LANGUAGE, COGNITION AND NEUROSCIENCE
https://doi.org/10.1080/23273798.2018.1501501

REGULAR ARTICLE

Interactionism in language: from neural networks to bodies to dyads

J. Benjamin Falandays, Brandon J. Batzloff, Samuel C. Spevack and Michael J. Spivey

Cognitive and Information Sciences, University of California, Merced, CA, USA

ABSTRACT
In a science of language, it can be useful to partition different formats of linguistic information into different categories, such as phonetics, phonology, semantics, and syntax. However, when the actual phenomena of language processing cross those boundaries and blur those lines, it can become difficult to understand how these different formats of linguistic information maintain their integrity while engaging in complex interactions with one another. For example, if the function of a cortical network that is known to process phonetics is immediately taking into account contextual influences from a cortical network that is known to process semantics, then it seems clear that this "phonetic cortical network" is doing more than just phonetics. In the neuroscience and cognitive science of language, the scope of analysis where different formats of linguistic information are seen to interact reveals a wide array of context effects in almost every possible direction. When one expands the scope of analysis to include nonlinguistic sensory modalities, such as vision and action, research is showing that even those lines are getting blurred. Visual perception and motor movement appear to influence various aspects of language processing in real time. As this scope of analysis expands further still, research is showing that two human brains and bodies exhibit various forms of synchrony or coordination with one another during natural conversation. Interactionism at all these levels of analysis poses a challenge to traditional frameworks that treat different components of language, perception, and action as operating via domain specific computations.

ARTICLE HISTORY
Received 24 January 2018
Accepted 6 July 2018

KEYWORDS
Psycholinguistics; interaction; context effects; networks; coordination

Introduction

In this article, we skim through a wide range of literature so that we can showcase the commonality across seemingly disparate phenomena in real-time language processing at short timescales of milliseconds, medium timescales of seconds, and longer timescales of minutes. The commonality observed is a general form of interactionism, whereby a cognitive mechanism that may seem mostly dedicated to one domain of information processing appears to nonetheless be sensitive to influences coming from a different domain of information processing. It could be semantics influencing phonetics, or pragmatics influencing syntax. It could be motor information influencing lexical processing. Or it could be one person's brain activity influencing another person's brain activity, via the transmission of linguistic information. The overarching lesson proffered from this overview is that, while it may be descriptively convenient to refer to a cognitive subsystem as "a syntax module," or as "a language module" or as "a motor system" or as "a muscle group" or as "a person," the fact is that such a label is a decidedly narrow approximation of the actual range of capabilities of that subsystem. And using the singular determiner, "the," to refer to that subsystem falsely implies that it is bounded and independent from its neighbouring subsystems. It is not.

As we run through several courses for this review, we also serve up some highly simplified network simulations, as a kind of case study in how interactionism might work among a set of generic subsystems that become connected. By delving into the dynamics of this network of networks, we can see how each net changes its behaviour a little bit as a result of becoming connected to the other net. As a result of our treatment, the reader may be tempted to relinquish the idea that context effects involve the daunting task of one subsystem (e.g, phonology, syntax, or semantics) requiring an extra "interface module" in order to translate influences from another subsystem that supposedly uses completely different rules and symbols (Jackendoff, 2002). Instead, context effects might be re-conceptualised as two different subsystems, with different internal dynamics, nonetheless using a common currency of information flow that allows them to interface directly, and they achieve agreement between each other in the form of an emergent harmony in their activity patterns. Uncertainty or conflict in one subsystem gets alleviated simply by its resonating with another subsystem

CONTACT Michael J. Spivey ✉ spivey@ucmerced.edu

that has less uncertainty or conflict. The two subsystems become one system -- at least a little bit, and at least for a little while.

Interactivity at the timescale of milliseconds

We begin our investigation at the timescale of milliseconds, which corresponds in the language system to the time-course of perceptual processing, in which acoustic information is categorised into meaningful linguistic units (McMurray, Tanenhaus, Aslin, & Spivey, 2003; Pisoni & Tash, 1974). Models of speech perception at this timescale fall into two main categories: modular, feedforward models (e.g. Fodor, 1983; Forster, 1979; Norris & McQueen, 2008), and parallel interactive models (e.g. Marslen-Wilson, 1987; McClelland & Elman, 1986). The former class of models contend that acoustic-phonetic processing is the first in a series of independent stages, with lexical processing occurring only after the phonetic module has completed its work, followed by semantic processing, and so on. While such an architecture makes intuitive sense, there is now a large body of behavioural and neurological evidence demonstrating that speech perception is readily influenced by lexical and semantic context (Borsky, Tuller, & Shapiro, 1998; Gow & Olson, 2016; Miller, Heise, & Lichten, 1951). Although it is possible to stretch a modular view to account for some of these findings, the proposed distinctions between phonetic, lexical, and semantic modules quickly lose their value in light of a speech perception system that rapidly integrates information of all three kinds. Furthermore, recent advances in neuroimaging techniques have allowed for more conclusive demonstrations of interactivity among subsystems than was previously possible with behavioural and electrophysiological methods alone (Gow & Olson, 2016). The current state of the literature now definitively points to a highly parallel, interactive architecture of speech perception.

Context effects on speech perception are most prominent when information from the acoustic signal is limited, such as acoustically ambiguous or perceptually degraded stimuli. Behavioural research demonstrating such effects dates back to the 1950s, beginning with work by George Miller and colleagues showing that sentential context improves the intelligibility of speech-in-noise (Miller et al., 1951). Contextual cues can also result in the *illusion* of intelligibility when stimuli are not just noisy, but actually missing acoustic information. For example, in demonstrations of the phoneme restoration effect, phonemes are deleted from a word and replaced with noise, such as a cough. Listeners typically fail to notice the deletion of the phoneme, instead perceiving the noise as the phoneme that would normally be present given the lexical context (Warren, 1970). In an ERP study of the phoneme restoration effect, Groppe et al. (2010) found that the electrophysiological response to deleted phonemes differed depending upon whether sentential context biased the likelihood of a particular phoneme in the gap, or was neutral with respect to the missing phoneme, and that this difference is evident as early as 192 milliseconds after the onset of the deleted phoneme. Given the rapidity with which context effects impinge upon speech perception, it seems unlikely that this influence occurs merely at a post-perceptual stage.

Effects of context on speech perception are also found with acoustically ambiguous stimuli. In work showing this, ambiguous stimuli are constructed by manipulating an acoustic cue, such as voice onset time, to create tokens spanning an acoustic continuum between two phonemes (e.g. /g/ and /k/) (Borsky et al., 1998). When presented in isolation, the tokens from the middle of the continuum are ambiguous such that they are categorised as either phoneme approximately 50% of the time. However, when these ambiguous tokens are embedded in a lexical or sentential context, participants are strongly biased towards the categorisation that results in a meaningful word or phrase (Connine, 1987; Ganong, 1980). For example, following a sentence such as "the man put on his … ", a token ambiguous between "goat" and "coat" is much more likely to be perceived as the latter (Borsky et al., 1998).

While the evidence reviewed above clearly supports interaction between acoustic-phonetic information and lexical-semantic information, it is still possible to account for some of these findings within a feed-forward, modular architecture. This is because behavioural data rely on overt, categorical judgments occurring at the end of a trial, at which point acoustic-phonetic, lexical, and semantic processing have all since completed their work. As such, it is difficult to determine the point at which interaction between different forms of information takes place; it may occur during acoustic-phonetic processing, or it may occur at a post-perceptual decision stage. While the ERP results of Groppe et al. (2010) demonstrate that semantic information is rapidly integrated, even this result is inconclusive, since feed-forward models need not make a commitment to the time-course of processing stages. In spite of decades of research, behavioural and electrophysiological measures have simply been unable to distinguish between competing accounts of speech perception.

Some researchers have turned to neuroimaging as a possible solution to this issue. Since BOLD imaging has the ability to localise brain activity with a high spatial resolution, fMRI methods have the potential to determine the relative activation of regions implicated in acoustic-phonetic processing, lexical processing, and semantic processing while participants respond to perceptually degraded or ambiguous stimuli with varying levels of context. Obleser and colleagues conducted a series of fMRI studies in which acoustic intelligibility and semantic predictability of sentences were varied (Obleser & Kotz, 2010; Obleser, Wise, Dresner, & Scott, 2007). They found that intelligibility was correlated with degree of activation in the superior temporal sulcus (STS) regardless of predictability, while predictability modulated activity in frontal regions. For degraded sentences in which predictability yielded the greatest improvement in performance, increased activity was also seen in left parietal regions. The authors interpreted these results as showing that when stimuli are successfully encoded in auditory cortex (STS), less input from regions associated with semantics (frontal and left parietal regions) is required. On the other hand, when stimuli have low acoustic intelligibility, functional connectivity between auditory and semantic regions is increased (Obleser et al., 2007; Obleser & Kotz, 2010). In another fMRI study, Guediche, Salvata, and Blumstein (2013) held intelligibility constant while manipulating sentential context along with the acoustic ambiguity of the final word in a sentence. Ambiguous tokens that were most influenced by sentential context resulted in increased activity in middle temporal gyrus (MTG), a region associated with lexical and semantic processing, and decreased activity in superior temporal gyrus (STG), which is associated with acoustic-phonetic processing. Consistent with the findings of Obleser and colleagues, Guediche et al. (2013) interpreted these results as supporting an interactive account of perceptual and semantic processing.

Despite these compelling results, due to the low temporal resolution of BOLD imaging, fMRI studies are unable to make causal inferences regarding the timing of activation in various brain regions, and therefore have failed to resolve the debate between modular and interactive accounts of speech perception. This difficulty was finally overcome in a study by Gow and Olson (2016), in which they used MRI-constrained MEG and EEG data to achieve both high spatial and temporal resolution of brain activity while participants heard acoustically ambiguous tokens embedded in sentential contexts. They applied Granger causality analysis to the time series data from several regions of interest. Granger causality analysis looks for statistical patterns in one time series that also show up in a different time series slightly later. The conservative interpretation is that the first time series can be used to *forecast* the second time series, but a bolder interpretation is that the process that generated the first time series may have some form of *causal influence* on the process that generated the second time series. Gow and Olson's causality analysis found that the temporal patterns of activity in left middle temporal gyrus and left supramarginal gyrus (MTG and SMG, two regions implicated in lexical and semantic processing) were correlated with slightly later patterns of activity in posterior superior temporal gyrus (pSTG, a region implicated in acoustic-phonetic processing) – suggesting that lexical-semantic processes were influencing acoustic-phonetic processes. These results convincingly demonstrate that interaction between acoustic-phonetic subsystems and lexical-semantic subsystems occurs *during* perceptual processing, and not in a later stage.

Interactivity at the timescale of seconds

We now expand our analysis to the timescale of seconds, which is the time course over which sentences are grammatically parsed. At this level, it can be seen that subsystems associated with syntax are coordinated with subsystems related to semantics and discourse information. The influence of context on syntax is often demonstrated through induction of the garden-path effect, which is the tendency to incorrectly parse a sentence when phrases are temporarily ambiguously attached. For example, in a sentence beginning "the defendant examined … ," the defendant may be parsed as the subject of the verb "examined," or as the object of it, as in "the defendant examined *by the lawyer* … " (Trueswell, Tanenhaus, & Garnsey, 1994). By recording readers' eye movements, Trueswell et al. (1994) found that the garden-path effect was influenced by whether the first noun was animate or inanimate. When the first noun was animate, participants were more likely to parse it as the subject, resulting in increased processing difficulty when contradictory information was encountered (i.e. when subjects first read "*by the lawyer*," and then realise that the initial noun is actually the object of the verb). This effect was much smaller for inanimate nouns, and was modulated by the semantic fit between the inanimate noun and the verb (Trueswell et al., 1994). McRae, Spivey-Knowlton, and Tanenhaus (1998) expanded upon these results to show that the garden-path effect can be induced by varying semantic fit alone while holding animacy constant. For example, in a sentence such as "the crook arrested … ", the crook is more likely to be parsed as the object of the verb "arrested" (McRae et al., 1998). These results suggest that semantic information has immediate effects on the syntactic parsing of sentences.

Syntactic processing has also been shown to interact with contextual information introduced by discourse context (Altmann & Steedman, 1988). One source of information available from discourse is the number of potential referents of a verb. Spivey and Tanenhaus (1998) had participants read sentences such as "The actress selected … ". Discourse history was manipulated by preceding these target sentences with context containing either one or two potential referents, and therefore biasing a relative or main clause interpretation, respectively. For example, a discourse context such as "The director chose one actress and not the other. The actress selected … " introduces two potential referents, therefore biasing an interpretation where "actress" is the object of "selected" (i.e. indicating which of the two actresses was selected).

Gender and order-of-mention of potential referents are two other variables in discourse context that have also been shown to influence syntactic parsing (Arnold, Eisenband, Brown-Schmidt, & Trueswell, 2000). Arnold et al. (2000) had participants listen to sentences containing personal pronouns ("he" or "she") while looking at visual displays containing male and female referents. When the discourse introduced two characters of different genders, subjects used gender information to determine the referent of the pronoun. When the context introduced two potential referents of the same gender, eye-movement data revealed that the personal pronoun was ambiguous. However, participants were more likely to interpret the pronoun as referring to the person who was mentioned first in the discourse.

In addition to the behavioural results reviewed above, EEG studies have also shown that semantic and discourse context can immediately influence syntactic processing. Work of this kind often relies on the N400 component of ERP signals. The N400 is a negative deflection in the EEG signal that is elicited in response to semantically surprising stimuli, such as when encountering the word "dog" in the sentence "I take my coffee with cream and *dog*" (Kutas & Hillyard, 1980). Given this property of the N400, it can be used as a measure of the degree to which participants expect a given word, which in turn is influenced by discourse context, as we have seen above. Using this method, Van Berkum, Hagoort, and Brown (1999) had participants listen to short stories containing a target word that was semantically sensible in the context of the local sentence, but semantically inconsistent relative to the discourse history. For example, a critical sentence such as "Jane told the brother he was exceptionally *slow*" would follow a context in which the brother was described as being very fast. Results showed a larger N400 component for inconsistent words relative to consistent ones, beginning around 200–250 milliseconds after word onset.

Van Berkum and colleagues later extended these findings to show that discourse context can result in highly specific predictions of upcoming words, and that these predictions contribute to the online parsing of a sentence (Van Berkum, Brown, Zwitserlood, Kooijman, & Hagoort, 2005). Participants listened to short stories in Dutch that indirectly supported a prediction of either a male or female noun in the upcoming discourse. After this context was introduced, the next segment of the story included a gender-marked adjective that was either consistent or inconsistent with the gender of the upcoming noun. EEG data showed a larger N400 in response to adjectives that were inconsistent with the predicted noun, showing that the predicted noun influenced parsing online *before* it was even introduced in the discourse.

In another study using this methodology, Nieuwland and Van Berkum (2006) showed that discourse history can also function to *overrule* the effects of semantic violations on parsing. Normally, when participants hear sentences that contain verb-object animacy contradictions (e.g. "the girl comforted the *clock*"), processing difficulty is increased for the object noun ("clock") and a larger N400 component can be seen, reflecting a semantic violation. However, Nieuwland and Van Berkum (2006) found that when preceding discourse history supported the personification of the object noun (e.g. a story in which a girl is speaking to a depressed clock), the N400 effect was completely eliminated. Moreover, in a discourse context supporting personification of the object noun, it was actually *easier* for subjects to process violations of animacy than predicates that returned to the canonical inanimacy of the noun (Nieuwland & Van Berkum, 2006).

The behavioural and electrophysiological evidence reviewed in this section clearly point to a high degree of interaction between linguistic subsystems specialised for syntactic processing, and those related to semantics and discourse context. Syntactic processing is inherently temporal in nature, as words must be integrated with one another while they are encountered one by one. Given this temporal quality of syntax, measures of processing difficulty and electrophysiological response as new words are encountered make it clear that semantic and discourse context influence syntactic processing directly, and not in some later stage of processing. Just as the previous section showed that speech perception interacts with semantic context on the timescale of milliseconds, one can now conclude that semantics and discourse context also interact with syntactic processing on the timescale of seconds.

Interactivity among subsystems in general

Before we expand the timescale further, allow us to explore a theoretical interlude here. We would like the reader to consider thinking of phonetics, phonology, semantics, and syntax as processed not by domain-specific modules that each use their own idiosyncratic rules and symbols, but instead as processed by neural subsystems who share the common currency of neuronal activation patterns. Based on the kinds of findings mentioned in the previous sections, it behooves one to no longer think of linguistic subsystems as lined up in sequence, with a phonology box sending its output only to a lexicon box which sends its output only to a syntax box, etc. (Forster, 1979). Instead, it appears that each of these different linguistic subsystems can send its output to any and all of the other subsystems (e.g. Onnis & Spivey, 2012; Spevack, Falandays, Batzloff, & Spivey, 2018).

One can think of it like Elman's (1990) simple recurrent network, where syntax and semantics both emerge out of the dynamic patterns of temporal order among words, co-existing in one undifferentiated matrix of synaptic strengths (see also Tabor & Tanenhaus, 1999). Alternatively, one could treat syntax and semantics as separate subsystems that have substantial interaction with each other, as in Spivey-Knowlton's (1996) integration-competition model, where (in the face of an ambiguity) syntactic/structural biases encode their probabilistic preferences, and semantic biases encode their own probabilistic preferences, and while those preferences compete against one another, a recurrent integration layer allows emerging confidence in semantics to "leak" into syntax to help resolve its ambiguity (see also McRae et al., 1998). Or one could think of it as a kind of blend between neural network accounts and symbolic accounts, such as a gradient symbol system (Smolensky, Goldrick, & Mathis, 2014). In Cho, Goldrick, and Smolensky's (2017) gradient symbol system simulation, the incremental parsing of a sentence is realised as a constantly changing "garden of forking paths" in the state space of the grammar (a term they borrow from Jorge Luis Borges). Whether it is a neural network (composed of unlabelled micro-features) or a set of probabilistic biases (composed of labelled features), or a constantly changing tensor product that combines disparate information formats, in every case the key to the solution is *interaction* across those seemingly different formats of information.

Once the processor for information format A has been systematically influenced by the processor for information format B, one can no longer truly claim that the former processor is a module that is solely dedicated to processing information of format A. Clearly, it is capable of incorporating information of format B in a systematic fashion as well. If one treats these processors as functioning via the same basic physical mechanisms, e.g. the common currency of neuronal activation patterns, then it is no longer the *format* of information that differs across the processors, it is merely the *content* of information. In other words, so long as two networks accept compatible formats of inputs and outputs (here being neuronal activity in both cases), then even if the pattern of activation or link structure differ between the two networks, in principle they may interact directly. Essentially, a context effect (like the ones discussed above) can be re-interpreted as nothing more than two neural systems that start out in conflict eventually finding a kind of resonance (Grossberg, 2013) or harmony (Smolensky, 2006) between them.

For one of these neural systems, imagine a fully-interconnected network of idealised neurons that are passing their activation back and forth to each other. It might use a logistic activation function that takes the current activation plus net incoming activation and squashes it into a range between 0 and 1 (e.g. Rumelhart & McClelland, 1986). With a random synaptic weight matrix connecting all 100 nodes, any individual synaptic strength might lie between about -3 and $+3$. We use this relatively abstract simplified form of network here so that the results have the potential to be applied to roughly any kind of network. Even though the weight matrix is random, and its starting activation values are random, such a network will often settle into a steady cyclic pattern of activation over time, within a couple hundred time steps. In this cyclic pattern, some nodes will be maxed out near 1.0 activation, exhibiting some subtle oscillation only in the third or fourth decimal place, and some nodes will be flat-lined near 0.0 activation, with oscillations again visible only in the third or fourth decimal places. Still other nodes will be oscillating steadily with activation changes in that first decimal place.

To examine the network's overall behaviour, one can plot over time the trajectory of the entire network's activation pattern in a reduced-dimension state space. With 100 nodes, Net A's activation over time is a trajectory in a rather opaque 100-dimensional state space. But when a third of the nodes' activations are averaged into one value over time, along with another third and another third, one can plot a 3-D data visualisation of Net A's behaviour over time (Figure 1(A)). Note how the trajectory loops on itself, approximately revisiting certain pathways repeatedly. When a separate network, Net B, is analyzed in this same manner, its behaviour is revealed to be a very precisely repeated pattern composed of 11 points in state space (Figure 1(B)). This limit-cycle pattern of behaviour is moderately common among these random networks. Net A could be a region of cortex that mostly processes syntax, while Net B could be a region of cortex that mostly processes semantics. Alternatively, Nets A and B could be two different people. In fact, Nets A and B could even be two different social networks of 100 people each. When we analyze all 200 nodes among the two separate nets as though they were one state space (Figure 1(C)), we are treating these two separate fully-intra-connected networks as though they were one not-well-connected network. This analysis reveals little or no repetition of trajectories, indicating that these two unconnected networks do not behave as one system. And why would they? They share absolutely no information with each other at this point.

This glaring disconnect between Nets A and B is further illustrated by single-cell recordings from these simulated networks. For example, Figure 2(A) shows the activation over time for node #92 in Network A. And Figure 2(B) shows the activation over time for node #84 in Network B. In this case, Net A is a "closed system," and Net B is a different "closed system," each of them experiencing no influence from outside of its own synaptic weight matrix. The activation pattern of node #92 in Net A is completely uninfluenced by, and uncorrelated with, the activation pattern of node #84 in Net B (Pearson correlation coefficient, $r = -.02$).

But what if one connects these two networks? Imagine that these two networks acquired a single bi-directional synaptic connection between them. Instead of being closed systems, they would now each be an "open system" that is subtly influenced by the other (Yoshimi, 2012). We have chosen these two, Nets A and B, as a case study for exploring what happens when networks get connected to each other. Node #18 in Net A and node #6 in Net B (not pictured here) each exhibits a sizeable range of activation, so we connected them to each other bi-directionally with synaptic strengths of 1.0.

Thus, amid the 10,000 synaptic connections between nodes inside Net A and the 10,000 synaptic connections between nodes inside Net B, there now exists *two* additional connections between Net A's node #18 and Net B's node #6. Among the torrent of activation influences that pass back and forth throughout Net A, a tiny sliver of it "leaks" into Net B, and vice versa. After a few hundred time steps, these two just-barely-connected networks begin to resonate with each other. Net A begins to speak a bit of Net B's "language" and Net A begins to speak a bit of Net A's "language." For example, node #92 in Net A, which used to be completely uncorrelated with node #84 in Net B, now produces a time course of activation that substantially resembles that of node #84, with a Pearson correlation coefficient of $r = .51$ (See Figure 2(C) and 2(D)).

We are not suggesting that synchronised neural firing patterns are the only way that interactions and context effects can happen. There may be a variety of ways that different brain areas exert their contextual influences on one another. However, recent EEG studies have found that phase synchrony between the large-scale neuronal activation patterns in two different brain regions may be fundamental to the occurrence of context effects between them (e.g. Hauk, Giraud, & Clarke, 2017; Lam, Bastiaansen, Dijkstra, & Rueschemeyer, 2017; Lewis, Schoffelen, Hoffmann, Bastiaansen, & Schriefers, 2017; Rommers, Dickson, Norton, Wlotko, & Federmeier, 2017). Importantly, in our little network

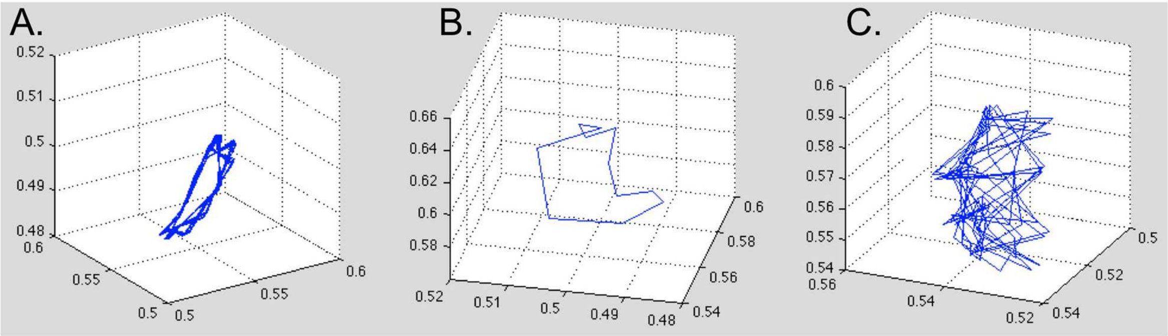

Figure 1. Before connected: One hundred time steps of state-space trajectories for Network A (panel A) and Network B (panel B). The shared state space for these two separate nets (panel C) shows little or no coherence, because they are unconnected. (Each dimension averages activation for a third of the nodes.)

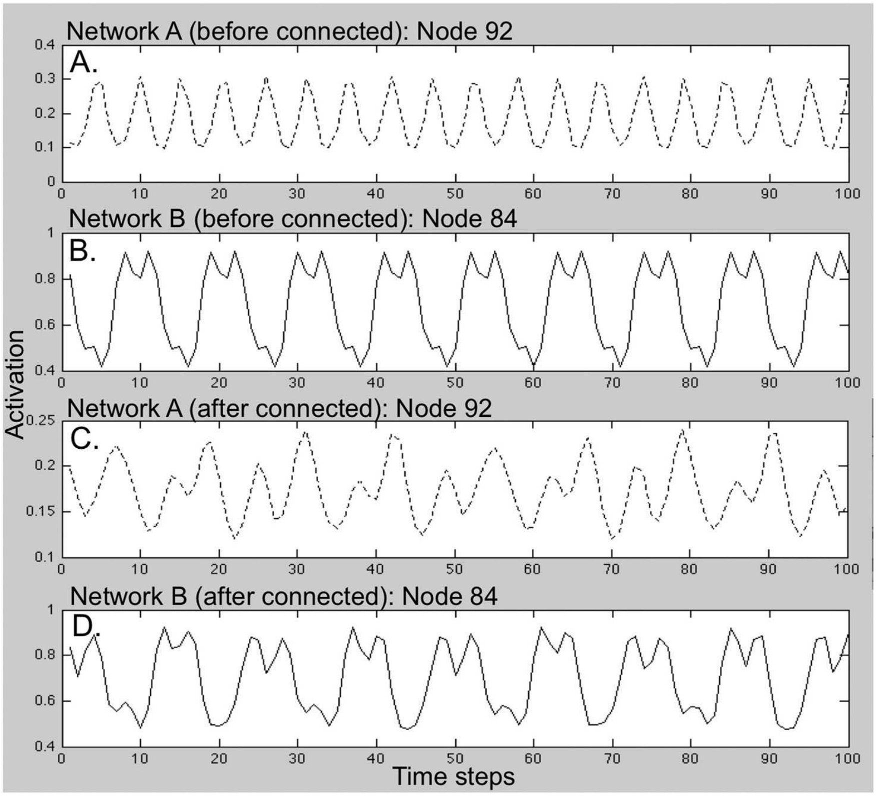

Figure 2. One hundred time steps of activation values for node #92 Network A (panel A) and node #84 in Network B (panel B), showing no correlation between the two ($r = -.02$). When Nets A & B are connected by two additional synaptic weights (panels C & D), the activation pattern for node #92 in Network A becomes significantly correlated with the activation pattern of node #84 in Net B ($r = .51$).

simulation, the emergence of this synchronisation is not instantaneous. It takes a few hundred time steps of activation flow for the correlations to show up. When we look at periods of 50 time steps at a time, we see that the correlation between node #92 in Net A and node #84 in Net B starts out with an r-value that wavers between −.2 and +.2, statistically non-significant. (See the solid line with asterisks in Figure 3). However, once the two nets acquire that tiny connection at time step 0, the correlation between these two nodes (who are not directly connected to each other) steadily rises to the .5 range. The same kind of behaviour emerges from the majority of nodes in Net A when compared to randomly selected nodes in Net B. Figure 3 (dashed lines with open circles) shows correlation functions for a random sample of 8 more pairs of nodes from Nets A & B (again, not directly connected to each other). In several cases, the pair of nodes from the two different networks gradually develops phase synchrony, a positively correlated activation pattern over time, just like Net A's node #92 did with Net B's node #84. Also, in a couple of cases, the pair of nodes develops a statistically significant *negative* correlation, indicating that they are rising and falling in anti-phase synchrony. Only in two instances does the pair of nodes settle into an uncorrelated pattern of activation over time (with the Pearson correlation coefficient near zero).

Now that these two networks have a tiny connection between them, when we look at their trajectories in state space, the overall behaviour has noticeably changed. Net A's mostly-regular limit cycle has become somewhat less regular, but still exhibits numerous repeated pathways, showing that it has maintained some coherence while being influenced by Net B (Figure 4(A)). Similarly, Net B's highly-regular limit cycle is now less precise and has taken on a second loop to its pattern (Figure 4(B)). Importantly, when the two nets are analyzed as one 200-node state space, it no longer shows quite the chaotic pattern that it did before the nets were connected. Figure 4(C) reveals a number of repeated pathways in the trajectory, suggesting that these two nets are now behaving a little bit like one coherent system – even though the only thing connecting them is node #18 in Net A and node #6 in Net B. In continuing explorations of these random networks, we are testing the effects of one-way communication, connecting different types of nodes across the nets, and non-random patterns of connectivity. Additionally, in discussing the synchrony of subnetworks shown here, it is important to note that the two networks had the same size, the same kind of connectivity (fully interconnected, with random weights), and random initial activation patterns. It is possible that with very different types of networks or inputs to those networks, the two would be less prone to achieving synchrony. This issue will be addressed in more detail in our section regarding constraints on interactionism.

This glimpse into a highly simplified exploratory simulation, though brief, is nonetheless revealing enough to provide some intriguing insights for how one might think about context effects in a network of interacting elements. Rather than developing complex stories for how one linguist's semantic calculus could interface with another linguist's syntactic structures, one can instead imagine having these different processors (that are roughly specialised for different types of linguistic information) using the same currency of information transmission, such as basic activation and inhibition. In this case, a context effect is imposed simply by one network interacting with another network, and the two of them finding new patterns of behaviour while they resonate together as one. The nodes in these networks are idealised and simplified enough that they could stand for neurons inside a brain region, or they could stand for different brain regions forming a network that stretches across cortex, or the two nets could even stand for two people in a dyadic conversation. When one network gets even slightly connected to another network, they often cannot help but influence one another and drift toward exhibiting some of the shared synchrony of one system (Buzsáki & Draguhn, 2004; Kelso, 1997; Strogatz, 2004).

Interactivity at the timescale of several seconds

If the reader is now thinking of phonetics, phonology, semantics, and syntax as processed not by domain-specific modules that each use their own idiosyncratic rules and symbols, but instead as neural subsystems that share the common currency of neuronal activation patterns and influence each other directly, then it should not be too big of a stretch to extend that thinking to other nearby brain regions, such as visual and motor areas of cortex. Indeed, there is much evidence that visual environment and bodily arrangement act as further contextual influences on language processing.

In the case of vision, the incremental word-by-word understanding of language guides, and is guided by, how one attends to their visual environment. Listeners are much more likely to look at objects that are related to what they hear than objects that are not (Cooper, 1974; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995). Thus, eye fixations around the visual environment provide insight into how participants are processing linguistic input. This is exactly what is done in the visual world experimental paradigm (Allopenna, Magnuson, & Tanenhaus, 1998; Tanenhaus et al., 1995). This paradigm has shown how visual attention is time-locked to linguistic phenomena at all levels and has revealed insights into phonological (e.g. Allopenna et al., 1998), syntactic (e.g. Tanenhaus et al., 1995), semantic (e.g. Yee & Sedivy, 2006), conceptual (Knoeferle & Crocker, 2006), pragmatic (e.g. Sedivy, Tanenhaus, Chambers, & Carlson, 1999), and predictive (Altmann & Kamide, 2007) linguistic processes.

One particular insight illustrated by the visual world paradigm is that syntactic ambiguity resolution can be constrained by visual context. For example, in a study by Tanenhaus et al. (1995), subjects viewed a scene and received instructions to manipulate objects within it, such as "Put the apple on the towel in the box." This sentence is temporarily syntactically ambiguous, since the prepositional phrase "on the towel" can be validly interpreted as a destination for the apple, or as a modifier (e.g. "the apple on the towel"). Eye fixations to the relevant objects in the scene revealed which interpretations were being considered as the sentence unfolded. The researchers showed that listeners are likely to initially misinterpret the phrase as a destination when there is one apple present in the visual scene, but as a modifier when there are two apples present (e.g. "the apple on the towel" distinguishes between two possible apples). Hence, the visual environment altered tendencies for sentence parsing in real time. Similar findings have been extended to other ambiguous sentences (Snedeker & Trueswell, 2004), anaphoric reference disambiguation (Kaiser & Trueswell, 2008), and pragmatic disambiguation of references (Sedivy et al., 1999).

Visual context has also been revealed to affect sentence processing via ERP studies measuring the P600 and N400 ERP components, the size of which reflect the presence of unexpected syntactic and semantic information, respectively. Knoeferle, Habets, Crocker, and Münte (2007) had participants listen to German sentences of the form noun-verb-noun. While the first nouns of such constructions are typically subjects of the verb, they may be objects in some instances. The grammatical case of the first noun is only fully disambiguated by the case marking of the final noun. Participants tend to assume the first noun is the subject until hearing the non-canonical object-case marker on the second noun, at which time a P600 effect is observed. However,

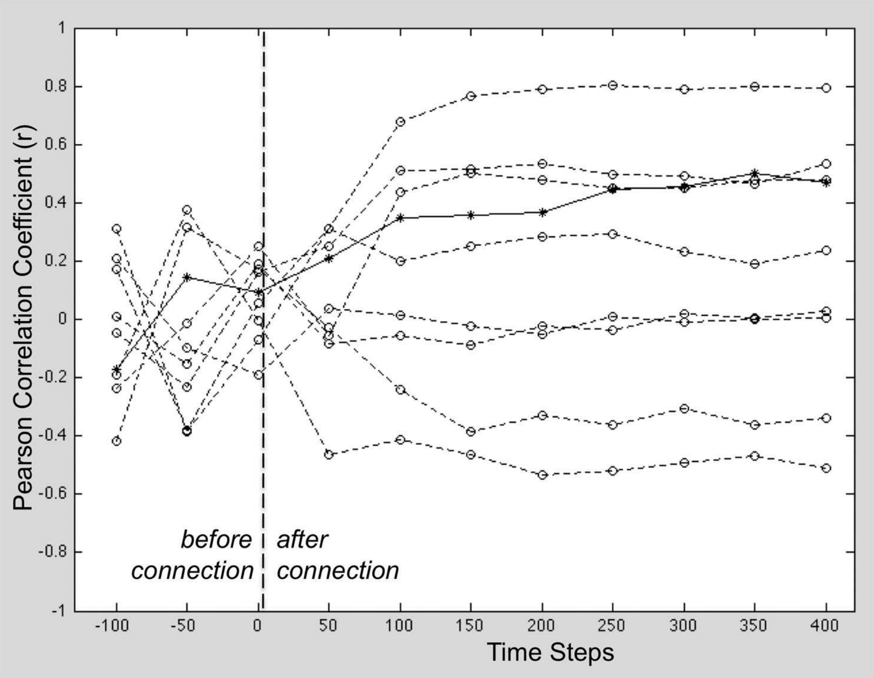

Figure 3. Correlation coefficients for pairs of nodes across Net A and Net B. Once Net A's node #18 is connected to Net B's node #6, any randomly selected node from Net A and any randomly selected node from Net B are reasonably likely to develop significantly correlated activation patterns over the course of a couple hundred time steps.

when sentences are accompanied by images that only match a non-canonical syntactic interpretation (e.g. the first noun is the object), a P600 effect was observed immediately after the verb instead of on the second noun, indicating that the visual context aided in an earlier disambiguation. In similar picture-sentence verification tasks looking at semantic disambiguation, an N400 effect has been observed immediately when there is a mismatch between the action depicted in the image and the verb in the sentence (Knoeferle, Urbach, & Kutas, 2011). These results demonstrate that visual information can exert immediate influences on syntactic and semantic processing.

Even without a co-present visual context for the linguistic input, there is evidence that sensorimotor systems are implicated in the processing of language (Barsalou, 1999; Richardson, Spivey, Barsalou, & McRae, 2003). For example, Zwaan and colleagues (Stanfield & Zwaan, 2001; Zwaan, Stanfield, & Yaxley, 2002) have shown that the mental simulation of language is sensitive to the visuospatial implications of linguistic input. Zwaan et al. (2002) had participants read sentences such as “an eagle is in the sky.” They were immediately shown a line drawing of an object and asked to judge if the object was present in the sentence. If participants utilise sensorimotor simulation to make these judgments, then the judgments should reflect compatibility between the way the object is depicted in the drawing and the way it was perceptually simulated. The characteristics of the simulation are in turn influenced by the sentential context. For example, an eagle that is in the sky would be flying, and therefore a simulation of it would be predicted to have spread wings, rather than folded. Indeed, Zwaan et al. (2002) found that participants were faster to report a match when the shape of the object shown matched the shape that was implied by the sentence. Similar examples of the sensorimotor simulation of language include the implied orientation of objects (Stanfield & Zwaan, 2001) and negation (Anderson, Huette, Matlock, & Spivey, 2010).

The contextual effects of non-linguistic information on language processing are not limited to the visual domain, but include the motor domain as well. Glenberg and Kaschak (2002) had participants judge if sentences were sensible by producing reaching movements either towards or away from their bodies. The sentences implied motion that was in a direction towards or away from oneself (e.g. “Close the drawer”). When the direction of the response was compatible with the implied direction of the sentence, the response was faster. This was also true of sentences which imply non-literal motion, such as "Liz told you the story." Zwaan and Taylor (2006) found similar results with clockwise and counterclockwise responses for sentences that imply circular motion such as "He turned down the volume." These results show that reading about actions primes activations of compatible actions, and interferes with incompatible actions.

Transcranial magnetic stimulation (TMS) research has added causal evidence for the involvement of the motor cortex in sentence processing. In a study by Buccino et al. (2005), participants received a single TMS pulse to the hand or feet areas of the motor cortex while listening to sentences involving foot/leg or hand/arm actions. Motor evoked potentials (MEPs) in response to these pulses were measured. Listening to action sentences, but not abstract sentences, resulted in reduced MEPs in corresponding muscles, indicating that activation of the motor cortex was modulated by the action sentences. Similar participation of motor cortex in language comprehension has been identified in phonological perception (Iacoboni, 2008; Pulvermüller et al., 2006), lexical decision tasks (Pulvermüller, Hauk, Nikulin, & Ilmoniemi, 2005) and with passive reading of handwritten words (Gordon, Spivey, & Balasubramaniam, 2017).

The studies in this section illustrate how sensorimotor systems influence syntactic and semantic aspects of sentence processing over the time course of several seconds during which we make sense of sentences in the context of their (real or imagined) sensorimotor consequences. Just as lexical processing is guided by sentential and discourse context, visual environment and bodily states may also serve as contextual influences in the larger interactive cognitive network that processes linguistic information.

Interactivity at the timescale of minutes

Over the timescale of minutes, communicating individuals can be seen as interconnected networks, whereby the understanding and production of language in each individual is affected by the activity of the other. Synchronisation of these sub-networks is displayed through coordination and mimicry across individuals of not only word choice and structural priming, but also motor activity. It is not just different cortical regions that engage in various forms of synchronisation to carry out their coordination (Hauk et al., 2017); different body parts and different people do so as well.

A common example of this coordination is speech alignment and phonetic convergence. As speaker A is listening to speaker B, speaker A will tend to unintentionally mimic the intonation, pitch, stress, and rhythm of speaker B (Goldinger, 1998; Pardo, 2013). Conversation partners also tend to coordinate their syntactic patterns and word choices (Dale & Spivey, 2006; Fusaroli & Tylén, 2016). Moreover, conversation can also induce non-linguistic motor coordination. Postural sway becomes coordinated in dyadic discourse (Shockley, Santana, & Fowler, 2003). Eye movements become coordinated during conversation about a shared visual display (Richardson, Dale, & Kirkham, 2007). Head and facial movements are frequently unintentionally mimicked by people during face-to-face conversation (Chartrand & Bargh, 1999; Louwerse, Dale, Bard, & Jeuniaux, 2012). And even brain activity becomes correlated between two people during spoken communication (Kuhlen, Allefeld, & Haynes, 2012; Pérez, Carreiras, & Duñabeitia, 2017).

Taken together, linguistic, motor, and neural synchronisation in a dyad are used to generate parameters for the shared meaning and purpose of the dialogue that is being co-created by the interlocutors (Clark, 2012). The cessation of synchrony and mimicry between communicating individuals is also informative. De-synchronisation occurs when people argue or disagree (Paxton & Dale, 2013) and when one party is attempting to deceive the other (Duran, Dale, Kello, Street, & Richardson, 2013). The cessation of synchrony is thus the cessation of productive communication and an indicator of failure to achieve a shared intention.

Given this synchrony between individuals, viewing the dyad as two interconnected sub-networks is a reasonable account. We can go further and state that for one party to understand and appropriately respond to the language of another, the two networks must be at least minimally synchronised. Imagine, at the extreme, attempting a conversation with a partner who cannot see, hear, or feel you. As each of these informational channels is opened, complementarity emerges; the activity of one individual becomes driven by what the other individual processes and produces. This is the very definition of a self-organising cognitive system in which multiple subsystems become interdependent (e.g. Coey, Varlet, & Richardson, 2012; Dale & Kello, 2018; Van Orden, Holden, & Turvey, 2003). The outcome of one subsystem can no longer be determined without reference to the other subsystem – and vice versa.

The work briefly reviewed in this section indicates that language processing on the timescale of minutes involves similar patterns of interactionism as seen at smaller scales. While the mechanics underlying these patterns may differ across scales, the end result remains the same: A non-linear dynamical interaction among subsystems such that they are often best treated as a single system. Interpersonal coordination of language, posture, facial expressions, eye movements,

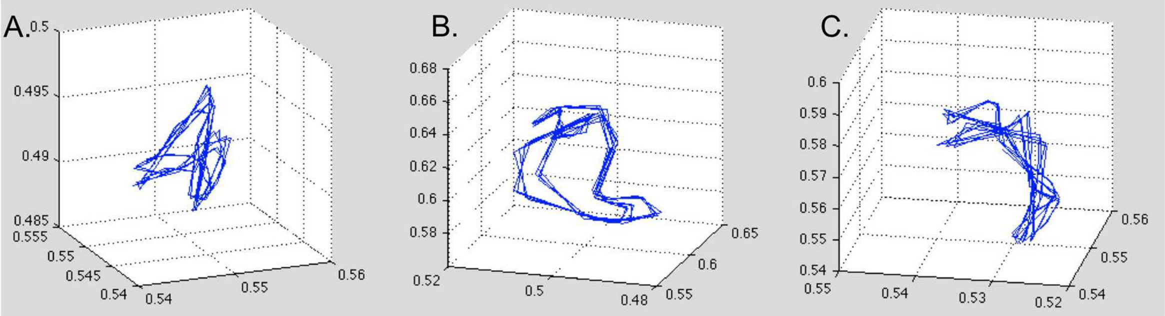

Figure 4. After connected: One hundred time steps of state-space trajectories for Network A (panel A) and Network B (panel B). The shared state space for these two nets (panel C) shows substantial coherence, with numerous repeated paths in the trajectory. (Each dimension averages activation for a third of the nodes.)

and neural activation patterns binds two parties into a single, self-organised communication network, just as separate brain regions and roughly-specialised processing modules become synchronised inside one individual.

Constraints on interactionism in language

While we have shown throughout this article that context effects can and do emerge at virtually every scale, it would be incorrect to conclude that context effects are *always* present or that there is no specialisation whatsoever in the cognitive system. However, given that context effects appear to be the rule and not the exception, it is now equally interesting to ask when such effects *do not* emerge. In examining such conditions, we can begin to understand the preconditions, flexibility, and constraints on interactionism.

If it were the case that there were no specialisation at all in language processing, then it would follow that damage to *any* region implicated in language processing would disrupt the entire system in the same way, with only degree of damage making a relevant difference. Of course, this is not the case. The classic work of Paul Broca and Carl Wernicke famously demonstrated a dissociation between speech perception and speech production. More recently, Kemmerer and Tranel (2000) found a neurological double dissociation between the use of linguistic versus visuospatial representations of locative prepositions (e.g. *in, on, above, below, etc.).* These types of evidence broadly indicate that simply because two processes *can* interact does not mean that they always must, or that their interaction is even functionally necessary (although it should be noted that a double dissociation does not necessarily entail architectural modularity in the system; Plaut, 1995; Van Orden & Kloos, 2003).

Furthermore, computational models have demonstrated that, in a purely feedforward, modular architecture, multiple modules can influence a given process without those modules influencing activity inside each other. For example, the Merge model of speech recognition (Norris, McQueen, & Cutler, 2000) demonstrated that lexical influences on phonological categorisation can occur, in principle, in the absence of feedback loops between lexical and pre-lexical activity. McQueen, Jesse, and Norris (2009) were able to argue that feedforward models best account for observed patterns of lexical influences on various pre-lexical processes – at least, before Gow and Olson's (2016) neuroimaging results. In sentence processing, results from a series of eye-tracking during reading studies by Staub (2011) were interpreted as supporting the predictions of the staged architecture of the EZ Reader 10 model, in which lexical access begins prior to syntactic integration. Moreover, McElree and Griffith (1995) used the speed-accuracy tradeoff method to show evidence that the use of thematic role information during reading may be delayed by 100 ms compared to the use of verb subcategorisation information. And Van Berkum, De Goede, Van Alphen, Mulder, and Kerstholt (2013) used ERPs to determine that, while being in a bad mood can interfere with the anticipation of semantic referential information during sentence processing, it may not interfere with syntactic processing. Findings like these clearly show that not every cognitive context can intervene with every linguistic process all the time. Different cognitive and linguistics processes have different time scales on which they operate, and thus certain delays in contextual influences are bound to be observed.

Similar arguments have been offered to explain embodied influences on language processing. Kemmerer and Gonzalez-Castillo (2010) and Kemmerer (2015) proposed a two-level cognitive architecture for verb meaning, in which one level represents the root meaning of the verb and is grounded in perceptuomotor representations, while another relatively amodal level represents the general class of verb. This proposal is supported by neuroimaging results showing dissociable activation related to each of these verb features (Kemmerer & Gonzalez-Castillo, 2010) as well as by behavioural results showing task dependence of perceptuomotor contributions to language processing (Kemmerer, 2015; Willems & Casasanto, 2011; Willems & Francken, 2012). One general conclusion reached thus far is that embodied effects in language processing depend upon the current salience of perceptual and motor features in the task (e.g. Lebois, Wilson-Mendenhall, & Barsalou, 2015; Richardson, Spivey, & Cheung, 2001). In fact, Zwaan (2014) has proposed that there may be a continuum of embodiment in language processing, such that perceptuomotor interactions with language processing are dependent upon the degree to which the context calls for detailed perceptuomotor representations.

Somewhat paradoxically, one possible conclusion of the evidence presented in this section is that even the lack of context effects is itself a type of context effect. Even a neutral context imposes its own set of biases (Steedman & Altmann, 1989). Still, it is important to temper our conclusions with the knowledge that not all processes interact with all other processes, and not under all circumstances. While it is beyond the scope of this article to fully resolve the issue of when and why synchronisation emerges or fails to emerge among sub-networks, this will be an important and interesting goal of future behavioural, computational, and neuroimaging work. For example, what specific components of task contexts drive flexibility in interactive effects? What is visible in the dynamics of subnetworks coupling and decoupling? And what are the connectivity and activation patterns that allow for synchronisation to emerge among subnetworks?

Conclusion

In this article, we have reviewed a wide range of findings regarding real-time language processing of various informational formats (e.g. phonology, semantics, syntax), at various timescales (from milliseconds to several minutes), and using various behavioural and neural methodologies. In all of this work, a clear pattern emerges: interaction and context effects are the rule, not the exception. At the briefest timescale, phonological processing has been shown to be rapidly influenced by semantic and syntactic context (Borsky et al., 1998; Miller et al., 1951), and advanced neuroimaging techniques have now conclusively demonstrated real-time interactions between neural subsystems specialised for acoustic-phonetic processing and lexical-semantic processing, respectively (Gow & Olson, 2016). At the timescale of seconds, behavioural and electrophysiological measures have demonstrated that syntactic parsing is readily influenced by semantic information (McRae et al., 1998; Trueswell et al., 1994) and wider discourse context (Altmann & Steedman, 1988; Van Berkum et al., 1999). Over the time course of several seconds, visual and embodied information can also be seen to interact with several aspects of language processing, including semantic (Yee & Sedivy, 2006), syntactic (Tanenhaus et al., 1995), lexical (Pulvermüller et al., 2005), and more general sentence comprehension (Glenberg & Kaschak, 2002; Zwaan et al., 2002). Finally, at the timescale of several minutes over which conversations take place, there can be found phonetic, linguistic, neural, and motor coordination between individuals (Dale, Fusaroli, Duran, & Richardson, 2013).

In light of the ubiquity of interaction effects, modular, stage-based accounts of language processing no longer seem tenable. When it can be seen that one subsystem is continuously causally impacted by another subsystem, the boundaries between these systems begin to blur. To more accurately capture the nature of the human language system, we have proposed a general model in which roughly-specialised subsystems are softly assembled, dynamically creating and dissolving patterns of effective connectivity (Anderson, Richardson, & Chemero, 2012; Kello & Van Orden, 2009).

As a proof of concept, we have presented some idealised network simulations which demonstrate that, by introducing a small amount of connectivity between entirely independent networks, these networks can become substantially synchronised, starting to behave as one coherent whole. By analogy to brain regions, we believe this serves as an argument that networks trained to process highly-specific patterns of information (e.g. phonetics, semantics, or syntax) may become linked directly without the need for an additional intermediating interface. Rather, when each of these systems is implemented on a common but flexible informational currency, such as simple neural activation patterns, they may communicate without a translator, and yet retain a degree of specialisation. This network analogy may easily be extended from interacting brain regions to interacting people, and even to much larger social networks. If a framework like this is adopted, the real work will be in understanding the specific dynamics of these networks embedded in networks, such as how connectivity is distributed and how subsystems dynamically couple and decouple to enable flexible interactions and context effects.

Based on the evidence reviewed here, we argue that it is problematic to make unqualified reference to "specialised modules" when in fact those modules rapidly and readily engage in dynamic interactions with other "specialised modules." Just as the interactions among phonological, syntactic, and semantic processing thwart attempts to delineate individual encapsulated modules for each of these functions, it is equally problematic to make reference to a "neural language system" given that language processing rapidly incorporates input from regions implicated in motor and visual processing. This creates a particular challenge for any researcher who accepts the arguments put forth here: Where does one focus when interaction is everywhere? Do we all need to become specialists in everything from neural networks to social networks? The proper solution to this challenge, we propose, is to understand that there is never one proper level of analysis, but rather that analysis should occur *between* levels (e.g. Spivey & Spevack, 2017). At whatever level a language researcher chooses to focus – be it speech perception, sentence processing, conversational dynamics, etc – one must consistently be aware of the interactions with "higher" and "lower" levels of analysis. And conversely, researchers should also seek to understand the limits of interactivity between their area of focus and other sources of information: How specialised is a given function, under what conditions, and when is more or less modularity beneficial in the cognitive system as a whole? For achieving these ends, we will all need to work toward a scientific process that is analogous to this interactive complex system, and even similar to our simple network simulation above. Our fields need to work toward developing strong collaborations between researchers focusing on different levels of analysis, learning to speak a bit of each other's scientific language, and achieving a resonance in the pursuit of team-based language science.

Disclosure statement

No potential conflict of interest was reported by the authors.

References

Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language*, *38*(4), 419–439.

Altmann, G. T., & Kamide, Y. (2007). The real-time mediation of visual attention by language and world knowledge: Linking anticipatory (and other) eye movements to linguistic processing. *Journal of Memory and Language*, *57*(4), 502–518.

Altmann, G., & Steedman, M. (1988). Interaction with context during human sentence processing. *Cognition*, *30*(3), 191–238.

Anderson, S. E., Huette, S., Matlock, T., & Spivey, M. J. (2010). On the temporal dynamics of negated perceptual simulations. In M. Turner, F. Parrill, & V. Tobin (Eds.), *Meaning, form, and body.* (pp. 1–20). Stanford: CSLI Publications.

Anderson, M. L., Richardson, M. J., & Chemero, A. (2012). Eroding the boundaries of cognition: Implications of embodiment. *Topics in Cognitive Science*, *4*(4), 717–730.

Arnold, J. E., Eisenband, J., Brown-Schmidt, S., & Trueswell, J. C. (2000). The rapid use of gender information: Evidence of the time course of pronoun resolution from eyetracking. *Cognition*, *76*, B13–B26.

Barsalou, L. W. (1999). Language comprehension: Archival memory or preparation for situated action? *Discourse Processes*, *28*, 61–80.

Borsky, S., Tuller, B., & Shapiro, L. P. (1998). 'How to milk a coat:' The effects of semantic and acoustic information on phoneme categorization. *The Journal of the Acoustical Society of America*, *103*, 2670–2676.

Buccino, G., Riggio, L., Melli, G., Binkofski, F., Gallese, V., & Rizzolatti, G. (2005). Listening to action-related sentences modulates the activity of the motor system: A combined TMS and behavioral study. *Cognitive Brain Research*, *24*(3), 355–363.

Buzsáki, G., & Draguhn, A. (2004). Neuronal oscillations in cortical networks. *Science*, *304*(5679), 1926–1929.

Chartrand, T. L., & Bargh, J. A. (1999). The chameleon effect: The perception–behavior link and social interaction. *Journal of Personality and Social Psychology*, *76*(6), 893–910.

Cho, P. W., Goldrick, M., & Smolensky, P. (2017). Incremental parsing in a continuous dynamical system: Sentence processing in gradient symbolic computation. *Linguistics Vanguard*, *3*(1), 20160105.

Clark, H. H. (2012). Spoken discourse and its emergence. In *Cambridge handbook of psycholinguistics* (pp. 541–557). Cambridge: Cambridge University Press.

Coey, C. A., Varlet, M., & Richardson, M. J. (2012). Coordination dynamics in a socially situated nervous system. *Frontiers in Human Neuroscience*, *6*, 164.

Connine, C. M. (1987). Constraints on interactive processes in auditory word recognition: The role of sentence context. *Journal of Memory and Language*, *26*(5), 527–538.

Cooper, R. M. (1974). The control of eye fixation by the meaning of spoken language: A new methodology for the real-time investigation of speech perception, memory, and language processing. *Cognitive Psychology*, *6*(1), 84–107.

Dale, R., Fusaroli, R., Duran, N., & Richardson, D. C. (2013). The self-organization of human interaction. *Psychology of Learning and Motivation*, *59*, 43–95.

Dale, R., & Kello, C. T. (2018). "How do humans make sense?" Multiscale dynamics and emergent meaning. *New Ideas in Psychology*, *50*, 61–72.

Dale, R., & Spivey, M. J. (2006). Unraveling the dyad: Using recurrence analysis to explore patterns of syntactic coordination between children and caregivers in conversation. *Language Learning*, *56*(3), 391–430.

Duran, N. D., Dale, R., Kello, C. T., Street, C. N., & Richardson, D. C. (2013). Exploring the movement dynamics of deception. *Frontiers in Psychology*, *4*, 140.

Elman, J. L. (1990). Finding structure in time. *Cognitive Science*, *14*(2), 179–211.

Fodor, J. A. (1983). *The modularity of mind: An essay on faculty psychology*. Cambridge, MA: MIT Press.

Forster, K. I. (1979). Levels of processing and the structure of the language processor. In W. Cooper, & E. Walker (Eds.), *Sentence processing: Psycholinguistic studies presented to Merrill Garrett* (pp. 27–85). Hillsdale, NJ: Erlbaum.

Fusaroli, R., & Tylén, K. (2016). Investigating conversational dynamics: Interactive alignment, interpersonal synergy, and collective task performance. *Cognitive Science*, *40*(1), 145–171.

Ganong, W. F. (1980). Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception and Performance*, *6*(1), 110–125.

Glenberg, A. M., & Kaschak, M. P. (2002). Grounding language in action. *Psychonomic Bulletin & Review*, *9*(3), 558–565.

Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review*, *105*(2), 251–279.

Gordon, C. L., Spivey, M. J., & Balasubramaniam, R. (2017). Corticospinal excitability during the processing of handwritten and typed words and non-words. *Neuroscience Letters*, *651*, 232–236.

GowJr., D. W., & Olson, B. B. (2016). Sentential influences on acoustic-phonetic processing: A Granger causality analysis of multimodal imaging data. *Language, Cognition and Neuroscience*, *31*(7), 841–855.

Groppe, D. M., Choi, M., Huang, T., Schilz, J., Topkins, B., Urbach, T. P., & Kutas, M. (2010). The phonemic restoration effect reveals pre-N400 effect of supportive sentence context in speech perception. *Brain Research*, *1361*, 54–66.

Grossberg, S. (2013). Adaptive resonance theory: How a brain learns to consciously attend, learn, and recognize a changing world. *Neural Networks*, *37*, 1–47.

Guediche, S., Salvata, C., & Blumstein, S. E. (2013). Temporal cortex reflects effects of sentence context on phonetic processing. *Journal of Cognitive Neuroscience*, *25*(5), 706–718.

Hauk, O., Giraud, A. L., & Clarke, A. (2017). Brain oscillations in language comprehension. *Language, Cognition and Neuroscience*, *32*, 533–535.

Iacoboni, M. (2008). The role of premotor cortex in speech perception: Evidence from fMRI and rTMS. *Journal of Physiology-Paris*, *102*(1-3), 31–34.

Jackendoff, R. (2002). *Foundations of language*. New York: Oxford University Press.

Kaiser, E., & Trueswell, J. C. (2008). Interpreting pronouns and demonstratives in Finnish: Evidence for a form-specific approach to reference resolution. *Language and Cognitive Processes*, *23*(5), 709–748.

Kello, C. T., & Van Orden, G. C. (2009). Soft-assembly of sensorimotor function. *Nonlinear Dynamics, Psychology, and Life Sciences*, *13*(1), 57–78.

Kelso, J. S. (1997). *Dynamic patterns: The self-organization of brain and behavior*. Cambridge, MA: MIT Press.

Kemmerer, D. (2015). Are the motor features of verb meanings represented in the precentral motor cortices? Yes, but within the context of a flexible, multilevel architecture for conceptual knowledge. *Psychonomic Bulletin & Review*, *22*(4), 1068–1075.

Kemmerer, D., & Gonzalez-Castillo, J. (2010). The two-level theory of verb meaning: An approach to integrating the semantics of action with the mirror neuron system. *Brain and Language*, *112*(1), 54–76.

Kemmerer, D., & Tranel, D. (2000). A double dissociation between linguistic and perceptual representations of spatial relationships. *Cognitive Neuropsychology*, *17*(5), 393–414.

Knoeferle, P., & Crocker, M. W. (2006). The coordinated interplay of scene, utterance, and world knowledge: Evidence from eye tracking. *Cognitive Science*, *30*(3), 481–529.

Knoeferle, P., Habets, B., Crocker, M. W., & Münte, T. F. (2007). Visual scenes trigger immediate syntactic reanalysis: Evidence from ERPs during situated spoken comprehension. *Cerebral Cortex*, *18*(4), 789–795.

Knoeferle, P., Urbach, T. P., & Kutas, M. (2011). Comprehending how visual context influences incremental sentence comprehension: Insights from erps and picture-sentence verification. *Psychophysiology*, *48*, 495–506.

Kuhlen, A. K., Allefeld, C., & Haynes, J. D. (2012). Content-specific coordination of listeners' to speakers' EEG during communication. *Frontiers in Human Neuroscience*, *6*, 266.

Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science*, *207* (4427), 203–205.

Lam, K. J., Bastiaansen, M. C., Dijkstra, T., & Rueschemeyer, S. A. (2017). Making sense: Motor activation and action plausibility during sentence processing. *Language, Cognition and Neuroscience*, *32*(5), 590–600.

Lebois, L. A., Wilson-Mendenhall, C. D., & Barsalou, L. W. (2015). Are automatic conceptual cores the gold standard of semantic processing? The context-dependence of spatial meaning in grounded congruency effects. *Cognitive Science*, *39*(8), 1764–1801.

Lewis, A. G., Schoffelen, J. M., Hoffmann, C., Bastiaansen, M., & Schriefers, H. (2017). Discourse-level semantic coherence influences beta oscillatory dynamics and the N400 during sentence comprehension. *Language, Cognition and Neuroscience*, *32*(5), 601–617.

Louwerse, M. M., Dale, R., Bard, E. G., & Jeuniaux, P. (2012). Behavior matching in multimodal communication is synchronized. *Cognitive Science*, *36*(8), 1404–1426.

Marslen-Wilson, W. D. (1987). Functional parallelism in spoken word-recognition. *Cognition*, *25*(1), 71–102.

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, *18*(1), 1–86.

McElree, B., & Griffith, T. (1995). Syntactic and thematic processing in sentence comprehension: Evidence for a temporal dissociation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(1), 134–157.

McMurray, B., Tanenhaus, M. K., Aslin, R. N., & Spivey, M. J. (2003). Probabilistic constraint satisfaction at the lexical/phonetic interface: Evidence for gradient effects of within-category VOT on lexical access. *Journal of Psycholinguistic Research*, *32*(1), 77–97.

McQueen, J. M., Jesse, A., & Norris, D. (2009). No lexical–prelexical feedback during speech perception or: Is it time to stop playing those Christmas tapes. *Journal of Memory and Language*, *61*(1), 1–18.

McRae, K., Spivey-Knowlton, M. J., & Tanenhaus, M. K. (1998). Modeling the influence of thematic fit (and other constraints) in on-line sentence comprehension. *Journal of Memory and Language*, *38*, 283–312.

Miller, G. A., Heise, G. A., & Lichten, W. (1951). The intelligibility of speech as a function of the context of the test materials. *Journal of Experimental Psychology*, *41*(5), 329–335.

Nieuwland, M. S., & van Berkum, J. J. (2006). When peanuts fall in love: N400 evidence for the power of discourse. *Journal of Cognitive Neuroscience*, *18*(7), 1098–1111.

Norris, D., & McQueen, J. M. (2008). Shortlist B: A Bayesian model of continuous speech recognition. *Psychological Review*, *115* (2), 357–395.

Norris, D., McQueen, J. M., & Cutler, A. (2000). Merging information in speech recognition: Feedback is never necessary. *Behavioral and Brain Sciences*, *23*(3), 299–325.

Obleser, J., & Kotz, S. A. (2010). Expectancy constraints in degraded speech modulate the language comprehension network. *Cerebral Cortex*, *20*, 633–640.

Obleser, J., Wise, R. J. S., Dresner, A. M., & Scott, S. K. (2007). Functional integration across brain regions improves speech perception under adverse listening conditions. *Journal of Neuroscience*, *27*, 2283–2289.

Onnis, L., & Spivey, M. J. (2012). Toward a new scientific visualization for the language sciences. *Information*, *3*(1), 124–150.

Pardo, J. S. (2013). Measuring phonetic convergence in speech production. *Frontiers in Psychology*, *4*, 559.

Paxton, A., & Dale, R. (2013). Argument disrupts interpersonal synchrony. *Quarterly Journal of Experimental Psychology*, *66*, 2092–2102.

Pérez, A., Carreiras, M., & Duñabeitia, J. A. (2017). Brain-to-brain entrainment: EEG interbrain synchronization while speaking and listening. *Scientific Reports*, *7*(1), 4190.

Pisoni, D. B., & Tash, J. (1974). Reaction times to comparisons within and across phonetic categories. *Perception, & Psychophysics*, *15*(2), 285–290.

Plaut, D. C. (1995). Double dissociation without modularity: Evidence from connectionist neuropsychology. *Journal of Clinical and Experimental Neuropsychology*, *17*(2), 291–321.

Pulvermüller, F., Hauk, O., Nikulin, V. V., & Ilmoniemi, R. J. (2005). Functional links between motor and language systems. *European Journal of Neuroscience*, *21*(3), 793–797.

Pulvermüller, F., Huss, M., Kherif, F., del Prado Martin, F. M., Hauk, O., & Shtyrov, Y. (2006). Motor cortex maps articulatory features of speech sounds. *Proceedings of the National Academy of Sciences*, *103*(20), 7865–7870.

Richardson, D. C., Dale, R., & Kirkham, N. Z. (2007). The art of conversation is coordination. *Psychological Science*, *18*(5), 407–413.

Richardson, D. C., Spivey, M. J., Barsalou, L. W., & McRae, K. (2003). Spatial representations activated during real-time comprehension of verbs. *Cognitive Science*, *27*(5), 767–780.

Richardson, D. C., Spivey, M. J., & Cheung, J. (2001). Motor representations in memory and mental models: Embodiment in cognition. In *Proceedings of the 23rd annual meeting of the cognitive science society* (pp. 867–872). Hillsdale, NJ: Lawrence Erlbaum Associates.

Rommers, J., Dickson, D. S., Norton, J. J., Wlotko, E. W., & Federmeier, K. D. (2017). Alpha and theta band dynamics related to sentential constraint and word expectancy. *Language, Cognition and Neuroscience*, *32*(5), 576–589.

Rumelhart, D., & McClelland, J. (1986). *Parallel distributed processing: Explorations in the microstructure of cognition, volume 1*. Cambridge, MA: MIT Press.

Sedivy, J. C., Tanenhaus, M. K., Chambers, C. G., & Carlson, G. N. (1999). Achieving incremental semantic interpretation through contextual representation. *Cognition*, *71*(2), 109–147.

Shockley, K., Santana, M. V., & Fowler, C. A. (2003). Mutual interpersonal postural constraints are involved in cooperative conversation. *Journal of Experimental Psychology: Human Perception and Performance*, *29*(2), 326–332.

Smolensky, P. (2006). Harmony in linguistic cognition. *Cognitive Science*, *30*(5), 779–801.

Smolensky, P., Goldrick, M., & Mathis, D. (2014). Optimization and quantization in gradient symbol systems: A framework for integrating the continuous and the discrete in cognition. *Cognitive Science*, *38*(6), 1102–1138.

Snedeker, J., & Trueswell, J. C. (2004). The developing constraints on parsing decisions: The role of lexical-biases and referential scenes in child and adult sentence processing. *Cognitive Psychology*, *49*(3), 238–299.

Spevack, S. C., Falandays, J. B., Batzloff, B. J., & Spivey, M. J. (2018). Interactivity of language. *Language and Linguistics Compass*, *12*(7), e12282.

Spivey, M. J., & Spevack, S. C. (2017). An inclusive account of mind across spatiotemporal scales of cognition. *Journal of Cultural Cognitive Science*, *1*(1), 25–38.

Spivey, M. J., & Tanenhaus, M. K. (1998). Syntactic ambiguity resolution in discourse: Modeling the effects of referential context and lexical frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*, 1521–1543.

Spivey-Knowlton, M. J. (1996). *Integration of visual and linguistic information: Human data and model simulations* (*Doctoral dissertation*). University of Rochester, Department of Brain and Cognitive Sciences.

Stanfield, R. A., & Zwaan, R. A. (2001). The effect of implied orientation derived from verbal context on picture recognition. *Psychological Science*, *12*(2), 153–156.

Staub, A. (2011). Word recognition and syntactic attachment in reading: Evidence for a staged architecture. *Journal of Experimental Psychology: General*, *140*(3), 407–433.

Steedman, M., & Altmann, G. (1989). Ambiguity in context: A reply. *Language and Cognitive Processes*, *4*(3/4), SI105–SI122.

Strogatz, S. (2004). *Sync: The emerging science of spontaneous order*. London: Penguin UK.

Tabor, W., & Tanenhaus, M. K. (1999). Dynamical models of sentence processing. *Cognitive Science*, *23*(4), 491–515.

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, *268*, 1632–1634.

Trueswell, J. C., Tanenhaus, M. K., & Garnsey, S. M. (1994). Semantic influences on parsing: Use of thematic role information in syntactic ambiguity resolution. *Journal of Memory and Language*, *33*, 285–318.

Van Berkum, J. J., Brown, C. M., Zwitserlood, P., Kooijman, V., & Hagoort, P. (2005). Anticipating upcoming words in discourse: Evidence from ERPs and reading times. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(3), 443–467.

Van Berkum, J. J., De Goede, D., Van Alphen, P., Mulder, E., & Kerstholt, J. H. (2013). How robust is the language architecture? The case of mood. *Frontiers in Psychology*, *4*, 505.

Van Berkum, J. J., Hagoort, P., & Brown, C. M. (1999). Semantic integration in sentences and discourse: Evidence from the N400. *Journal of Cognitive Neuroscience*, *11*(6), 657–671.

Van Orden, G. C., Holden, J. G., & Turvey, M. T. (2003). Self-organization of cognitive performance. *Journal of Experimental Psychology: General*, *132*(3), 331–350.

Van Orden, G. C., & Kloos, H. (2003). The module mistake. *Cortex*, *39*(1), 164–166.

Warren, R. M. (1970). Perceptual restoration of missing speech sounds. *Science*, *167*(3917), 392–393.

Willems, R. M., & Casasanto, D. (2011). Flexibility in embodied language understanding. *Frontiers in Psychology*, *2*, 116.

Willems, R. M., & Francken, J. C. (2012). Embodied cognition: Taking the next step. *Frontiers in Psychology*, *3*, 582.

Yee, E., & Sedivy, J. C. (2006). Eye movements to pictures reveal transient semantic activation during spoken word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*(1), 1–14.

Yoshimi, J. (2012). Active internalism and open dynamical systems. *Philosophical Psychology*, *25*(1), 1–24.

Zwaan, R. A. (2014). Embodiment and language comprehension: Reframing the discussion. *Trends in Cognitive Sciences*, *18*(5), 229–234.

Zwaan, R. A., Stanfield, R. A., & Yaxley, R. H. (2002). Language comprehenders mentally represent the shapes of objects. *Psychological Science*, *13*(2), 168–171.

Zwaan, R. A., & Taylor, L. J. (2006). Seeing, acting, understanding: Motor resonance in language comprehension. *Journal of Experimental Psychology: General*, *135*(1), 1–11.