Connecting Twenty-First Century Connectionism and Wittgenstein

Charles W. Lowney II[1] · Simon D. Levy[2] · William Meroney[3] · Ross W. Gayler[4]

Abstract

By pointing to deep philosophical confusions endemic to cognitive science, Wittgenstein might seem an enemy of computational approaches. We agree (with Mills 1993) that while Wittgenstein would reject the classicist's symbols and rules approach, his observations align well with connectionist or neural network approaches. While many connectionisms that dominated the later twentieth century could fall prey to criticisms of biological, pedagogical, and linguistic implausibility, current connectionist approaches can resolve those problems in a Wittgenstein-friendly manner. We (a) present the basics of a Vector Symbolic Architecture formalism, inspired by Smolensky (1990), and indicate how high-dimensional vectors can operate in a context-sensitive and object-independent manner in biologically plausible time scales, reflecting Wittgenstein's notions of language-games and family resemblance; we (b) show how "soft" symbols for such a formalism can be formed with plausible learning cycles using Sparse Distributed Memory, resolving disputes surrounding Wittgenstein's private language argument; and (c) show how connectionist networks can extrapolate meaningful patterns to solve problems, providing "ways to go on" without explicit rules, which indicates linguistic plausibility. Connectionism thus provides a systematicity and productivity that is more than a mere implementation of a classical approach, and provides Wittgenstein-friendly and Wittgenstein-illuminating models of mind and language for cognitive science.

Keywords Connectionism · Neural networks · Wittgenstein · Cognitive science · Philosophy of cognitive science · Philosophy of mind · Philosophy of language · Linguistics · Semantics · Computer modeling · Language games · Private language argument · Vector symbolic architectures · Smolensky · Emergentism · Reductionism

---

William Meroney is an independent philosopher and systems analyst at Meadows of Dan, Virginia, USA.
Ross W. Gayler is an independent cognitive scientist and applied statistics consultant at Melbourne, Australia.

✉ Charles W. Lowney, II
   lowneycw@hollins.edu

Extended author information available on the last page of the article

"All models are wrong, but some are useful." [1]

Burton Dreben, the Boston University and Harvard professor who was a moving force behind the New Wittgenstein interpretation of the *Tractatus* and the Emerson Hall reading of the so-called "private language argument,"[2] once said, with gruff rhetorical flourish, "Cognitive science is either *bad* philosophy or *bad* science."[3] Ludwig Wittgenstein's later work on mind and language has indeed cast much doubt upon cognitive science in general and computational approaches in particular. This doubt can be due to philosophical confusions endemic to the project, or it can be due to the functional limitations of a computational approach.

Dreben made his remark in the mid-90s, when the "classical" symbols and rules approaches of Noam Chomsky and Jerry Fodor dominated, and connectionist models did not seem to be viable alternatives for modeling the systematicity and productivity of thought and language (Fodor and Pylyshyn 1988; Pinker and Prince 1988; Fodor 1997). Since then, however, P.M.S. Hacker, a Wittgensteinian philosopher, teamed with Maxwell Bennett, a neurophysicist, to provide a role for philosophy in clearing away some of the misconceptions Wittgenstein would attack (Bennett and Hacker 2003; Bennett et al. 2009). Also since then, connectionism evolved to address many of its functional challenges (Calvo and Symons 2014),[4] perhaps, in the process, also clearing away what would be most inimical to Wittgenstein's understanding of language and thought in twentieth century cognitive science.

In part I of this paper, we demonstrate a general affinity between Paul Smolensky's (1987, 1988, 1990) distributed representation connectionist approach and Wittgenstein's ideas about symbol constitution, language-games, family resemblance, rule-following, logic, and language learning. At this point we rely mainly on Stephen Mills (1993) and we begin building analogies connecting connectionism and Wittgenstein. We then briefly describe how classicists countered connectionist claims, and *a fortiori* a Wittgensteinian approach to language in cognitive science, by pointing out some failures of 1980s connectionism to model important features of language and language learning. By the early 1990s it could indeed seem that connectionism was biologically, pedagogically and linguistically implausible, and what value it could have would merely be in providing insight on the implementation of a symbols and rules approach.

In part II, we proceed to more definitively overcome limitations set out in part I by presenting a family of approaches called Vector Symbolic Architectures (VSAs) that build on Smolensky's (1990) work and align with a dynamic systems approach.

---

[1] Box, G. E. P. 1979.

[2] New Wittgensteinians typically espouses the notion that inside the framing remarks all the propositions of the *Tractatus* are nonsense and there is nothing left even to *show* once the say-show distinction is surpassed and the ladder is thrown away. See, Cora Diamond (1995) and, Crary and Read, eds. (2002). The "Emerson Hall Reading" is Warren Goldfarb's verbal gloss what might also call a "third blush" reading (Goldfarb 1997), both are discussed below. See also footnote #20

[3] Our (Lowney) source here is a comment that Burton Dreben made in a class on Wittgenstein at BU. It was in direct response to the release of Steven Pinker's *The Language Instinct* in 1994. The gist was that, for Wittgenstein, the idea of a language of thought, and the idea of internal causes here, were not merely untenable, but were the result of philosophical confusions.

[4] New connectionist approaches are sometimes called "post-connectionist" or "non-classical connectionist." Some of these favor of a marriage with a dynamic systems approach (Calvo and Symons 2014).

Correlatively, we also more firmly establish connections between connectionism and Wittgenstein's descriptions of language acquisition and use. In (II.1) "Beetle in the Box" we introduce a VSA symbolism that can model language use in plausible time scales, and connect it with features of language-games; in (II.2) "Boxing the Beetle" we use a VSA-related approach called Sparse Distributed Memory to show how, from non-interpretable contextual iterations, symbols with the potential for systematicity and productivity can be constructed, and how this conforms to Wittgenstein's discussions of private language; and in (II.3) "Enter the Matrix" we use Raven's matrices to show how VSA connectionism can produce rule-like behavior without relying on the force of rules, which squares with Wittgenstein's notion of rule-following and logic. As we further strengthen the connectionist-Wittgenstein connection, we show how models in the VSA family effectively restore biological, pedagogical and linguistic plausibility without being a mere implementation of a classical approach. We thus display the promise of twenty-first century connectionism for cognitive science in a Wittgenstein-friendly and Wittgenstein-illuminating manner.

1 Part I: Early Connections between Connectionism and Wittgenstein

1.1 Connectionism *v.* a Classical Approach

The enthusiasm for the classical approach in cognitive science traces back to Chomsky's linguistics and the notion of generative grammar (Chomsky 1965). This approach emerged as a reaction against the behaviorism that began to dominate the fields of psychology and linguistics with B.F. Skinner. Chomsky showed that behaviorism and its associationist mechanics could not account for the productivity of language. His generative grammar hearkened back to a Fregean picture of a language of thought (LOT) that built new meaningful units up from symbols and basic logical-linguistic relationships (specified by rules). Current implementations of the classical architecture are thus said to employ "Fregean compositionality" (Jackendoff 2003). This approach advanced the aims of cognitive science, since, thanks to the ability to treat semantics as syntactically encoded (as in a formal logic), cognitive scientists could work to model language on standard digital computers.

In contrast to the classical approach, connectionism worked with associations and activation patterns based on parallel processing. This approach, popular at the time of Skinner, was overshadowed by the work of classicists such as Chomsky, Fodor and Stephen Pinker, but received renewed attention in the 1980s, especially after the technique of backpropagation was introduced (Calvo and Symons 2014). This technical innovation demonstrated the promise of connectionism for cognitive science by allowing connectionist networks to effectively learn regular and irregular verb patterns (see Rumelhart and McClelland 1986).

The resurgence of interest in connectionism was also due to the rise of neuroscience. Connectionism was a "neurologically inspired" architecture: connectionist nodes and networks were said to model neurons and the synapses created through the connection of axons and dendrites. Connectionism thus showed promise for unifying brain biology and perception with higher cognitive activities such as thought and language. It also went beyond the concatenation of signs used by the classical approach by using the

power of parallel distributed processing (PDP) in its computations, which seemed to better mimic brain architecture and speed. It thus seemed that a connectionist approach might eventually show us how thought and language actually worked. For connectionists, the higher level of description that employed linguistic categories and a Fregean compositionality (if it weren't eliminated as a remnant of folk psychology) would likely find its proper explanation in the architecture of networks and activation patterns (See Paul Smolensky's notion of Integrated Connectionist/Symbolic Systems or ICS, Smolensky 1988).[5]

1.2 Wittgenstein *v.* a Classical Approach

Several authors see connectionism as reflecting a new understanding of language and mind that emerged in Wittgenstein's later work (e.g., Stern 1991, Mills 1993, Dror & Dascal 1997, Goldstein and Slater 1998; Elman 2014). These authors also recognize the value of Wittgenstein's insights and so they see this confluence as support for the idea that connectionist approaches access something more basic to how we really think and use language.

One reason for noting the similarity between Wittgenstein and connectionism is the context for Wittgenstein's later insights: his comments often criticize or undermine the classicist notions of Gottlob Frege, Bertrand Russell and his earlier *Tractarian* self.[6] The classicist approach advocated by Fodor, Pylyshyn and Brian McLaughlin is shot through with assumptions presented by these early analytic philosophers regarding the connection of symbols/representations to objects/relations in the world, and regarding the existence of a logical or linguistic structure that accounts for meaningful composition from primitive symbols.[7]

In the classical analytic view, basic components received meaning from a direct connection to objects or relations considered to be primitive. A string of signs such as "aRb" could be a formalization of the English sentence "John loves Marie," which also may be parsed out symbolically by "LOVES (John, Marie)". The classical view also encouraged belief in the existence of mental states that could act on propositions and explain behavior; for example, "Sally believes that John loves Marie" or "BELIEVES [Sally, LOVES (John, Marie)]" and other such modern day renderings of Russell's notion of propositional attitudes. The later Wittgenstein criticized this classical picture of language and also criticized the correlative "language of thought" idea that reified linguistic usage into a propositional structure with variable constituents that could be filled in with names for objects and relations.

One of the ways that language could mislead us is by habituating us to believe that grammatical structure, and the entities it posits, were constitutive of meaning even in cases where they are initially absent. For Wittgenstein, just as "Slab!" is not an ellipsis of the sentence "Bring me a slab" that, in turn, carries the real meaning of the expression (*PI* #19), one need not go through the intermediary of rules, atomic

---

[5] Here already, in collapsing language to brain, a Wittgensteinian might sense something going awry: aren't we confusing causal processes together with grammatical relations?

[6] Wittgenstein might have been building up this classical picture in order to show how it falls apart when pressed, as Dreben and New Wittgensteinians argue.

[7] McLaughlin (2014) provides a good illustration of these classical assumptions by building to meaning/semantics in a Tractarian way.

symbols, and compositional grammar for the word to have its proper use or meaning. Likewise, we understand which piece in chess is the king not merely from pointing to a physical object but from what sort of moves the piece can make and how it relates to other pieces.

To summarize features of language for Wittgenstein, we note four main points. First, (a) signs require context to become meaningful, and they have soft rather than rigid definitional boundaries. Signs are mere marks or sounds that become symbols in the context of a language-game. Iterations of the sign in the right context allows for it to gain meaning. This language-game approach accounts for the family resemblance of a term in various uses. There need be no essential feature that is common throughout every meaningful use of a word. This moves against the idea of a primitive direct association of sign to referent, as well as against essentialisms, Fregean formalisms, and any simple psycho-linguistic parallelism. Second, (b) rule-like behavior emerges, but semantics is not necessarily governed by a rule. Language-games are prior and they are first learned by training (*abgerichtet*) or exposure to what we will call situated utterances of the sign or word. This primacy extends even to (c) logic and mathematics and their rules, and (d) the systematicity and productivity of language arrive without the prior need for either genetically preprogrammed or experientially internalized general syntactic rules.

1.3 Connectionism and Wittgenstein *v.* a Classical Approach

Smolensky's connectionism, which he dubbed PTC or the Proper Treatment of Connectionism (Smolensky 1988), was part of an Integrated Connectionist-Symbolic approach (ICS) which allowed for a higher-level description of symbolic processing in terms of the classical approach (what he called the functional or f-level), but whose real motors were in connectionist networks (the computational or algorithmic c-level).[8] For Smolensky, the higher level descriptions that the classists use are post hoc and rough approximations, reifying the results of connectionist processes in a useful way. There are several ways in which twentieth century connectionist approaches like Smolensky's are friendly to Wittgenstein.

For connectionists, signs can become symbols by virtue of their appearance in a variety of contexts. Smolensky's symbolic units, represented by vectors, are distributed representations as opposed to local. They are not one to one correspondences with items, i.e., objects or events that one might independently point to, such as the coffee in the cup or the sipping of a liquid, but instead they display an original context-dependence and can accommodate family resemblances to different uses of a term. As Mills explains, citing Wittgenstein, "The various uses of a word are unified, not by something they have in common, but by 'a complicated network of similarities overlapping and criss-crossing', similarities which are analogous to 'the various resemblances between members of a family'"(PI #67#68; Mills, 139).

Similarly, Smolensky, in discussing connectionist representations, talks about "a family of distributed activity patterns" in which the meaning of the term is not some feature common to all its uses. Smolensky even uses Wittgenstein's term "family

---

[8] There are effectively four levels to consider in Smolensky's ICS architecture that range from the higher cognitive processes that we are aware of to neural processes: 1) the functional level; the computational level, which is split into a higher (2) and lower (3); and (4) the neural level.

resemblance" to describe the distributed relation of different uses of "coffee" (Smolensky 1991, 176), and Mills, citing Smolensky's "coffee" example, strengthens the analogy between Smolensky and Wittgenstein's uses of "family resemblance" (140).

The symbol COFFEE is built up in a connectionist network in which the term "coffee," conceived as a physical mark or utterance, is distributed over various contextual iterations in which the sign appears. [Note that we will use all capitals, e.g. COFFEE, when we mention a symbol (a sign with representational meaning) to distinguish it from a mere sign, e.g. "coffee," as a string of marks or sounds, which we also call an "iteration," since it generally takes more than one instance of a sound/mark for it to become a sign, and hence be a viable candidate for becoming a symbol). Also, as in standard analytic practice, when we put a word in quotes, we will be referring to the word/sign/marks/sound rather than its meaning as a symbol.] In other words, the contextual iterations, with whatever sensory input the world provides, collect together from their various stored locations to produce the symbol as it appears in a meaningful use of language. The symbol, though it may develop some context-independence (as we describe below in "From Symbols to Systematic Language"; part II, section 2.3), carries its context with it internally; so, as Mills states for Smolensky, "One cannot then individuate a given representation of 'coffee' without making reference to the context in which the coffee occurs. Precisely the same is true for any of the various uses of a word on Wittgenstein's account" (Mills, 141).

Just as Wittgenstein showed how the various uses of the term in language-games—rather than a direct correspondence with an item—provided signs with their symbolic meanings, Smolensky and Mills show how the symbol gains its representational power from instances of use, such as "tree with coffee," "can with coffee," etc. (Smolensky 1988, 67; Smolensky 1991, 174–176; Mills 1993, 140). Hence the symbolic level that we can understand, is built up from the impression of iterations in a connectionist network that form a sub-symbolic level, whose associations are processed algorithmically.

The iterations, what we have called "situated utterances," are microfeatures that constitute a subsymbolic level (lower-c level, for Smolensky), and their interactive patterns of activation build to nodes at the symbolic level (higher-c level). Smolensky will represent the particular distributed instances as uninterpretable vectors, and represent the connectionist symbol as a vector that is an activation pattern or product of those sub-system vectors.

The development of a sign/symbol, which retains associations from experiences, forms a strong analogy between the families of distributed activity patterns that build a symbol from particular utterances and the family resemblance of various linguistic constructions that use the same sign. This is because constituents at the subsymbolic level can also be part of networks that activate different but related symbols (e.g., CAPPUCCINO). Also, the symbolic and sub-symbolic relation must be considered as relative, so that what are symbols in one activation pattern might be iterations or mere signs in another. In general connectionist terminology, one may say that weights established by use can activate nodes, and the resulting ensembles of node activations are interpreted as symbols.[9] Symbols can also combine to form meaningful

---

[9] It is important to note that this picture of weights and nodes does not entail a specific location in the brain correlating to a specific symbol or set of symbols.

propositions. As Smolensky states, "high level conceptual entities such as propositions are distributed over many nodes, and the same nodes simultaneously participate in the representation of many entities" (Smolensky 1991, 171).

As we see with family resemblance, distributed representations result in connectionist symbols with soft rather than hard boundaries (Mills, 141). Hence a connectionist approach does not suffer a breakdown when a familiar word is used in an unfamiliar way; it can work out a similarity based on a family resemblance to other uses and to proximally used words. So while in a classical symbols and rules approach a deviation in the meaning of a symbol might cause the breakdown of the program, connectionist architectures are not so brittle.[10] As Wittgenstein pointed out in analogy, the command to "stand roughly there" makes sense, though it lacks sharp borders (*PI* #71).

For connectionists, some form of training through iterations builds to language just as, for Wittgenstein, training and exposure build to language-games, and with training comes rule-like regularities. 1980s era connectionists arrived at a grammatical formulations by training networks. The groundbreaking work on this form of implicit rule generation is the formation of past-tense rules through exposure to linguistic use (Rumelhart and McClelland 1986). The main insight illustrated by this result is that correction of mistakes on individual words can, over many "training" iterations, lead to rule-like generalizations without recourse to explicit rules (such as "add the suffix 'ed' to create past tense") or lists of exceptions (change "go" to "went"). This training is important in obtaining the proper use of a term for connectionists, while explicit rules for word use are not. The key technical innovation enabling this result was the back-propagation of an error measure through intermediate ("hidden") layers of nodes. This technique of feeding information back as well as forward through a network to adjust the weights of various activations allows the network to learn.

While "training" for Wittgenstein covers more in the way of action and behavior than the training of a network, there is a similarity in the initial need for the rote repetition of iterations. Connectionist training is like the network experiencing the proper use of the words in particular language-games. Mills emphasizes this training and how it creates not only "soft" symbols but "soft" rules with fluid boundaries (Mills, 145). In connectionist networks, rules emerge that can change in response to experience. The defeasible and changing rules that Wittgenstein saw in language-games are thus more naturally modeled in connectionist networks. The graceful degradation in a network's ability to perform when faced with different or new usages mimics the actual behavior of human speakers. This shows how networks can mimic the gray area between, e.g., the use of "coffee" as a beverage and the use of "coffee" as a color, or how the term can fade to nonsense when the use is entirely inappropriate and out of confluence with all prior uses.

As we have begun to see, the parallels are so strong between connectionist approaches and Wittgenstein that Mills presents –and defends– a "complementarity thesis" whereby connectionism forms an *explanation* of the linguistic phenomena that Wittgenstein *describes*.

---

[10] The use of, e.g., statistical approaches has helped classicists become "softer" with regard to both symbols and rules, but this is somewhat artificial. For connectionists these features come "for free" (Smolensky 1988, 69).

1.4 Classicist Rebuttals to Twentieth Century Connectionism, and Mere Implementation

Advocates of the classical approach have pointed out a variety of difficulties with twentieth century connectionisms. As Fodor and Pylyshyn (1988) already began to note, connectionists could not easily replicate linguistic compositionality and productivity. Without something like a grammatical structure with placeholders for symbols, or rules for the productive recombination of symbols, it seemed that connectionist models would need some exposure to every single possible sentence, or might even need all possible new combinations of words neurally wired into the brain in advance. As long as connectionism had not dealt effectively with the creativity of language, it was *linguistically implausible*.

Also, while connectionism's use of backpropagation appeared to develop a level of linguistic competence, and so seemed to bypass the need for rules, these connectionist models did not seem scalable to real-world learning. Pinker and Prince (1988) pointed out that the cases of past-tense production were so information-saturated that statistical machines could produce the same regularities. Networks using backpropagation did not learn from a few positive examples, the way humans do (see Chomsky 1980, on the poverty of stimulus); instead, they required a large number of supervised training sessions to correct weights for the proper activation or inhibition of nodes. This hardly correlated with a child's ability to learn ten to fifteen new words a day. Connectionist programmers were thus building rule-like behavior into the networks in *pedagogically implausible* ways.

While Fodor and Pylyshyn sharply criticized connectionism for failing to model important systematic features of language, which the symbols and rules approach could capture directly, they still allowed that it might be used to model the biological or implementational level. Others went further in their criticisms, however, pointing out how, at second glance, connectionist processing fell short of what one would expect of a neurologically-inspired architecture, since neurons did not appear to display the sort of wiring that backpropagation would suggest.[11] Localist, rather than distributionist, models especially encountered a scalability problem, since it appeared they might require a number of "neurons" surpassing the number of particles in the universe (Stewart and Eliasmith 2012), which could further suggest *biological implausibility.*

While Smolensky's own approach in the 1990s did not rely on backpropagation and could begin to circumvent some of these critiques (as we shall see with VSA in part II), near the end of the twentieth century, classicists could, and did, renew arguments that connectionism was linguistically, pedagogically, and even biologically implausible.

Those classicists who did appreciate connectionism relegated its value to being a mode of implementation for a LOT approach, modeling the biological level. This raises philosophical as well as technical questions.

1.4.1 Implementation, Reduction, Elimination

The three plausibility arguments might be considered technical or functional problems in the ability to model properties that we associate with higher

---

[11] Even serious connectionists saw a problem here, causing some to retreat from claims to model neurophysiology and instead claim that connectionism modeled an algorithmic level (using Marr's distinctions) above biology but below the linguistic. See Bechtel and Abrahamsen (2002), 78–79; 342–344.

cognitive capacities, but the deeper philosophical conflict at stake concerned reductionism and eliminativism. Early classicists like Frege and Russell were rationalists who saw language and thought as independent of any biological structure that might implement them. Similarly, the functionalist approach adopted by the classicists (Fodor and Pylyshyn 1988) saw the classical structures as multiply realizable and thus as properly amenable to independent study. This, however, made it seem like the various ways in which the higher level of language and cognition was realized by a lower level were irrelevant, since classicists' linguistic level set all the desiderata that needed to be satisfied. Even if a connectionist approach might be improved to better model brain architecture this would be, for Fodor and Pylyshyn, a *mere* implementation. Connectionists, in contrast, saw themselves as dealing with linguistic processes in a more fine-grained way and from that perspective the coarse-grained approach of the classicists could be criticized as a reifying distortion.

This rationalist impetus may be why classicists criticized connectionists for not being able to account for the necessity of logic (Fodor and Pylyshyn 1988). Smolensky was able to mimic the behavior of the logical operators using connectionist architectures (Smolensky 1990), but while this might have been good enough for later Wittgenstein's normative conception of logic and mathematics (Hacker and Baker 2009) it did not have the sort of universality and necessity that the rationalist predilections of the classicists required.

We note that connectionism seems in confluence with Wittgenstein's insights into language and mind, but the classicists, at this point, seem to better reflect the anti-reductionist impetus of Wittgenstein's later thought. A successful connectionism that repudiated the classical understanding of language and propositional attitudes would appear to facilitate the elimination of "folk" conceptions born of ordinary use and facilitate the reduction of language and mind to brain (Ramsey et al. 1990). It seems a delicate line between a reduction, which (reduced up) would make connectionism appear an implementation or (reduced down) would make higher-level experience seem illusory, and an elimination, which would sacrifice entirely the viability of many of our ordinary descriptions of mental life (Churchland 1989). We believe that connectionists can avoid the philosophical pitfalls of reductivism and eliminativism through an emergentist approach. For example, there is an emergence from uninterpretable signs to symbols, and an emergence from context-dependent symbols to a context independence that allows for further higher-order interactions (as we shall see). For Wittgenstein, there is no reduction of the grammatical to the causal, and language works perfectly well just as it is. Taking those cues, the desiderata for a good connectionist emergentism are that it does not too easily slide into a reduction *up* to an implementation of symbols and rules, or *down* into lower level causal processes.[12] It should also forestall any full elimination of our ordinary language concepts. We shall not go much further into

---

[12] While Bechtel criticizes connectionist emergentists such as McClelland et al. (2010), the thrust of his criticism seems to be that they are not emergentist enough. He and co-authors seem to worry that without the emergence of "encapsulated representations" that operate with some context independence, a connectionist approach could not resist a reduction (down) (Borensztajn et al. 2014). See footnote #16 below and part II section 2.2.3 "From Symbols to Systematic Language" for a bit more on issues related to emergentism.

emergentism here, except to note that in an emergentist approach, higher levels properties must be considered epistemologically or ontologically insolvent. Here we primarily endeavor to show how connectionism can successfully respond to the rebuttals of the classicists, overcome its functional limitations, and coordinate even better with many of Wittgenstein's insights.

2 Part II: Twenty-First Century Connectionism: Modeling Wittgenstein's Insights

In response to classicist rebuttals, we now introduce a family of connectionist architectures, which, following Ross Gayler (2003), we call Vector Symbolic Architectures or VSAs (Kanerva 1994; Plate 2003; Rasmussen and Eliasmith 2011). VSAs follow the connectionist tradition of Smolensky into high-dimensional vector space. While other connectionist approaches have risen to meet the challenges of the classicists in various ways (Calvo and Symons 2014), we show how VSA and related approaches begin to resolve functional limitations of earlier connectionist efforts in a Wittgenstein-friendly, Wittgenstein-illuminating way.

2.1 Bracketing the Beetle: A Wittgenstein-Friendly VSA

Vector Symbolic Architectures describe a class of connectionist models that use high-dimensional vectors to encode systematic, compositional information as distributed representations. VSAs can represent complex entities such as multiple role/filler relations or attribute/value pairs in a way that every entity – no matter how simple or complex – is represented by a pattern of activation distributed over all the elements of the vector. This is because it does not assign different types of representations for components and for compositions.

In Levy et al. (2014), we provide a detailed computational description of how VSA can be used to build associations and answer queries about them. We set up a formalism with just three operations on vectors: an elementwise multiplication operation, $\otimes$, that associates or binds vectors of the same dimensionality (i.e., similar to a tensor product operator[13]); an elementwise vector-addition operation, +, that superposes such vectors or adds them to a set; and a permutation operator, $P()$, that can be used to encode precedence relations and other asymmetric relations like containment.

If the vector elements are taken from the set $\{-1,+1\}$, then each vector is its own binding inverse.[14] Binding and unbinding can therefore both be performed by the same operator, thanks to its associativity:

---

[13] In VSAs $\otimes$ is equivalent to a tensor product followed by projection onto a subspace of the same dimensionality as the operands.

[14] After adding vectors, the magnitudes of the elements may no longer be one. However, the information in the vectors is carried mainly by the signs of the elements, which collectively identify the hyperquadrant of the vector while the magnitudes of the elements locate the vector within the hyperquadrant. For typical VSA vectors, the number of hyperquadrants (i.e., the resolution due to considering only the signs of the elements) is vastly more than the number of protons in the observable universe. Memory constraints economize this by collapsing the vector to +1 or − 1 elements (this works like attractors in dynamic systems) and bring back scalability.

$$X \otimes (X \otimes Y) = (X \otimes X) \otimes Y = Y$$

Because these vector operations are also commutative and distribute over addition, another interesting property holds: the unbinding operation can be applied to a set of associations just as easily as it can to a single association:

$$Y \otimes (X \otimes Y + W \otimes Z) = Y \otimes X \otimes Y + Y \otimes W \otimes Z = X \otimes Y \otimes Y + Y \otimes W \otimes Z$$

$$= X + Y \otimes W \otimes Z$$

If the vector elements are chosen randomly, then we can rewrite this equation as

$$Y \otimes (X \otimes Y + W \otimes Z) = X + noise$$

where *noise* is a vector orthogonal (i.e., completely dissimilar)[15] to any of our original vectors W, X, Y, and Z. If we like, the noise can be removed through a "cleanup memory" that stores the original vectors in a neurally plausible way (Stewart et al. 2011). In a multiple-choice setting, the cleanup is not even necessary, because we can use the vector dot-product to find the answer having the highest similarity to (X+ *noise*).

The appeal of VSA lies in its flexibility for modeling all sorts of relations in a general way. We can even encode precedence and hierarchy relations by exploiting the aforementioned third simple operation, permutation P(), which allows us to express asymmetric relations like containment. Because $X \otimes P(Y) \neq Y \otimes P(X)$, we can represent the idea of X coming before (or containing) Y, but Y not containing or coming before X. So we can model "John loves Marie" as an asymmetrical relation by using permutation to order the relation and give it directionality.

Permutation also allows us to successfully search our network for structured information. For example, if we want to represent our knowledge of a beetle anatomy as a body (B) consisting of a head (H), thorax (T), and abdomen (A), with eyes (E) in the head and legs (L) on the thorax, we could use the following vector:

$$B \otimes P(H + T + A) + H \otimes P(E) + T \otimes P(L)$$

Now to answer the query *What is in the beetle's head?* we multiply this vector by the vector H and perform the inverse permutation:

$$P^{-1}(H \otimes (B \otimes P(H + T + A) + H \otimes P(E) + T \otimes P(L)))$$

$$= P^{-1}(H \otimes B \otimes P(H + T + A) + H \otimes H \otimes P(E) + H \otimes T \otimes P(L))$$

$$= P^{-1}(H \otimes B \otimes P(H + T + A) + P(E) + H \otimes T \otimes P(L))$$

$$= P^{-1}(P(E)) + P^{-1}(H \otimes B \otimes P(H + T + A) + H \otimes T \otimes P(L))$$

$$= E + P^{-1}(H \otimes B \otimes P(H + T + A) + H \otimes T \otimes P(L)) = E + noise$$

---

[15] (Dis)similarity can be quantified by a common function like dot product or vector cosine. The concept of noise requires a set of vectors of interest. Any vector in the subspace spanned by that set of vectors is signal and vectors in the complement subspace are noise.

VSA provides a principled connectionist alternative to classical symbolic systems (predicate calculus, graph theory) for encoding and manipulating a variety of useful structures. The biggest advantage of VSA representations over other connectionist approaches is that a single association (or set of associations) can be quickly recovered from a set (or larger set) of associations in a time that is independent of the number of associations. VSA thus answers the scalability problem raised by classicists with regard to biologically plausible real-time processing.

In a typical VSA model, each symbol is distributed numerically in a vector space of 10,000 (an arbitrary number that is a large enough upper limit, yet biologically plausible). By using random vectors of several thousand dimensions, we obtain a tremendous number of mutually orthogonal vectors (distinct signs/potential symbols) that are highly robust to distortion (Kanerva 2009). With these highly distributed representations there is no 'brittle-ness" that would cause the system to break down when the use of a term is different, e.g., in a different language-game. VSAs thus retain the soft symbols and family resemblances important to Wittgenstein. They also indicate the dispensability of grammatical rules. They build systematically with weights $\{+1, -1\}$ and sorts of association, without requiring either atomic symbols or the rules that classical approaches rely upon.

As Mills noted, for Wittgenstein there is not typically an atomic content or correspondence that one can point to in order to explicate the meaning of a term. A symbol is a product of "a complicated network of similarities overlapping and criss-crossing" (*PI* #66; Mills, 139). VSAs use multidimensional vectors and numerical weights, randomly assigned at the most basic level, in the actual processing of the networks constructed. There is no one-to-one correspondence to an entity or item for representation. A symbol is represented in signs/vectors that are distributed across a vector space, and operations with symbols, in turn, use these distributed representations to establish proximity relations that model thought and language use.[16]

Because VSA symbols are vectors that are constituted sub-symbolically with vectors, they do not function as classical symbols are thought to behave, for instance, VSA symbols can be highly context dependent. A VSA program can use weights established by the various iterations of the word in specific contexts to recognize a proper use and accommodate a new use.[17] The new use influences the very structure by which a word is assessed, as the new use becomes another iteration/sign in the distributed symbol.

---

[16] Symbols therefore are not tokened at the algorithmic level (as noted for Smolensky in McLaughlin 2014), so there is not an isomorphism between a higher cognitive and a lower algorithmic levels. This could interfere with Smolensky's attempt to integrate Connectionist and Symbolic architectures by making symbolic manipulations otiose and by denying the part-whole constituency relation that compositionality seems to require (Fodor and Pylyshyn 1988, 105). A full isomorphism, on the other hand, would present the opposite danger of a full reduction of the symbolic to the algorithmic level, in which case the connectionist structure could, in effect, be seen as a mere implementation (McLaughlin 2014). Of course, if the symbolic level is emergent upon the VSA operations, we have no need for the symbols to be directly tokened for them to be meaningfully dealt with at their own level of significance. This may be the thrust of Smolensky and McClelland's emergentist approach (McClelland et al. 2010) but the slide to (a) full reduction or (b) elimination may be made too easy in Smolensky's approach if (a) its symbols lack context independence (Borensztajn et al. 2014) or if (b) the symbolic level is completely disjunctive from the algorithmic processes (McLaughlin 2014). We will discuss some variations on this move to context independence in Part II section 2.2.3.

[17] To get an idea of how one can judge whether a word is used in one context rather than another (i.e., its family resemblance cousin that takes part in a different language-game), see how connectionist weights can force an outcome to one of two (or more) possible representations via the connectionist resolution of the duck-rabbit gestalt (see our Levy et al., 2014; and Rumelhart et al. 1986).

Together, the iterations encountered in social contexts provide us with common language-games about, e.g., beetles, but there need be no specific entity directly represented by a VSA symbol.

This approach conforms to what Wittgenstein says at *PI#293,*

> Suppose everyone had a box with something in it: we call it a "beetle". No one can look into anyone else's box, and everyone says he knows what a beetle is only by looking at *his* beetle.—Here it would be quite possible for everyone to have something different in his box. One might even imagine such a thing constantly changing.—But suppose the word "beetle" had a use in these people's language?—If so it would not be used as the name of a thing. The thing in the box has no place in the language-game at all; not even as a *something:* for the box might even be empty.—No, one can 'divide through' by the thing in the box; it cancels out, whatever it is.

VSA's use of random vectors virtually guarantees that "my beetle" will be different from "your beetle," thus mirroring the variations in the uses of the word that we each have encountered. Further, as the concept or symbol BEETLE evolves in the experience of an individual, or the word "beetle" changes its use by that individual in different contexts, the random vectors that constitute the symbol itself may be "constantly changing," but these differences, at what becomes the sub-symbolic level, do not interfere with the informative communication that can take place at the symbolic level due to the regularities in the use of the words.

One might see this regularity in the use of a word or a symbol as the functionalists do, i.e., that what is happening at the symbolic level is independent of the constitution that implements symbolic meaning, however, this would neglect the continued context-dependence of the symbol that connectionists retain. VSAs retain the distributed uninterpretable signs at the sub-symbolic level, i.e., the "micro-features," and this context-dependence provides for both the common symbolic meaning and individual nuances. Understanding a sentence or word is dependent on the games we have heard before that have included that word. The symbols depend primarily on situated utterances, and the "rules" associating symbols come from patterns of use.

To make sense, a symbol need not divorce entirely from games in which it has been used, as it would if it were entirely context-free, but, at the same time, those iterations are public and accessible, they are physical marks and sounds that we recognize and replicate. Communication is possible because the iterations and the proximity of the values of the symbols allow us to produce similar responses. So one can ask a question, such as "What is the color of your Beetle?" and the common language-games in which we participate will facilitate the response, "black," but how do I know my BLACK is the same as yours? BLACK would be equivalent to <black>, where the angle brackets < > around an item stand for the vector representation of that item, typically the vector sum (or superposition) of all of the individual instances of experience of that item (e.g., "coffee" in its uses) which contains all the vectors produced in the distributed micro-structure. Your BLACK need not be identical to mine, but the word "black" has been used in similar contexts and so it has acquired through training similar weights in my network as in yours (just as "beetle" or "coffee" have), i.e., there is a strong proximity

relation between it and "dark gray" for both you and me. The cumulative weights of the vectors create something like an attractor basin pulling us both towards common communal uses and the meaning that comes with them. This does not require that the internal composition of your network and mine be identical. The branches of Quine's elephantine bushes (Quine 1960, 8) (the bushes being <black> or the symbol BLACK in our example) might be trimmed with corrected use or by its failures/ successes in living up to predicted behavior in a given use, as in Andy Clark's predictive processing framework (Clark 2013).

While the VSA formalism described here does not show precisely how a symbol is constituted from the associations of contextual iterations, one can see how a Wittgensteinian notion of symbol-constitution is displayed at the functional-operational level of symbol manipulation. It shows that the VSA language can display and manipulate systematic relations with several simple connectionist operators, and this all takes place within an upper bound that provides biological plausibility. In the next section (2), we will show how symbols for this formalism may be built from limited and low-precision iterations[18]; this counters the classicist objection that connectionist training of networks requires an implausible number of training utterances to model human learning. Concurrently, this example will show in what sense there can indeed be a private language for Wittgenstein. Here our interpretation of Wittgenstein is similar to that of Hacker and Baker (2009, Hacker 1989 & Hacker 2001) and contrasts with that of P.F. Strawson (1954) or Norman Malcolm (1954).[19] Section three, building on Wittgenstein's work on rule-following, will provide an indication of how systematic and productive relations can develop, lending linguistic plausibility.

2.2 Boxing the Beetle: Soft Symbols from Sparse Data

In the *Philosophical Investigations* (1953), Wittgenstein discusses the possibility of a private language for an "inner" sensation (primarily in #256 & #258–260). One interpretation is that Wittgenstein here denies the possibility of a private language altogether and with it any meaningful discussion about private sensations. But if there can be a sort of private language that Wittgenstein is not denying, e.g., the sort that a Robinson Crusoe could invent and speak, even before the arrival of Friday, as Hacker and Baker (2009) surmise, then Wittgenstein's private language discussions present a riddle that connectionists can help solve.

2.2.1 Three Blushes

Although the meaning of these passages on private language is highly contested, a main point that Wittgenstein seeks to convey is that the Cartesian picture, whereby we can

---

[18] We do not claim to solve the connectionist symbol-grounding problem, but provide one indication of how that might proceed. Fruitful work on the grounding problem is underway. See, for example, Barsalou (1999) and Eliasmith (2013).

[19] Strawson thinks Wittgenstein believed that criteria for verification were required to make sense of language about private sensations, and —contra his understanding of Wittgenstein— Strawson believes such verification was possible. Malcolm believes that the payoff of the "private language" argument is the need for a community of speakers to convey sense. Hacker argues that a criterion is needed but it might not involve the community at large.

access directly an internal state or sensation and name it, is mistaken. Just as the idea of ostension as the primary procurer of meaning is false for the external world of coffee, chess pieces and stone slabs, so the picture of an internal ostension is likewise distorting. The intriguing difference here is that with private sensation, we have very little to build from. These cases are unlike "black" or "beetle" which have common uses that build the signs into meaningful symbols and give them a post or position in language-games (*PI#* 257). Some sensation words can be even more private than "pain," since we do have many common pain responses, and so can learn the meaning of the word as we learn our common language. But without the privileged access of an intentional conscious act, how is it that when I *feel* something that is not publicly accessible my name for it can make any sense at all—even to myself? And so Wittgenstein asks, "How do words *refer* to sensations?" (*PI#* 244).

At first blush, it appears that they cannot. Wittgenstein says...

> Let us imagine the following case. I want to keep a diary about the recurrence of a certain sensation. To this end I associate it with the sign "S" and write this sign in a calendar for every day on which I have the sensation.——I will remark first of all that a definition of the sign cannot be formulated.—But still I can give myself a kind of ostensive definition.—How? Can I point to the sensation? Not in the ordinary sense. But I speak, or write the sign down, and at the same time I concentrate my attention on the sensation—and so, as it were, point to it inwardly.—But what is this ceremony for? for that is all it seems to be! A definition surely serves to establish the meaning of a sign… But in the present case I have no criterion of correctness. One would like to say: whatever is going to seem right to me is right. And that only means that here we can't talk about 'right'. (*PI* #258)

In what Warren Goldfarb (1997) might call a "first blush" reading, this thought experiment does indeed seem to deny an individual's ability to talk about a private sensation, and thus denies the ability to construct a private language, *tout court.*[20] Strawson (1954), for example, sees Wittgenstein calling for some sort of public check or verification of what any word could refer to in order for it to communicate meaningfully, but not even the mark on a calendar could count. And so Strawson, like other first blushers, takes Wittgenstein's [no] "private language argument" at face value.

---

[20] I adapt somewhat Goldfarb's use of "blushes" and cite particular authors who are likely to fall roughly into these categories. Goldfarb (1997) uses the notion of progressively deeper readings or "blushes" to discuss the purported need for something other than normal use to fix the meaning of a general term. I extend his use to cover progressively deeper readings of the private language argument. What I latch onto is that a first blush allows us to see that the justification of the use of the term does not come from the objects designated, or some "property, essence, or attribute which underlies our ascription of a general term" (78), but that the use is "parasitic on the highly variegated features of individual cases" (78, 79). Second blushers, recognizing that there must be some fixity for meaningful use, provide a more complex understanding of how that might be possible for Wittgenstein. For them, according to Goldfarb, "There is, nonetheless, *something* to be grasped" (79). For third blushers, there is no underlying level that prescinds from practices that guarantees fixity to the use of a term (82). This reading, like Goldfarb's Emerson Hall reading, can involve recognizing that the sort of conceptions we bring to bear, and the perceived needs we have, are seductive but mistaken.

At "second blush," however, it seems that Wittgenstein is not denying the possibility of a private language, but correcting our understanding about how such a language might function. The Hintikkas, for example, point out that Wittgenstein does not deny our ability to use sensation words, or other "internal" words, meaningfully (see Hintikka and Hintikka 1986, 249, where they cite *PI* #243, 244, 257, 270, & 300 as evidence). They argue that Wittgenstein came to realize that public, or "physical," language-games constitute the semantics of even "phenomenal" language-games—and not the other way around. We do not start from something like sense-data as basic atoms and then work up towards meaningful language, as Russell believed and as early Wittgenstein may have believed.[21] With this reversal comes a shift in our understanding of language. Wittgenstein noted that "If we construe the grammar of the expression of sensation on the model of 'object and designation [*Bezeichnung*]' the object drops out of consideration as irrelevant" (*PI* #293). The Hintikkas argue that if we distinguish Wittgenstein's use of the *Bezeichnungen,* or *designations,* from his use of *Namen,* or *names*, we see that Wittgenstein was not denying our ability to identify and even name (*nennen*) a private sensation (Hintikka and Hintikka 1986, 254–256). The beetle drops out of the box on the designation model, but that does not mean there are no experiences of beetles and we cannot correctly identify and name a beetle; similarly we can successfully identify and name a toothache when we say we have one.

"Third blushers," like Goldfarb and Dreben, agree that Wittgenstein was not *denying* anything here regarding phenomena (broadly construed) that we encounter, nor was he *denying* our ability to use sensation language meaningfully. They would insist, however, that second-blushers are making the same sort of mistakes that Wittgenstein was trying to steer us away from. To make sense of any argument for or against a private language, one would need to have a grip on the basic terms of the argument, and for third-blushers Wittgenstein is unraveling those terms in order to show up the nonsense of our efforts, pro *or* con. For third-blushers, what Wittgenstein shows is that we don't have a grasp on an "immediate" (with no mediating contextual structures) "private" (completely sealed from any expressing behavior) "sensation" (construed along the lines of a physical object). This is the heart of the Emerson Hall reading: Wittgenstein is attempting to take the Cartesian picture seriously and finds that none of those three words have any sense in this special picture of meaning constitution; if sense relies on the three notions Wittgenstein undermines, any purported argument *for* or *against* a private language is nonsensical.

We agree to an extent with first blushers in that some form of criteria for meaningful use must be developed. We also see Wittgenstein as rejecting any form that would rely on an intentional pointing, i.e., the sort that would have to precede marking "S" on the calendar. Certainly, according to Wittgenstein, if we deny access to any bodily expression of a sensation or any references to (internal or external) context, a language cannot get off the ground. But we don't see Wittgenstein rejecting the possibility that criteria of sorts may develop by which a sensation word may be used meaningfully and appropriately (whether or not it is parasitic on "external" language or would communicate anything to others

---

[21] There is controversy in Wittgenstein scholarship as to whether or not Wittgenstein intended his simple objects to be Russellian sense data, and there is controversy regarding whether or not his efforts were genuinely aimed at building a better *Begriffsschrift* (concept language) than Frege or, as New Wittgensteinians claim, he was showing that any such attempt was nonsense.

would be further questions). So we agree with second blushers, like the Hintikkas, who believe that Wittgenstein was not denying the possibility of referring to sensations nor a private language outright. The intentional "pointing" or designation (*Bezeichnung*) relation misconstrues language learning and functioning and is parasitic on a more primary stage-setting for the understanding of a word's meaning. But what can do the job here, with sensation words, that language-games and external cues perform for the naming words we commonly use? What could set that stage? Like third-blushers, we agree that— though denying nothing our common use affirms—Wittgenstein may have seen no way out; he could not, himself, resolve the dilemma surrounding how our ordinary sensation words develop meaning. Like second-blushers often do, however, here we will use Wittgenstein's insights to craft a solution to a real problem Wittgenstein raises but does not resolve. We will present a connectionist solution that also accommodates some third-blush insights, since it questions the notions of an *immediate* grasp of a *sensation* (conceived as an original object) that is purely *internal* and private.

2.2.2 Sparse Distributed Memory

PI #258 clearly supports the idea that there can be no way of tying words directly to private sensations. It indicates that one needs more than attention to a raw sensation, memory and a designating mark in order to develop a language-game in which, e.g., "S" can be the name of a sensation.

Wittgenstein is denying that one can build an identification of something raw into an independent criterion, i.e., a basis for knowledge or the *definiens* of a definition. He is emphasizing the need for a background of reliable cues and uses to be in place before a sign can be meaningful. And he questions the use of memory, as well as intention, as an accurate original source of a criterion of correctness. A connectionist approach can show how one might start to form a schema that can act as a general criterion for judging when some such sensation is present. Such a schema can be the basis of the conceptual content that would make a sign a meaningful symbol. There is little to go on, as Wittgenstein points out, but a connectionist approach helps explicate the manner in which private experiences *can* develop into meaningful uses of language even here where the information can be sparse.

Sparse Distributed Memory or SDM (Kanerva 1988) is a technology for content-based storage and retrieval of high-dimensional vector representations like that used in Latent Semantic Analysis (LSA) and VSA.[22] An SDM consists of some (arbitrary) number of address vectors, each with a corresponding data vector (Addresses and data can be binary 0,1 values, or + 1,-1 values as in VSA). The address values are fixed and chosen at random, and the data values are initially zero. To enter a new address/data pair into the SDM, the Hamming distance (count of the elementwise differences) of the new address vector with each of the existing address vectors is first computed. If the new address is less than some fixed distance from an existing address, the new data is added to the existing data at that address. To retrieve the item at a certain "probe" address, a similar comparison is made between the probe address and the existing addresses, resulting in a set of addresses less than a fixed distance from the probe. The

---

[22] LSA (Landauer and Dumais 1997) models individual word meaning in a VSA and Wittgenstein-friendly way, while VSA models compositionality.

data vectors at these addresses are summed, and the resulting vector sum is converted to a vector of 0's and 1's by converting each non-negative value to 1 and each negative value to 0.

Such a memory is called *sparse* because the set of actual addresses is a tiny fraction of the set of possible addresses; e.g., for a million addresses of a thousand bits each the fraction of addresses used is $10^6 / 2^{1,000}$ or 9×10^{-296}. As well as supporting the storage and retrieval of distributed representations, the memory of an SDM itself is distributed in the sense that the storage and retrieval of each item takes place over a set of locations.

As illustrated by Denning (1989), the distribution of each pattern across several locations produces a curious property: given a set of degraded exemplars of a pattern (such as the pixels for an image with some noise added), an SDM can often reconstruct the "ideal" form of the pattern through retrieval, even though no example of this ideal was presented to it. One can expect, then, that in cases where there is no original exemplar (e.g., no unitary object-like sensation that can be identified) but there are contextually situated iterations presenting similar patterns—which would give them proximal addresses—the patterns distributed across memory can be conjoined to produce an exemplar or schema. The development of a schema or "ideal" pattern heralds some criteria by which to judge. Smolensky would describe such schemata as "harmonious peaks" in similarity relations (Smolensky 1988, 78–80).

On this picture, the sensation itself would not be like a "thing" we could "point" to, but more of a pattern that we could begin to recognize. Over time, when different but similar enough combinations of contextually situated cues accumulate (i.e., when several proximal addresses are built), a paradigmatic instance of the sensation could emerge (the proximal vectors would constitute a sparse address). A situation emerges where one could say, "*that* was like *the feeling I had before*," where both occurrences display features distributed proximally in the same vector space, but neither is quite the ideal or schema.

Even before language emerges, the ability to feel and respond to an "internal" sensation can build on the model of a SDM in us and in other animals. The ability to represent in language could ride upon these motor representational abilities (Langacker 2008) but might need further cues, or a further development of cues, to place the sensation in a language. The "internal" naming with "S" is likely modeled, and parasitic upon, semantic habits developed in "external" language-games (Robinson Crusoe, after all, also learned how to use language in community before he become isolated). But, as with "external" naming words, stage setting precedes any meaningful designation. Also the internal-external distinction breaks down; the contextually situated cues we need to identify and name sensations need not be restricted to events inside the body.

Wittgenstein suggests that one cannot tell, even privately, whether one is using the word correctly if there is nothing but the sensation going on, so one does not have a special internal or intentional grip on the *right* use of the term. One needs something like a context and other behavioral cues that give consistent sense to the sign, as second blushers like the Hintikkas insist. What SDM shows, consistent with *PI# 258*, is how an ideal or proper use for the sign can be built up from variegated uses of "S" even if the iterations of the "S" tying it to the sensation, and the internal or external cues for doing so, were not "correctly" used (from some imagined God's eye point of view that we do not have). The sensation need not be raw and self-identical, the memory need not be

accurate, and there need be no necessarily concomitant feature or cue to act as a hard criterion, but the emergence of an exemplar that we can use as a criterion for judging when the sensation is appropriately named is successfully achieved. So, as with third blushers like Goldfarb, there need be no originary sensation conceived on the basis of a material object, to which we have immediate access by some form of internal mental pointing.

SDM shows how there need be no immediate access to a unique private sensation that the sign "S" captures in order for a meaningful language-game containing the sign "S" to get off the ground (This would result in what the Hintikkas called a primary, as opposed to secondary language-game (Hintikka and Hintikka 1986, chapter 11)). The iterations and contextual cues come to form a schema or exemplar that can act as a soft constraint on the meaning of the word. Each identifiable criterion adds weight and proximity connecting the term and the signs/cues, but no particular aspect of the collective criteria is necessary; a "criterion" is typical, but it is also defeasible, e.g., it is still a pain, and "I am in pain" can be used properly, even if I control myself so that I do not grimace. The cues that point us to the presence of the sensation (or the collection of behaviors typically concomitant with the experience we name "S") are part of the distributed representation. They make sense only in the context of the wider language-game in which we use the term "S".

With the example of beetles in boxes, Wittgenstein is not denying that there are beetles nor that they play any role in our use of the word; Similarly, Wittgenstein is not denying that there are sensations that may help constitute the meaningful use of a sensation word. But before our ability to identify a sensation and name it emerges with a schema, we do not start with an isolated thing, like a beetle. What we start with is "not a something, but not a nothing either!" (*PI*#304) And once we have a schema, and "S" can take up a role in our language, the *something* that would be at the other end of an individual designation is no longer required for the word to be meaningfully used.

Without any "external" checks to help provide a use for the sign we could not reliably establish a meaning, but from a fairly small number of instances, we can accumulate occurrences with their associations, e.g., my stomach growling and it being around 12 o'clock when I say "S," and I can begin, so to speak, to build a box around the beetle—and then we can bracket the beetle: the private sensation itself need not function to provide the rule by which even I will successfully use the word. And the rule by which I use the word is not fixed permanently, nor is it a rule of ordinary English grammar, or a rule of Mentalese that such a grammar may purport to approximate.

Wittgenstein questioned how a schema can come to guide us. How does one come to use a particular leaf, or the overlap of a bunch of leaf samples, as the schema of leaf (*PI*# 73, 74), or how do we judge that one rod in Paris is a meter and the criterion of the proper use of "meter" (*PI*# 50)? When examined, the usual sort of representational identifications we are tempted to employ break down. They show themselves as tautological or fall to regressive arguments. SDM shows one way that such exemplars can emerge and be employed even where the cues are primarily internal and successive similarities in their patterns are elusive. In an emergent stochastic process like the kind Denning describes (Denning 1989), schemata can develop and act as attractor basins in memory. Circularity and regression are avoided as a pattern emerges and receives

reinforcement in the architecture of memory storage and in the predictive adequacy of the sign in use (Clark 2013).

2.2.3 From Symbol to Systematic Language

SDM is an example of a VSA memory storage technique that shows how symbols might be grounded in difficult cases; we can even develop a symbol from a few distributed occurrences of iterations with obscure contextual cues. We see also how this is done efficiently, without the need for many rounds of exposure or training, which provides VSAs with pedagogical plausibility. Classicists, however, would see the symbols as needing to fit into a system of linguistic rules before such symbols could be systematically combined in productive uses of language. Before we show how a VSA approach can begin to handle systematicity and productivity through the development of analogical structures, we will consider some related approaches to systematicity and productivity that allow us to reinforce the biological plausibility of VSA connectionism.

One reason for thinking that connectionist systems must lack a capacity for systematicity and productivity is the context dependency of the symbols, which would seem to belie compositionality. It was believed that symbols would need to be discrete atoms in order to participate in the variable places left open in a grammatical structures. Additionally, it was believed that the experience-based development of connectionist structures could not produce the sort of rules that assure the systematicity and productivity of language, without either implicitly relying on classicist rules or without simply being a mere implementation of a classicist architecture (McLaughlin 2014).

A move towards systematicity happens with the notion of distributed representations that shows symbols to be distinct without being atomically discrete, but also discrete enough to enter into higher-order relations; that is, the symbols can become "encapsulated" and further operations can be performed on them (Borensztajn et al. 2014). Connectionist memory here plays an important role. Architectures that include features like SDM or Hawkins's Memory Prediction Framework (MPF) (Hawkins and Blakeslee 2004; used in Borensztajn et al. 2014) display the features of dynamical systems in that a symbol or schema can act as an attractor basin that streamlines the organization and retrieval of information. On our preferred VSA model, such encapsulated representations are not fully context independent. The volume of the information is greatly reduced in memory storage, i.e., the dimensionality of the tensor product is of the same dimensionality as the constituents, but even after further manipulation the constituent information can be retrieved from memory with very little degradation (Plate 2003) and symbols/ representations can continue to change to some extent as more related instances are added to a VSA memory.[23]

Contrary to classicist assumptions, there is evidence that linguistic structure can be built in connectionist architectures via the development of symbols without the

---

[23] As more related instances are added, the parts that are common across instances effectively reinforce each other, effectively increasing their magnitude relative to the idiosyncratic parts. Given that a VSA memory has fixed capacity, in the long run the magnitude of the idiosyncratic parts is so low that they effectively disappear from the memory.

need for anything like grammatical rules (Elman 2014; Frank 2014). These lexicon approaches work with the ability of symbols to bind together into meaningful units on the basis of their own complex internal structure (Bod et al. 2003). This move from symbols to grammatical patterns can be seen in Jeff Elman's work in Simple Recurrent Networks (SRN), in which the context that sets the basis for distributed representations can extend out to produce rule-like grammatical regularities (Elman 2014). So we see here that context dependence does not automatically rule out systematicity, but even shows a way for meaningful language to get off the ground without preexisting rules or a rigid lexicon.

These approaches require a degree of context independence that allows the symbols to function meaningfully when connected with other independent units. In this, they mimic Wittgenstein's own early approach. In the *Tractatus* picture, atomic objects had something like valences that allowed for particular sorts of connections with other atomic objects. Objects hung together to form atomic facts like the links in a chain (*Tractatu*s, 2.03) without the "third man" of external rules. These facts were described by atomic sentences that connected the names of objects (objects here included relations for Wittgenstein). Connectionist constructions add to this picture by incorporating the context dependence of names/symbols that develops through the situated utterances that constitute them. The distributed representations initially set the ground for the possibility of meaningful connections. Regularities in use reinforced through training; and the subsequent ability to combine in different ways provides the compositionality that allows for productivity without classical symbols and rules. Thus Stephan Frank, building on Elman's work shows a way to develop what Hadley (1994) described as "strong systematicity" with connectionist symbols, although Frank notes that a symbols and rules approach combined with statistical probability still worked slightly better in mimicking our language use than his connectionist model (Frank 2014).

VSAs can have an advantage over Frank's or Elman's approach here because VSAs can dispense with a reliance on backpropagation. To get from a lexicon to language, some training can be required, but as we see with SDM, there is no need for tens of thousands of epochs of backpropagation.[24] In building to schemata, our model here shows how meaningful signs and exemplars can be formed and used more efficiently. But, as with backpropagation, we can get to the same phenomenal results we see in language learning and use without positing a deep grammar or logic.

These advances in memory storage and retrieval techniques may also provide a more biologically plausible account than backpropagation, since according to Borensztajn et al. (2014) connectionist memory storage and processing techniques can produce hierarchically structured encapsulations, which can better model the progress of a network through different areas of the brain.

The expansion of soft symbols into a full language provides one approach to productivity. Another approach would be to see the symbols as capable of filling something like roles that develop with analogical structures. Both these approaches suggest a systematicity that depends on what looks like a part-whole

---

[24] Hebbian learning (Hadley 1994), can also be used to train networks with fewer examples in a more psychologically plausible way. Kanerva (2010) provides an elegant example of this sort of "one-shot" learning for analogies.

constituency relation at the coarse level, with variable places that can be filled so that the sentential unit can be meaningful. But while they have the productivity of a symbols-grammatical rules approach, they do not have nor need empty places in a structure at the processing level. We will provide an indication of how this rule-like behavior with variable-like positions can develop in connectionist architectures in the following section.

2.3 Enter the Matrix: Soft Rules and Systematicity

Wittgenstein sees that one is misled to think that when "anyone utters a sentence and *means* or *understands* it he is operating a calculus according to definite rules" (*PI*#81). Wittgenstein would acknowledge that "Slab!" can be made to fit into a classicist linguistic structure after the fact, but a requirement to view "Slab!" as an ellipsis of "Bring me a slab" would essentially be a distortion. The rules come post hoc and can summarize our linguistic behavior. They do not normally provide the source of linguistic productivity much less an explanation of why we have it. Connectionist networks can produce meaningful utterances without the need for a deep grammatical structures or rules, and connectionists can see such rules as a coarse-grained account of how meaningful language is produced and actually operates.

2.3.1 Raven's Matrices: Rule-Following without Rules

Having discussed the way in which VSA provides a formalism that can support compositionality and systematicity (in 1, "Bracketing the Beetle"), and how SDM (in the VSA family) can provide the soft symbols and exemplars (in 2, "Boxing the Beetle"), we now illustrate how VSAs can generalize rule-like behavior from an incomplete pattern. This example is due to Rasmussen and Eliasmith (2011), who show how a VSA-based neural architecture can solve the Raven's Progressive Matrices task. In this task, subjects are given a figure like the one below (simplified for our purposes here), and are asked to complete the missing piece at lower right:
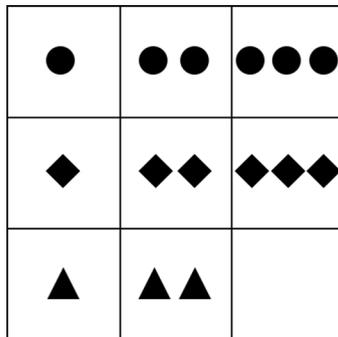
Some candidate solutions are shown in the figure below. The first candidate is the correct one.
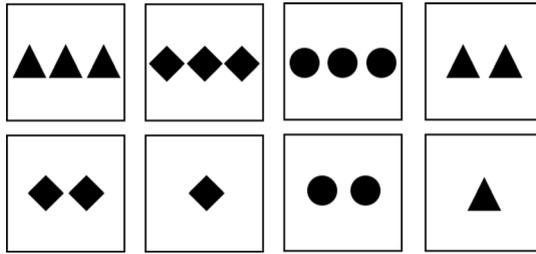
Asked how they arrived at this solution, a person might report that they followed these two rules:

1. *There is one item in the first column, two in the second, and three in the third.*
2. *There are circles in the first row, diamonds in the second, and triangles in the third.*

As Rasmussen and Eliasmith argue, however, a VSA can solve this problem without recourse to such explicit rules.[25] Each element of the matrix can be represented as a set of attribute/value pairs; for example, the center element would be

$$< shape > \otimes < diamond > + < number > \otimes < two >$$

Solving the matrix then corresponds to deriving a mapping from one item to the next. As Rasmussen and Eliasmith show, such a mapping can be obtained by computing the vector transformation from each item to the item in the row or column next to it. The overall transformation for the entire matrix is then the average (or sum) of such transformations.[26]

For Wittgenstein "no course of action could be determined by a rule, because any course of action can be made out to accord with the rule" (*PI*# 201). He thus undermined the notion that we have some special grip on what ought to be the next number in a sequence, or "how to go on" more generally, due to the grasping of an underlying rule. This Raven's Matrices example shows the power of a network to predict the next item in a succession without some special spark of understanding. This pattern identification also heralds the beginning of an ability to deal with counterfactuals, i.e., what should or could be next given the accumulated information stored, and this provides a principled approach to demonstrating linguistic productivity. While rooted in experience, the language develops the means to go beyond experience in a meaningful way.

The VSA approach to Ravens Matrices provides the ability to generalize from examples without the need for variables, rules, or excessively repeated training. The ability to form analogical patterns and use them in learning provides an avenue of response to the systematicity and productivity objections posed by classicists. Roles

---

[25] Rasmussen and Eliasmith use a different binding operator (circular convolution) than ours (elementwise multiplication), and hence derive a different inverse operator, but this does not affect our argument or results.
[26] In Levy et al. (2014) we provide details of these transformations.

present structures that ask what sort of symbol might meaningfully fill or follow; actors present solutions with high degrees of probability for filling or following. It becomes intuitively plausible to see how a language with roles and actors can develop in a connectionist architecture, which VSA formalism can track.

2.3.2 Against Mere Implementation

The analysis of language into rules and symbols is what Smolensky would call a psychological rather than a causal explanation (Smolensky 1995, 260). The coarse-grained account is helpful to our understanding and catches emergent relations between distinct linguistic units, but such an analysis can take several different forms (rules-symbols, concept-object, role-filler; subject-verb-object, etc.) without catching the underlying mechanisms at work in producing what, from the coarse-grained level, look like hard symbols and necessary rules. In his recognition that "uniqueness is not a problem" (Smolensky 1991, 175), Smolensky displays a change from the early Wittgenstein to the late. Whereas in the *Tractatus* Wittgenstein thought that one analysis provided a sufficient explanation, i.e., the analysis from facts/sentences to objects/names, later he saw that many different analyses were possible and the sort of constituents one would find was contingent on the analysis that one employed (*Philosophical Grammar*, 210–212).[27]

Connectionism might be considered an implementation of the symbols and rules approach if there were a seamless bridging or construction of classical structures from connectionist algorithmic processes. Seamlessness would allow one to argue that the upper reduced to the lower or the lower merely implements the upper. But as we build up from symbols to language in a lexicalized grammar approach, or up from analogical structures to roles in a rule-like approach, we see that there is a disjunction between the causal processes that connectionist architectures employ and the sort of distinctions we would expect to find from any classical linguistic analysis. While the symbols and rules approach might provide an emergent level of description, we see a disjunction, though not a sharp one, at the level of symbols (as encapsulated representations or dynamic attractors) and then at the level of roles (in which rule-like features come from connectionist ability to build analogical structures). These transitions suggest a more fluid picture of language than the classical approach allows and also suggests an emergentist rather than a reductionist or an implementational account.

3 Conclusion: Wittgenstein's Challenges for Twenty-First Century Connectionism

While other avenues have been developed to meet the challenges posed to connectionism in the twentieth century (see e.g., Calvo and Symons 2014), we have shown how VSA architectures perform exceptionally well.[28] As Rasmussen and Eliasmith say, "In

---

[27] For Dreben, defending a New Wittgensteinian view, the change here was that Wittgenstein recognized his approach to showing the nonsensicality of a general analysis in the *Tractatus* was too a priori and dogmatic; he had thought one analysis could display the line between sense and nonsense once and for all, and later he saw that the line constantly shifted and many different analyses could be put to use.

[28] Another promising model and serious contender to the VSA approach is the LISA model of Hummel and Holyoak (2005).

order to represent a Raven's matrix in neurons and work on it computationally, we need to translate the visual information into a symbolic form. Vector Symbolic Architectures (VSAs; Gayler 2003) are one set of proposals for how to construct such representations. VSAs represent information as vectors, and implement mathematical operations to combine those vectors in meaningful ways" (Rasmussen and Eliasmith 2011, 143). While other approaches can be consistent with VSA, Kanerva emphasizes that with SDM the interesting properties only begin to emerge when you have high-dimensional vectors (Kanerva 1993). An equally important feature is the distributed nature of the vector representation; that is, the fact that each entity is represented by a pattern of activity over many computing elements, and each computing element is involved in representing many different entities (Hinton et al. 1986). Our examples, together, add weight to a high dimensional vector approach in which lower-level processes are necessary but mean nothing in themselves for linguistic understanding until higher-order holistic qualities emerge.

We have shown ways that connectionism meets the challenges of biological, pedagogical and linguistic plausibility and, in the examples we have brought to bear, we also show that VSAs present promising architectures for advancing twenty-first century connectionisms. Also, by showing similarities between Wittgenstein's insights and connectionist approaches, we save Wittgenstein's insights regarding (a) soft symbol construction (b) soft rule-following, and (c) soft logic in language, while retaining (d) systematicity and productivity. In "Bracketing the Beetle" we showed how a connectionist formalism can work to capture important features of language, without relying on a fundamental connection with objects for meaning, in a biological plausibility time scale. In "Boxing the Beetle" we demonstrated how symbols and schemata represented in such a formalism can emerge in a pedagogically plausible way from limited information, while accommodating, to some extent, a Drebenian third blush reading of the private language argument. And in "Enter the Matrix" we showed how we can gain rule-like behavior without following external rules. This, along with connectionist circuitry, provides the basis for the operators in our formalism and indicates how we might go on to develop roles for fillers through the development of analogical structures. By this approach, or by a lexicon approach that builds on the SDM development of symbols, we thus indicate routes to non-classical compositionality and linguistic plausibility. Furthermore, we noted how the disjunction between the from-the-ground-up constructive approaches and the distinctions we commonly bring to bear in linguistic analysis indicate that connectionist models are not a mere implementation of a classical LOT.

Dreben, however, would not be satisfied. "Those connections to Wittgenstein are all superficial," he might say. "There are *deeper* reasons why the project of cognitive science is *totally at odds* with Wittgenstein's views." Some of the deeper reasons involve interpretations of connectionism that promote a mechanistic and reductivist understanding of thought and language. Diane Proudfoot, for example, sees connectionist approaches as replicating elements of a Cartesianism that Wittgenstein attacks (Proudfoot 1997); and reductions that would make connectionist models an implementation (reduced up) or neurological (reduced down) would violate the "insulation thesis" that separates the grammatical from the causal for Wittgenstein (see Klagge 2011). The promise of connectionist's "neurologically inspired" architectures to better model thought and language have animated hopes for a more thorough reduction of

higher cognition to biology—and even bolstered eliminativist hopes to undermine folk psychology (Churchland 1989). While Smolensky's approach is consistent with an emergentist account (McClelland et al. 2010) and neither reductivist nor eliminativist, it can still seem to conflate the mind with the brain (see Bennett and Hacker 2003, on the mereological fallacy) and conflate a mathematical model with what we are and do. For instance, that last conflation seems implicit when Smolensky says, "*The mind is a statistics-sensitive engine operating on structure-sensitive (numerical) representation*" (Smolensky 1991, 176). Computational approaches (classic and connectionist) in general tend to take a mechanistic approach to thought and language, which can be construed as reductivist (Shanker 1998), but they need not be.

Wittgenstein would, no doubt, recoil against both the mechanistic and reductivist interpretations of thought and language, but Wittgenstein does allow for the construction of mind-models for scientific purposes (e.g. *Brown Book*, 1933–5, 117, 118; *Philosophical Grammar* Wittgenstein 1978, 48) and connectionists are not committed to the philosophical difficulties that Smolensky's interpretations can sometimes insinuate. We believe that connectionists can avoid conceptual confusions that Wittgenstein would warn against, such as (1) the conflation of the causal and the grammatical, (2) the (mis)application of scientific techniques to grammatical fictions, and (3) confusion surrounding the notion and use of representation. One need not conflate the model with the mind, or the model with the brain, or the mind with the brain, or any of the above with what we are as persons (Bennett and Hacker 2003).

This last point raises the question of the value of a computational model as itself providing an explanation. Unlike Mills, Wittgenstein would not see connectionism as an *explanation* for what he *describes*, but he might see it as a helpful avenue toward a proper explanation. As a mind-model, it can help guide inquiry into the workings of the phenomena and can dispel some misconceptions, but as close as it may come to analogically portraying some important features, it should not be mistaken for the only or the actual way that language works.

By respecting Wittgenstein's insights and providing a VSA account that displays linguistic compositionality, integrates soft symbols, and develops analogical structures that can be systematic and advance productively, we have shown how twenty-first century connectionism can address what appeared to be limitations in the functionality of its operation, limitations in learning, and limitations in biological plausibility that might have thwarted connectionism's ability to be a better mind-model for language and cognitive science.

References

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

Bechtel, W., & Abrahamsen, A. (2002). *Connectionism and the mind*. Oxford: Blackwell.

Bennett, M., & Hacker, P. (2003). *Philosophical foundations of neuroscience*. Oxford: Blackwell-Wiley.

Bennett, M., Dennett, D., Hacker, P., & Searle, J. (2009). *Neuroscience and philosophy: Brain, mind and language*. New York: Columbia UP.

Bod, R., Scha, R., & Sima'an, K. (2003) *Data-oriented parsing*. Center for the Study of Language and Information.

Borensztajn, G., Zuidema, W., & Bechtel, W. (2014). Systematicity and the need for encapsulated representations. Calvo and Symons, 2014, 165–189.

Box, G.E.P. (1979). Robustness in the strategy of scientific model building. In R. L. Launer & G. N. Wilkinson (Eds.), *Robustness in statistics* (pp. 201–236). Cambridge: Academic Press.

Calvo, P., & Symons, J. (Eds.). (2014). *The architecture of cognition: Rethinking fodor and pylyshyn's systematicity challenge*. Cambridge: MIT Press.

Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge: MIT Press.

Chomsky, N. (1980). *Rules and representations*. Oxford: Basil Blackwell.

Churchland, Paul (1989) Folk psychology and the explanation of human behavior. In *A Neurocomputational Perspective* Cambridge: MIT Press, 111–127.

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences, 36*(3), 181–204.

Crary, A., & Read, R. (Eds.). (2002). *The New Wittgensteinian*. London: Routledge.

Denning, P. J. (1989). Sparse distributed memory. *American Scientist, 77*, 333–335.

Diamond, C. (1995). *The realistic spirit: Wittgenstein, philosophy and the mind*. Cambridge: MIT Press.

Dror, I., & Dascal, M. (1997). Can Wittgenstein help free the mind from rules? In Johnson, D., Erneling, C., (eds.) The philosophical foundations of connectionism (pp 217-226). Oxford: Oxford UP.

Eliasmith, C. (2013). *How to build a brain: A neural architecture for biological cognition*. Oxford: Oxford UP.

Elman, J. (2014). Systematicity in the lexicon: Having your cake and eating it too. Calvo & Symons, 2014, 115–146.

Fodor, J. (1997). Connectionism and the problem of systematicity: Why Smolensky's solution still doesn't work. *Cognition, 62*, 109–119.

Fodor, J., & Pylyshyn, Z. (1988). Connectionism and cognitive architecture: A critical analysis. In C. MacDonald & G. MacDonald (Eds.), (1995). *Connectionism: Debates on psychological explanation (Vol. 2). Oxford: Blackwell* (pp. 90–163). New York: Springer.

Frank, S. (2014). Getting real about systematicity. Calvo & Symons, 2014, 147–164.

Gayler, R. (2003). Vector symbolic architectures answer Jackendoff's challenges for cognitive neuroscience. In P. Slezak (Ed.), *ICCS/ASCS international conference on cognitive science* (pp. 133–138). Sydney, Australia, University of New South Wales: CogPrints.

Goldfarb, W. (1997). Wittgenstein on the fixity of meaning. In W. W. Tait (Ed.), *Early analytical philosophy: Frege, Russell, Wittgenstein. Essays in honor of Leonard Linsky* (pp. 75–89). Chicago: Open Court.

Goldstein, L., & Slater, H. (1998). Wittgenstein, semantics and connectionism. *Philosophical Investigations, 21*(4), 293–314.

Hacker, P. M. S. (1989). Wittgenstein on language and rules. *Philosophy, 64*, 5–28.

Hacker, P. M. S. (2001). *Wittgenstein: Connections and controversies*. Oxford: Clarendon Press.

Hacker, P. M. S., & Baker, G. P. (2009). *Wittgenstein: Rules, grammar and necessity*. Oxford: Blackwell-Wiley.

Hadley, R. F. (1994). Systematicity in connectionist language learning. *Mind & Language, 9*(3), 247–272.

Hawkins, J., & Blakeslee, S. (2004). *On intelligence*. New York: Times Books.

Hintikka, J., & Hintikka, M. B. (1986). *Investigating Wittgenstein*. Cambridge: Blackwell.

Hinton, G.E., McClelland, J.L. and Rumelhart D.E. (1986). Distributed representations. In McClelland, J. L., & Rumelhart, D. (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition. Cambridge: IT Press* vol. 1, pp. 77-109. New York: Springer.

Hummel, J. E., & Holyoak, K. J. (2005). Relational reasoning in a neurally plausible cognitive architectures: An overview of the LISA project. *Current Directions in Psychological Science, 14*(3), 153–157.

Jackendoff, R. (2003). *Foundations of language*. Oxford: Oxford UP.

Kanerva, P. (1993). Sparse Distributed Memory and related models. In M. H. Hassoun (Ed.), *Associative neural memories: Theory and implementation* (pp. 50–76). New York: Oxford University Press.

Kanerva, P. (1994). The spatter code for encoding concepts at many levels. In M. Marinaro & P. G. Morasso (Eds.), *ICANN '94, Proceedings of the international conference on artificial neural networks (Sorrento, Italy)* (pp. 226–229). London: Springer–Verlag.

Kanerva, P. (1988). *Sparse Distributed Memory*. Cambridge, Mass: MIT Press

Kanerva, P. (2009) Hyperdimensional computing: An introduction to computing in distributed representation with high-dimensional random vectors. *Science and Business Media* (published online January).

Kanerva, P. (2010) What we mean when we say "What's the dollar of Mexico?": Prototypes and mapping in concept space. *Quantum Informatics 2010: AAAI-Fall 2010 Symposium on Quantum Informatics for Cognitive, Social, and Semantic Processes.*

Klagge, J. C. (2011). *Wittgenstein in exile*. Cambridge: MIT Press.

Landauer & Dumais. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review, 104*(2), 211–240.

Langacker, R. W. (2008). *Cognitive grammar: A basic introduction*. New York: Oxford UP.

Levy, S., Lowney, C., Meroney, W. & Gayler, R. (2014) Bracketing the beetle: how Wittgenstein's understanding of language can guide our practice in AGI and cognitive science. Artificial General Intelligence, 2014 Lecture Notes in Artificial Intelligence (LNAI), 8598 Proceedings.

MacDonald, C., & MacDonald, G. (1995). *Connectionism: Debates on psychological explanation* (Vol. 2). Oxford: Blackwell.

Malcolm, N. (1954). Wittgenstein's Philosophical Investigations. *Philosophical Review, 63*(4), 530–559.

McClelland, J., Botvinick, M., Noelle, D., Plaut, D., Rogers, T., Seidenberg, M., and Smith, L. (2010). Letting structure emerge: Connectionist and dynamical systems approaches to cognition. *Trends in Cognitive Science, 14* , 348–356.

McLaughlin, B.P. (2014) Can an ICS architecture meet the systematicity and productivity challenges? In Calvo & Simons, 2014, 31-76.

Mills, S. (1993). Wittgenstein and connectionism: A significant complementarity? *Royal Institute of Philosophy Supplement, 34*, 137–157.

Pinker, S., & Prince, A. (1988). On language and connectionism: Analysis of a parallel distributed processing model of language acquisition. *Cognition, 23*, 73–193.

Plate, T. (2003). *Holographic reduced representations*. Stanford: CSLI Publications.

Proudfoot, D. (1997). On Wittgenstein on cognitive science. *Philosophy, 72*(280), 189–121.

Quine, W. V. O. (1960). *Word and object*. Cambridge: MIT Press.

Ramsey, W., Stich, S., & Garon, J. (1990). Connectionism, eliminativism and the future of folk psychology. In C. MacDonald & G. MacDonald (Eds.), (1995). *Connectionism: Debates on psychological explanation (Vol. 2). Oxford: Blackwell* (pp. 311–338). New York: Springer.

Rasmussen, D., & Eliasmith, C. (2011). A neural model of rule generalization in inductive reasoning. *Topics in Cognitive Science, 3*, 140–153.

Rumelhart, D.E. & McClelland, J.L. (1986) On learning the past tense of English verbs. In J.L. McClelland & D. Rumelhart (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition. Cambridge: IT Press* vol. 2, pp. 216-271. New York: Springer.

Rumelhart, D.E., Smolensky, P., McClelland, J. and Hinton, G.E. (1986) Schemata and sequential thought processes in PDP models. In J.L. McClelland & D. Rumelhart (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition. Cambridge: IT Press* vol. 2, pp. 7–57. New York: Springer.

Shanker, S. (1998). *Wittgenstein's remarks on the future of AI*. London: Rutledge.

Smolensky, P. (1987). Connectionist AI, symbolic AI, and the brain. *Artificial Intelligence Review, 1*, 95–109.

Smolensky, P. (1988). On the proper treatment of connectionism. Behavioral and Brain Sciences, 11 (1), 1–23 reprinted in MacDonald, C., & MacDonald, G. (1995). *Connectionism: Debates on psychological explanation (Vol. 2). Oxford: Blackwell* (pp. 28-89). New York: Springer.

Smolensky, P. (1990). Tensor product variable binding and the representation of symbolic structures in connectionist systems. *Artificial Intelligence, 46*, 159–216.

Smolensky, P. (1991). Connectionism, constituency and the language of thought. In C. MacDonald & G. MacDonald (Eds.), (1995). *Connectionism: Debates on psychological explanation (Vol. 2). Oxford: Blackwell* (pp. 164–198). New York: Springer.

Smolensky, P. (1995). Reply: Constituent structure and explanation in an integrated connectionist/symbolic cognitive architecture. In C. MacDonald & G. MacDonald (Eds.), (1995). *Connectionism: Debates on psychological explanation (Vol. 2). Oxford: Blackwell* (pp. 223–290). New York: Springer.

Stern, D. (1991). Models of memory. *Philosophical Psychology, 4*(2), 203–217.

Stewart, T. & Eliasmith, C. (2012). Compositionality and biologically plausible models. Hinzen, W., Machery, E. & M. Werning (eds.) (2012) *Oxford handbook of compositionality*. New York: Oxford.

Stewart, T., Tang, Y., & Eliasmith, C. (2011). A biologically realistic cleanup memory: Autoassociation in spiking neurons. *Cognitive Systems Research, 12*(2), 84–92.

Strawson, P. F. (1954). Wittgenstein's "Philosophical Investigations". *Mind, 63*, 70–99.

Wittgenstein, L. [1922] (1995) *Tractatus logico-philosophicus*. C.K. Ogden (trans.) London: Routledge.

Wittgenstein, L., (1953) *Philosophical investigations [PI]*. G.E.M. Anscombe (trans.) Oxford: Blackwell.

Wittgenstein, L., [1933–5] (1958). *The blue and brown books*. New York: Harper and Row.

Wittgenstein, L. (1978) *Philosophical grammar*. Rush Rhees (ed.), Anthony Kenny (trans.) Berkeley: University of California Press.

Affiliations

Charles W. Lowney II[1] · Simon D. Levy[2] · William Meroney[3] · Ross W. Gayler[4]

Simon D. Levy
levys@wlu.edu

William Meroney
meroneyw@aol.com

Ross W. Gayler
ross@rossgayler.com

[1]　Department of Philosophy, Hollins University, 7916 Williamson Road, Roanoke, VA 24020, USA

[2]　Department of Computer Science, Washington and Lee University, 204 W. Washington Street, Lexington, VA 24450, USA

[3]　Meadows of Dan, VA 24120, USA

[4]　Melbourne, Australia