisingly, this controversy has had only limited impact on linguistics. Connectionism has contributed to Optimality Theory (Prince & Smolensky 1993/2004) and inspired analogical models of language (Pierrehumbert 2001; Bybee & McClelland 2005). But many linguists rarely consider the connectionist debate and how it speaks to the fundamental theoretical tenets that shape their daily research practices.

Pater's (2018) expansive review is a significant step towards bridging the disconnect with connectionism, and as such, it is an important service to the field. Pater concludes his piece with a call for integration, collaboration, and fusion, and the appeasing tone is certainly welcome.

Still, in our view, Pater's efforts for inclusion and reconciliation may have gone TOO far, inasmuch as they have obscured crucial substantive disagreements. The gist of Pater's article seems to be that the disagreements between connectionism and generative linguistics are more apparent than real. Pater asserts that the two traditions emerging from the generative work of Chomsky (1957) and the neural network approach of Rosenblatt (1957) are distinct, and yet, when he is through, it's hard to see exactly where he thinks the difference lies. Evidence to the contrary, as in the work of Pinker and colleagues (Pinker & Prince 1988; Prasada & Pinker 1993; Kim et al. 1994; Marcus et al. 1995; Berent et al. 1999; Pinker 1999; Pinker & Ullman 2002) is brushed aside as one that is 'not inherent to a generative analysis' as if differences in opinion are matters of emphasis, not substance.

In our view, the differences are real, and cannot be effectively bridged, unless they are first understood and acknowledged. Of course, we certainly agree with Pater's view that both Chomsky and Rosenblatt are ultimately concerned with the same thing: how complex behavior emerges from bounded computational systems. But sharing that premise doesn't mean the two approaches are aligned in their details.

One way in which the two approaches apparently differ is a red herring; ostensibly, Rosenblatt's tradition is more concerned with the neural substrates of cognition. So-called neural networks are often described as being inspired by the brain, while Chomsky and researchers in his tradition often make no such claims (other than to note that all linguistic behavior originates in the brain). But here the difference really is more apparent than real; as Francis Crick pointed out long ago, most neural networks aren't all that neural (Crick 1989). There is a school of researchers that try to build models of neural responses that are deeply rooted in details about neurotransmitters and brain wiring, but the 'neural networks' that try to capture linguistic phenomena do nothing of the sort. Nobody has (yet) proposed a detailed brain simulation that bridges between, say, syntactic representations and the detailed mechanisms of synaptic potentiation. It may be an advance when that happens, but that's not what is on the table.

The real differences, we believe, lie elsewhere, in foundational matters related to the representational commitments made by the two approaches. Here, we articulate these notions and demonstrate their far-reaching consequences with respect to learning and generalization.

2. REPRESENTATIONAL COMMITMENTS. Researchers on both sides of the divide will agree that speakers readily generalize their knowledge to novel forms—I say *blix,* you pluralize it as *blixes;* I go *gaga*, tomorrow you'll invent *dada.* The two approaches differ sharply in how they account for such generalizations. It is worth exploring at some length how the two approaches account for language, given how much is at stake.

The ALGEBRAIC approach (Chomsky & Schützenberger 1963; Fodor & Pylyshyn 1988; Pinker & Prince 1988; Pinker 1991; Marcus 2001) attributes such generalizations to operation on abstract categories, such as 'Noun' and 'X' ("any syllable").

The associationist approach, realized in many neural networks, denies that such categories play a causal role in cognition, and that abstract operations over those categories (rules) play a role in language. Instead, the associationist approach asserts that learners induce ONLY associations between specific lexical instances and their features— those of *dada* and *blix*, for instance (Rumelhart & McClelland 1986; Plunkett & Juola 1999; Ramscar 2002; Bybee & McClelland 2005).

Each of these proposals assumes that regularities are partly learned from experience, and as Pater points out, each form of learning also commences with some innate endowment and generates some abstract structure. There are nonetheless substantive differences in what TYPE of structure is given, what is learned, and how far knowledge generalizes (discussed in detail in Marcus 2001).

1.1 THE ALGEBRAIC HYPOTHESIS. The algebraic account assumes that the CAPACITY to operate on abstract categories is available innately. This does not necessarily mean that SPECIFIC categories and principles (Noun, the head parameter) are innate. Rather, it means that learning mechanisms that operate algebraically, over abstract rules, are present in advance of learning. These algebraic mechanisms chiefly include the capacity to form EQUIVALENCE CLASSES—abstract categories that treat all their members alike—and to operate over such classes using variables.

In this view, algebraic mechanisms are critical for generalization. A noun is a noun is a noun, no matter whether it is familiar or novel (e.g. *dog* vs. *blix*), and indeed, even if some of its elements are entirely nonnative to the language (e.g. *Bach*). Because the regularities extracted by the learner concern such abstract equivalence classes, not lexical instances, they are bound to generalize ACROSS THE BOARD**,** to any potential member of the class.

Algebraic operations over variables also provide the means to combine these classes to form a hierarchical structure that maintains two lawful relations to its components. One relation is SYSTEMATICITY. If you know about *blue dogs*, and *daxes*, you can readily understand alternatives built on similar bedrock, as such *blue daxes*; likewise, knowledge about the well-formedness of *baba* and *dada* will readily allow you to form *gaga*. In each case, you extract a formal structure (Adjective+Noun; XX) which you can apply to novel instances. Furthermore, knowledge of these complex forms entails knowledge of their components. Knowledge of *blue dogs* entails knowledge of *dogs*, and if a language allows the geminate in *agga*, it must also allow the singleton *g* (Berent 2013). This relation between complex forms and their parts reflects COMPOSITIONALITY.

Across-the-board generalizations, systematicity, and compositionality are three of the main hallmarks of algebraic systems. In a seminal paper, Fodor and Pylyshyn (1988) show how these properties follow from the structure of mental representations and the operations that manipulate them.

First, representations are discrete symbols, inasmuch as they link form and meaning, akin to the Saussurean notions of a signifier and a signified—the meaning of DOG is “canine”, whereas the meaning of a phoneme (/g/) is the information it conveys within the phonological system. Second, Fodor and Pylyshyn contrast between atomic and complex representations. The concept of a *blue dog*, for instance, has complex meaning that is comprised of two semantic atoms—for *blue* and *dog*, respectively. Each such semantic value, in turn, is expressed by a signifier-- simple or complex. A geminate /gg/ is likewise semantically complex, distinct from the atomic /g/, akin to the relation between the complex plural *blixes* and the simplex base *blix*. Third, and crucially, the meaning of complex representations is lawfully linked to its syntactic form. So if we assume that *blue*

dog has a complex meaning, its form must be likewise complex, rather than atomic; and if the notion *blue* is atomic, its form must be atomic as well. The converse--complex meaning expressed by atomic form, or atomic form expressing complex meaning --are typically avoided. Finally, mental operations are STRUCTURE-SENSITIVE—they operate only on the form of representations and ignore their meaning.

In light of these assumptions about structure-sensitive operations, systematicity, compositionality, and unbounded generalizations follow automaticity. Knowledge of the semantically complex *blue dog* entails knowledge of the atomic *dog* (and *blue*) because the latter is literally part and parcel of the former. And because *brown dog* and *blue dog* have identical structure, and it is this structure that determines their semantic interpretation, knowledge of *brown dog* will automatically allow you to envision what the novel *blue dog* means. The same holds for *gaga* and *blixes*—each is a symbol with a complex meaning and a complex form (XX, and Noun+S_{plural}) respectively, and for this reason, knowledge of the complex form entails knowledge of its constituents. In fact, this consequence is *guaranteed*--it follows mechanically from the structure of the representations. In other words, the structure of representations plays a CAUSAL role in computations.

Linguists readily recognize many of these assumptions in their own work: the constituent structure of representation matters precisely because it is the putative cause of linguistic processes. And it is for this reason that linguists carefully attend to the formal structure of their accounts. Fodor and Pylyshyn articulate why structure is necessary: form ensures that semantic relations between mental representations are preserved by the brain--a physical machine. How the brain encodes form (or meaning, for that matter) is unknown (Gallistel 2017), but it is not unreasonable to assume that the brain represents formal structure (for specific proposals, Marcus 2001). And if these categories are open-ended, then this machinery also ensures productivity. So if linguistic operations are sensitive to the constituent structure of forms, then it is possible to envision how, in principle, the brain could give rise to linguistic productivity, systematicity, and compositionality. It is this innate capacity to exhibit structure-sensitive operations over equivalence classes that we refer to as ALGEBRAIC, following Marcus (2001). The view is distilled in 1.

(1) The algebraic hypothesis.

a. Structured representations.

(i) Categories (e.g., Noun) form equivalence classes, distinct from their members (e.g., *dog*)

(ii) Mental representations are symbols (either simple or complex).

(iii) The meaning of complex representations depends on the syntactic structure of their form and the meaning of their simple constituents.

b. Structure sensitive processes

(i) Mental processes manipulate the syntactic form of representation in a manner that is blind to their semantic content.

(ii) Mental processes operate on variables.

Notice that the notion of algebraic operations (or algebraic rules) is broader than the standard notion of 'rules' in linguistics. While linguists typically use 'rules' to refer to 'recipes' for mapping inputs (a head and a complement) onto outputs (an X-bar), algebraic rules also encompass structure sensitive constraints on outputs ('A projection has a head'). Both views commonly assume equivalence classes, structured representations, and structure-sensitive operations, as summarized in 0.

1.2. ASSOCIATIONISM. Associationism outright rejects each of the foundational assumptions in 1. For example, in Rumelhart and McClelland's past tense model, learning begins with two arrays of feature-triplets, one serving as input, and one serving as output, and a set of connections between the input and output layers. By design, as part of the challenge to classical approaches, there are no systematic links between the form of representations and their meaning: the form of *liked* (semantically complex) is no different from the form of *like* (simple) or the irregular *went*—so called because, in the algebraic account, the meaning of *went* is complex, but its form isn't. In the associationist hypothesis, distilled in 2, these representations do not differ in kind:

(2) The associationist hypothesis.

a. Mental operations consist of associations between inputs and outputs, induced by experience.

b. No abstract categories distinct from their instances.

c. No systematic links between the structure of mental representations and their meaning.

This is not to say that associationism singlehandedly rejects all forms of 'abstraction' and 'structure'. As Pater points out, the past tense model includes abstract features (not sensory impressions or motor commands), and the model also has some measure of structure (e.g. the triplet structure of its representation, and the learned associations between inputs and outputs). What is critically eliminated from this account is the systematic link between syntactic form and meaning along with structure sensitive operations.

Additionally, not all forms of connectionism subscribe to associationism, just like not all 'generativist' models are algebraic. Outside of neural networks, associationism has inspired linguists to explore other computational approaches that seek to induce knowledge of language by relying on minimal innate structure. For example, Hayes and Wilson's 2008 Maximum Entropy model induces phonological constraints from strings of feature matrices; there are otherwise no innately structured representations or operations over variables. But despite the elimination of algebraic mechanisms, these associationist networks (connectionist or otherwise) have been shown to learn and generalize.

How is this possible? How could minimalist representations give rise to such powerful learning outcomes? Rumelhart and McClelland believe that the answer lies in the richness of linguistic experience—a claim that deliberately challenges Chomsky's assertions regarding the poverty of the input. Indeed, Rumelhart and McClelland envision that their research program will ultimately eliminate any innate linguistic knowledge altogether.

> We chose the study of acquisition of past tense in part because the phenomenon of regularization is an example often cited in support of the view that children do

> respond according to general rules of language. Why otherwise, it is sometimes asked, should they generate forms that they have never heard? The answer we offer is that they do so because the past tenses of similar verbs they are learning show such a consistent pattern that the generalization from these similar verbs outweighs the relatively small amount of learning that has occurred on the irregular verb in question. We suspect that essentially similar ideas will prove useful in accounting for other aspects of language acquisition. We view this work on past-tense morphology as a step toward a revised understanding of language knowledge, language acquisition, and linguistic information processing in general. (Rumelhart & McClelland 1986, p. 267-8).

The apparent success of connectionist models should give linguists reasons to pause and ponder. If a model that eschews the algebraic machinery, that is standard to generative models, can learn and generalize, then perhaps there are no structured representation—syntactic constituents, syllables or morphemes, and no rules or constraints. And if such structural representations are eliminated from the initial state of learning, then learners obviously could not encode innate universal constraints on language structure either. Associationism would thus deny the learner the representational mechanisms necessary to represent Universal Grammar. On this view, the entire research program of generative linguistics is seriously off track.

3. IS ASSOCIATIONISM A MERE NOTATIONAL VARIANT OF ALGEBRAIC RULES? Although people have often imagined that the algebraic and associationist hypotheses are mutually incompatible, Pater seems to believe that the distance between the generative and connectionist traditions is not as large. Referring to Elman's associationist RNN model of syntax, Pater notes that:

> A hidden layer can form abstract representations of the data, and there are some hints in Elman's results that those representations may do the work of

> explicit categories and constituent structure, but much research remains to be done, even today, to determine the extent to which they can (p. 23).

The key to bridging the gulf separating the two traditions is presented by the promise of 'emergentism'. On this view, the initial state of learning does not encode structured representations and rules. Thus, as an account of the initial state of learning, this view sides with associationism, and sharply differs from the algebraic hypothesis. But this may not be the case for the final state. On Pater's formulation, as we understand it, algebraic mechanisms might spontaneously EMERGE.

It is for this reason that Pater presents the contrast between 'innatism' (in the algebraic hypothesis) and 'emergentism' (in 'associationism') as a false dichotomy. And if algebraic mechanisms can spontaneously arise, then the two hypotheses—algebraic and associationist—would be not only compatible and complementary; they would also be essentially isomorphic. Pater, then, would certainly be right to encourage the fusion of the two traditions. As an account of the final state, associationism would be merely a notational variant of algebraic rules.

But as we will show, the promise of 'emergentism' does not seem to materialize, and associationism does not beget rules. When one looks carefully at the nature of linguistic generalizations, it becomes apparent that associationist networks systematically fail to capture the empirical facts.

4. THE SCOPE OF LINGUISTIC GENERALIZATIONS. Generalization presents the quintessential test of learning, and it initially appeared that associationist models passed it with flying colors. When Rumelhart and McClelland first presented their model with *mate*--a regular verb that the model has not previously encountered--the model's most frequent response was *mated* – generalization without algebraic representation. And as Pater points out, subsequent models, with improved (more realistic) phonological representations produced even better outcomes. These results would seem to suggest that an algebraic machinery 'emerges' during the learning process. But a closer inspection suggests that this conclusion was premature.

The hallmark of algebraic rules is not simply the capacity to generalize. Rules generalize ACROSS THE BOARD. They can extend generalizations to any member of a category, irrespective of its similarity to training items, and they obey systematicity and compositionality. For example, a model trained on the English past tense should be able to generalize regular inflection not only to *jake* (similar to the regular verbs *bake, fake*) but also to [x]*ake* (with a nonnative English phoneme)—an exemplar that is dissimilar to English verbs. Similarly, a reduplication model trained on [ba] ([ba]→[baba]) will generalize [xa] to [xaxa].

Why are algebraic rules so powerful? The reason is simple, and, as noted earlier, it follows directly from the representational commitments of algebraic models. Because algebraic representations are systematically structured (e.g., *baked* and [x]*aked* share the same syntactic form, Verb+suffix) and compositional (the –*ed* suffix makes the same contribution to *baked* and [x]*aked*), and because mental operations are structure sensitive, generalizations depend ONLY on the structure of mental representations; they are literally blind to the idiosyncrasies of *bake* and [x]*ak*e. Across the board generalizations, then, are the inevitable reflex of algebraic machinery.

Generalizations, then, offer a concrete litmus test for computational properties of a model. If algebraic machinery could 'emerge' spontaneously in connectionist models, then such models should not merely generalize; they should generalize across the board, irrespective of whether test items are similar or dissimilar to training items, and these generalizations

should respect systematicity and compositionality. On the other hand, if these models track the statistical structure of specific instances (in line with associationism), then generalizations should depend on the similarity of test items to training items.

Before we proceed to evaluate this prediction, however, we need a more precise definition of 'similarity'. And indeed, what counts as 'similar' critically depends on the phonological representation employed by the model and the properties of the training and test items. To see this, compare the generalization of the reduplication function to two novel test items—[pa] and [xa] in two conditions. In both conditions, the model is trained on the same two items, [ba] and [ta]. But in one condition, these items are represented using segments, whereas in the other, the representation encodes features (for simplicity, we consider only a small subset of the consonantal features). The potential challenge to the learner in the two cases is vastly different.

(3) Generalization based on segmental representations.

		b	t	p	x
Train	[ba]	+			
	[ta]		+		
Test	[pa]			0	
	[xa]				0

When the representation is segmental, [pa] and [xa] are equally similar to training items (see 3); this is evident from the overlap between test items and training items (shared elements are indicated by a plus sign; elements that are not shared are marked by 0). The potential challenge to the learner changes drastically if the same items are encoded using features (see 4). Now, [pa] can be exhaustively described by features that have all been trained on, so this item is quite similar to the training items. In contrast, [xa] includes a feature that was never encountered during training, so this test item is far less similar to the training set. Marcus (1998; 2001) refers to the former test item ([pa]) as one that is situated within the training space, whereas the latter ([xa]) falls outside the training space.

(4) Generalizations based on featural representations.

		Labial	Voicing	Velar	Fricative
Train	[ba]	+			
	[ta]		+		
Test	[pa]	+	+		
	[xa]		+	0	0

For the algebraic hypothesis, the notion of the training space is irrelevant—generalization depends only on whether or not the test item belongs to the relevant class (X=syllable), and the answer, in both cases, is decidedly 'yes'. But if the model only extracts the statistical co-occurrence between the features encountered during training, then performance in the cases should differ. An associationist model should be able to generalize within the training space, but fail to extend generalization to test items that fall outside it. And these contrasting predictions allow us to determine whether algebraic rules can emerge in the course of learning.

Marcus (1998; 2001) systematically evaluated this question in various associationist connectionist models (feedforward network and simple recurrent network) using two distinct functions—reduplication and the past tense. Recent research by Loula, Baroni, and Lake (Lake & Baroni 2017; Loula et al. 2018) extended this investigation to explore the capacity of various recurrent connectionist networks to exhibit systematicity and compositionality. One set of simulations examined whether a network trained on 'jump twice' and 'sing twice' will systematically generalize to 'dax twice' (Lake & Baroni 2017). Another set of experiments examined whether knowledge of complex forms, such as 'jump around right' entails knowledge of its component 'jump right' (Loula et al. 2018).

The results across these distinct models and numerous case studies were quite clear. Associationist models were able to generalize within the training space, but consistently failed to systematically generalize to items that fall beyond it. For example, in the study of Loula and colleagues (Loula et al. 2018), a network trained on *jump around left*, *jump left* and *walk around right* readily generalized to *jump around right*. This is only expected, given

that all components of the test item (e.g. __*around right*) formed part of the training set. But when this specific bit of information was withheld (e.g., when trained on *jump left*, *jump around left*, *walk right*), generalization accuracy (to *jump around right*) dropped to 2.46%. It thus appears that these models have not induced an algebraic rule. When test items differ markedly from the training items, generalization fails.

As Pater correctly reminds us, connectionist networks are certainly able to implement algebraic mechanisms that are hardwired in the model 'innately', in advance of learning. For example, Smolensky (2006) has shown how the tensor product could be used to represent syllable structure in a connectionist model. Models equipped with operations over variables are demonstrably able to extend generalizations beyond the training space (Marcus 2001). On the other hand, associationist models that are not connectionist are not guaranteed to succeed. For example, the original Maxent model (Hayes & Wilson 2008) lacked the capacity to operate over variables, and for this reason, it failed to extend generalizations across the board. Once this capacity was added to the model, across-the-board generalizations followed (Berent et al. 2012).

Summarizing then, generalizations are not all alike. While test items that fall within the training space can be readily mastered by associationist models, generalizations outside the training space present a serious challenge for such models.

5. MOVING FORWARD. Pater's review calls for a fusion of generative linguistics with connectionism. He believes that the historic tensions between these two research traditions reflect mere differences in focus (on structured representations vs. learning, respectively), that the two perspectives are complementary, and that their integration could be fruitful.

In our view, these two approaches are largely antithetical. Most current connectionist models reject the fundamental representational commitments of generative linguistics. They eschew the notions that mental representations have formal constituent structure and that mental operations are structure-sensitive. These assumptions concerning the initial state shape the scope of learning. There is no hierarchical organization of sentences,

morphemes, or syllables; such formal constituents play no causal role in mental processes. Instead, learners only extract the statistical structure of the lexicon. Productivity, then, is limited to lexical analogies; no linguistic generalizations can extend across the board. It is difficult to see how such mutually exclusive perspectives could be integrated.

Moving forward, we see only two options: either associationism (in the strong sense of an alternative to algebraic rules) is right, and generative linguistics must be radically revised, or the strong associationism hypothesis must be replaced by a weaker version that is compatible with the algebraic hypothesis that has guided the generativist tradition. To adjudicate between these possibilities, there is a need for both computational and empirical research effort.

At the computational level, we need a more targeted investigation of generalization. Most researchers still evaluate their models by examining WHETHER they can generalize to new test items, rather than examining in detail which generalizations are and are not made. The results of Marcus (1998; 2001) and his followers (Berent et al. 2012; Lake & Baroni 2017; Loula et al. 2018) suggest that this is too coarse of a test. Generalizations falling within the training space are no guarantee that a model can freely generalize. So to evaluate the algebraic hypothesis, the SCOPE of generalizations is paramount, and so is the investigation of SYSTEMATICITY and COMPOSITIONALITY**.** It is only through such a targeted research program that one could determine whether algebraic machinery is emergent (as implied by Pater) or whether it must be hardwired in the model innately, in advance of learning (as suggested by Marcus).

Equally important is the evaluation of generalizations in human learners. Informed by their own intuitions, generativist linguists have assumed that humans can generalize freely, beyond the training space. But analytical judgments obtained leisurely, off-line, hardly demonstrate that people can extend such generalizations systematically in on-line language processing. While there is a handful of results that are consistent with this possibility (Berent et al. 2002a; Berent et al. 2014; Berent & Dupuis 2018), the scope of linguistic generalizations is rarely considered. This remains an urgent question for further empirical evaluation.

Before closing, we wish to briefly touch on innateness. Pater seems to brush the innateness question aside, suggesting that all models assume some measure of innateness, and in a sense, he is of course right. But this truism doesn't mean that the innateness question is inconsequential; what is innate matters. Associationist systems typically assume only some intrinsic phonological features along with machinery for analyzing correlation, and wind up unable to capture the richness of compositionality; generative approaches typically presume that, at the very minimum, the machinery of compositionality is innate, and seek to understand nuanced linguistic relationships as a function of such machinery.

Where the associationist approach has yielded relatively little in the way of specific characterizations of the sort of linguistic phenomena that are the bread and butter of generative linguistics, one might well wonder where the attraction to the more impoverished associationist view lies; in our view, it lies in an oft-held allergy to nativism. Researchers such as Elman (1996) Everett (2016) and Evans (2014) often suggest that innate ideas (of any kind) are biologically implausible, and so are innate linguistic primitives and constraints. Associationism eliminates the tensions surrounding innateness. If there are no rules, then there could be no innate universal rules either. And although de facto, connectionist networks encode linguistic knowledge, not brain activity, much of the excitement surrounding connectionism has to do with the hope of reducing the cognitive (mentalistic) level of explanation to the body—either the brain or sensory organs.

We do not believe that these concerns have any scientific merit. The notion of innate ideas is perfectly compatible with modern biology (Marcus 2004), and it is in line with the large literature on infant core cognition (Bloom 2004; Spelke & Kinzler 2007; Bloom 2013). In fact, a recent line of research suggests that the resistance to innate ideas could well be grounded in core cognition itself (Berent et al. 2018). To be clear, this does not show that scientists are biased, and it certainly doesn't demonstrate that language is innate. But these results do suggest that the promise of connectionism to minimize innate knowledge and ground it in the body resonates with common biases that lie deep within the human mind.

Finally, some words about integration. Our discussion so far has considered associationism—a view that, by definition, is incompatible with the algebraic hypothesis.

But a weaker claim that SOME linguistic generalizations are formed by associations could certainly live side by side with the algebraic view—this is precisely the approach presented by Pinker and colleagues.

In our view, this integration is not only possible but necessary. The large literature on statistical learning shows that humans (including young infants) can generalize by relying on mechanisms that are clearly NOT algebraic (MacDonald et al. 1994; Saffran et al. 1996). For example, people demonstrably generalize irregular inflection to novel forms (e.g. *bouse-bice*). As Pater notes, one could, of course, try to capture these generalizations by rules (Chomsky & Halle 1968; Yang 2002; Albright & Hayes 2003), but this move seems unmotivated. Irregular generalizations are exquisitely sensitive to similarity—the greater the phonological and semantic similarity to *mouse* the more likely people are to choose *bice* (Prasada & Pinker 1993; Berent et al. 2002b; Ramscar 2002). Such generalizations have all the hallmarks of an associative, rather than an algebraic, process; the distinct neural underpinnings are also in line with this view (Sahin et al. 2009).

A full account of linguistic productivity would likely require the synthesis of associative mechanisms along with algebraic rules. But this unification must maintain the representational commitments of the algebraic hypothesis that have guided the generative tradition from its inception. Without such structured representations as a bedrock, there can be no adequate integration.

Address for correspondence:

Iris Berent

Department of Psychology

Northeastern University

i.berent@neu.edu

125 Nightingale

360 Huntington Ave

Boston MA 02115

Phone (617) 373 4033

Fax: (617) 373-8714

6. REFERENCES

ALBRIGHT, ADAM & BRUCE HAYES. 2003. Rules vs. analogy in English past tenses: a computational/experimental study. *Cognition* 90.119-61.

BERENT, IRIS. 2013. *The phonological mind* Cambridge: Cambridge University Press.

BERENT, IRIS & AMANDA DUPUIS. 2018. The unbounded productivity of (sign) language. *The Mental Lexicon* 12.309-41.

BERENT, IRIS;AMANDA DUPUIS & DIANE BRENTARI. 2014. Phonological reduplication in sign language: Rules rule. *Frontiers in Language Sciences* 5.560.

BERENT, IRIS;GARY F. MARCUS;JOSEPH SHIMRON & ADAMANTIOS. I. GAFOS. 2002a. The scope of linguistic generalizations: evidence from Hebrew word formation. *Cognition* 83.113-39.

BERENT, IRIS;STEVE PINKER & JOSEPH SHIMRON. 2002b. The nature of Regularity and Irregularity: Evidence from Hebrew Nominal Inflection. *Journal of Psycholinguistic Research* 31.459-502.

BERENT, IRIS;STEVEN PINKER & JOSEPH SHIMRON. 1999. Default nominal inflection in Hebrew: Evidence for mental variables. *Cognition* 72.1-44.

BERENT, IRIS;GWENDOLYN M. SANDOBOE & MELANIE PLATT. 2018. How we reason about innateness. *Manuscript submitted for publication.*

BERENT, IRIS;COLIN WILSON;GARY MARCUS & DOUG BEMIS. 2012. On the role of variables in phonology: Remarks on Hayes and Wilson. *Linguistic Inquiry* 43.97–119.

BLOOM, PAUL. 2004. *Descartes' baby: how the science of child development explains what makes us human* New York: Basic Books.

BLOOM, PAUL. 2013. *Just babies : the origins of good and evil*: New York : Crown.

BYBEE, JOAN & JAMES L. MCCLELLAND. 2005. Alternatives to the combinatorial paradigm of linguistic theory based on domain general principles of human cognition. *Linguistic Review* 22.381-410.

CHOMSKY, NOAM. 1957. *Syntactic structures* Gravenhage: Mouton.

CHOMSKY, NOAM & MORRIS HALLE. 1968. *The sound pattern of English* New York: Harper & Row.

CHOMSKY, NOAM & M. P. SCHÜTZENBERGER. 1963. The Algebraic Theory of Context-Free Languages. Studies in Logic and the Foundations of Mathematics, ed. by P. Braffort & D. Hirschberg, 118-61: Elsevier.

CRICK, FRANCIS. 1989. The recent excitement about neural networks. *Nature* 337.129.

ELMAN, JEFFERY L.;ELIZABETH A. BATES;MARK H. JOHNSON;ANNETTE KARMILOFF-SMITH;DOMENICO PARISI & KIM PLUNKETT. 1996. *Rethinking Innateness: A connectionist perspective on development* Cambridge: MIT press.

EVANS, VYVYAN. 2014. *The language myth : why language is not an instinct*: Cambridge, United Kingdom.

EVERETT, DANIEL LEONARD. 2016. *Dark matter of the mind: the culturally articulated unconscious* Chicago: University of Chicago Press.

FODOR, JERRY & ZENON PYLYSHYN. 1988. Connectionism and cognitive architecture: A critical analysis. *Cognition* 28.3-71.

FRANK, ROBERT;DONALD MATHIS & WILLIAM BADECKER. 2013. The Acquisition of Anaphora by Simple Recurrent Networks. *Language Acquisition* 20.181-227.

GALLISTEL, CHARLES R. 2017. The Coding Question. *Trends in Cognitive Sciences* 21.498-508.

HAYES, BRUCE & COLIN WILSON. 2008. A maximum entropy model of phonotactics and phonotactic learning. *Linguistic Inquiry* 39.379-440.

KIM, JOHN J.;GARY F. MARCUS;STEVEN PINKER;MICHELLE HOLLANDER & MARIE COPPOLA. 1994. Sensitivity of children' s inflection to grammatical structure. *Journal of Child Language* 21.173-209.

LAKE, BRENDEN M. & MARCO BARONI. 2017. Generalization without systematicity: On the compositional skills of sequence-to-sequence recurrent networks.

LOULA, JOÃO;MARCO BARONI & BRENDEN M. LAKE. 2018. Rearranging the Familiar: Testing Compositional Generalization in Recurrent Networks.

MACDONALD, MARYELLEN C.;NEAL PEARLMUTTER & MARK S. SEIDENBERG. 1994. The lexical nature of syntactic ambiguity resolution. *Psychological Review* 101.676-703.

MARCUS, GARY. 2001. *The algebraic mind: Integrating connectionism and cognitive science* Cambridge: MIT press.

MARCUS, GARY. 2018. Deep Learning: A Critical Appraisal.

MARCUS, GARY F;URSULA BRINKMANN;HARALD CLAHSEN;RICHARD WIESE & STEVEN PINKER. 1995. German inflection: the exception that proves the rule. *Cognitive Psychology* 29.189-256.

MARCUS, GARY, F. 1998. Rethinking eliminative connectionism. *Cognitive Psychology* 37.243-82.

MARCUS, GARY F. 2004. *The birth of the mind : how a tiny number of genes creates the complexities of human thought* New York: Basic Books.

PATER, JOE. 2018. Generative linguistics and neural networks at 60: foundation, friction, and fusion. *Language*.

PIERREHUMBERT, JANET B. 2001. Stochastic phonology. *GLOT* 5-9.1-13.

PINKER, STEVEN. 1991. Rules of language. *Science* 253.530-35.

PINKER, STEVEN. 1999. *Words and rules: The ingredients of language* New York: Basic Books.

PINKER, STEVEN & ALLAN PRINCE. 1988. On language and connectionism: Analysis of a parallel distributed processing model of language acquisition. *Cognition* 28.73-193.

PINKER, STEVEN & MICHALE T. ULLMAN. 2002. The past and future of the past tense. *Trends in Cognitive Science* 6.456-63.

PLUNKETT, KIM & PATRICK JUOLA. 1999. A connectionist model of English past tense and plural morphology. *Cognitive Science* 23.463-90.

PRASADA, SANDEEP & STEVEN PINKER. 1993. Generalization of regular and irregular morphological patters. *Language and Cognitive Processes* 8.1-55.

PRINCE, ALAN & PAUL SMOLENSKY. 1993/2004. *Optimality theory: Constraint interaction in generative grammar* Malden, MA: Blackwell Publishing.

RAMSCAR, MICHAEL. 2002. The role of meaning in inflection: why the past tense does not require a rule. *Cognitive Psychology* 45.45-94.

ROSENBLATT, FRANK. 1957. The perceptron, a perceiving and recognizing automaton (Project PARA). Cornell Aeronautical Laboratory.

RUMELHART, DAVID, E. & JAY MCCLELLAND, L. 1986. On learning past tense of English verbs: Implicit rules or parallel distributed processing? Parallel distributed processing: Explorations in the microstructure of cognition, ed. by D. Rumelhart, E., J. McClelland, L & T.P.R. Group, 216-71. Cambridge MA: MIT Press.

RUMELHART, DAVID, E.;JAY MCCLELLAND, L & THE PDP RESEARCH GROUP (eds) 1986. *Parallel distributed processing: Explorations in the microstructure of cognition***2**). Cambridge MA: MIT Press.

SAFFRAN, JENNY, R;RICHARD N. ASLIN & ELISSA L. NEWPORT. 1996. Statistical learning by 8-month-old infants. *Science* 274.1926-298.

SAHIN, NED T.;STEVEN PINKER;SYDNEY S. CASH;DONALD SCHOMER & ERIC HALGREN. 2009. Sequential Processing of Lexical, Grammatical, and Phonological Information Within Broca's Area. *Science* 326.445-49.

SMOLENSKY, PAUL. 2006. Optimality in Phonology II: Harmonic completeness, local constraint conjunction, and feature domain markedness. The harmonic mind: From neural computation to Optimality-theoretic grammar, ed. by P. Smolensky & G. Legendre, 27-160. Cambridge, MA: MIT Press.

SPELKE, ELIZABETH S. & KATHERINE D. KINZLER. 2007. Core knowledge. *Developmental Science* 10.89-96.

YANG, CHARLES D. 2002. *Knowledge and learning in natural language* Oxford, UK: Oxford University Press.