TOPICS
TOPICS IN COGNITIVE SCIENCE

Topics in Cognitive Science 8 (2016) 761–779

ISSN:1756-8757 print / 1756-8765 online
DOI: 10.1111/tops.12219

This article is part of the topic "Memory and Common Ground Processes in Language Use," Sarah Brown-Schmidt, William S. Horton and Melissa C. Duff (Topic Editors). For a full listing of topic papers, see: http://onlinelibrary.wiley.com/doi/10.1111/tops.2016.8.issue-4/issuetoc.

# Toward Integrative Dynamic Models for Adaptive Perspective Taking

Nicholas Duran,[a] Rick Dale,[b] Alexia Galati[b,c]

[a]*School of Social and Behavioral Sciences, Arizona State University*
[b]*Cognitive and Information Sciences, University of California, Merced*
[c]*Department of Psychology, University of Cyprus*

Received 4 January 2015; received in revised form 20 July 2015; accepted 7 August 2015

**Abstract**

In a matter of mere milliseconds, conversational partners can transform their expectations about the world in a way that accords with another person's perspective. At the same time, in similar situations, the exact opposite also appears to be true. Rather than being at odds, these findings suggest that there are multiple contextual and processing constraints that may guide when and how people consider perspective. These constraints are shaped by a host of factors, including the availability of social and environmental cues, and intrinsic biases and cognitive abilities. To explain how these might be integrated in a new way forward, we turn to an adaptive account of interpersonal interaction. This account draws from basic principles of dynamical systems, principles that we argue are already expressed, both implicitly and explicitly, within a broad landscape of existing research. We then showcase an initial attempt to develop a computational framework to instantiate some of these principles. This framework, consisting of what we argue to be important mechanistic insights rendered by neural network models, is based on a promising and long-standing approach that has yet to take hold in the current domain. We argue that by bridging this gap, new insights into other theoretical accounts, such as the connections between memory and common ground information, might be revealed.

*Keywords:* Perspective taking; Interaction; Learning; Memory; Dynamical systems; Neural networks

Correspondence should be sent to Nicholas D. Duran, School of Social and Behavioral Sciences, Arizona State University, DynamiCog Lab, Glendale, AZ, 85306 E-mail: nicholas.duran@asu.edu

## 1. Introduction

Much of our day-to-day expression of thought and action occurs in rich social interaction. This was likely true in our evolutionary history, as it certainly is in modern society (Beckner et al., 2009; Clark, 1996). Humans have been interpreting others, and acting with and among others, in a way that may be constitutive of our species (Tomasello, 2008). When each of us assesses the significance of a communicative event, and seeks to contribute in kind, we reveal a non-trivial cognitive process, which depends on a number of factors. These factors range from the perception and use of immediate cues in the environment, to drawing on previous histories of social interaction to guide language use and understanding (Galati & Brennan, 2010; Gibbs & Van Orden, 2012). In many instances, these factors are processed against a reciprocal appreciation of others' needs or mental states, where simple attributions about what another knows or believes, or simple memory associations about the contents of shared knowledge, can quickly shape interaction (Brennan, Galati, & Kuhlen, 2010). This focus on others, "other-centric" or "common ground" processing, does not necessarily hold—or need to hold—in all cases. Interlocutors, at least initially, do not always consider or integrate the mental states of others. Instead, they have been shown to draw from their own knowledge and perspective, and often rely on non-social cues and heuristics to mitigate potential sources of confusion (i.e., "egocentric" processing; Keysar, Lin, & Barr, 2003).

At first blush, these findings may indicate a lack of consensus, or be perceived as a contradictory state of understanding in our field. This may be true if the assumption is that there *should* be an ego- or other-centric "default" across all interactions. But when viewed in another light, namely, that egocentric and other-centric behaviors arise from a highly adaptive cognitive system, it is not surprising to see such variation. To be adaptive, sensitivity to context is essential, and given that contexts vary in any number of ways, from the attributions that can be made, the saliency of associated cues, and the intentions to be expressed, the resulting behavior should be richly complex. This has led researchers on both sides of the issue to acknowledge that there are likely multiple strategies for when and how common ground is used (Barr, 2014; Brown-Schmidt & Hanna, 2011). It also places greater importance on the contextual constraints that are present during interaction (Schober & Brennan, 2003), as well as the attentional and memory resources that are needed to process these demands (Horton & Gerrig, 2005).

One of the challenges, however, for any account of perspective taking, is in explaining how these many constraints and processes are integrated, and why resulting behaviors are seemingly "inconsistent"—or, put differently, whether so much variation across individuals and contexts can be explained more systematically. A promising way forward, as we discuss here, is to consider how basic principles from dynamical systems theory might provide a conceptual framework for bridging ideas, and how these principles can be instantiated as models for systematic exploration.

In general, dynamical systems theory is a theoretical account of how a complex system changes over time. This process is marked by interactions among component parts, with

particular emphasis on environmental constraints, the timescales at play, and the multicausal underpinnings of stable patterns. There is no need for a "central executive" or other hardwired control process for explanatory purposes; rather, the locus of control is distributed across the many interactions present. Given these occur as embedded in a larger environment, new configurations of the system are always at the ready to meet the needs of the moment. And with no single deterministic factor dictating outcomes, multiple causes can produce similar patterns.

Extending these principles to an *adaptive account of perspective taking,* we assume that during communication, available perspective-taking information is organized across multiple timescales, both in how it is instantiated and how it is expressed during use. This information interacts in real time and within diverse social environments, and thus the relative saliency of one source of information might give way to, or perhaps enhance, the saliency of other sources of information. In addition to available information, the dynamics expressed are also influenced by learning and online processing capacities, that is, cognitive architectural constraints. As these various components vary, so too will the likelihood of ego- and other-centric behaviors.

In Sections 2 and 3 of this study, we begin by examining how dynamical principles are already expressed across existing perspective-taking studies, despite many of these studies originally designed for other purposes. The goal here is to discuss an adaptive, dynamically inspired account against the backdrop of familiar research, and critically, to draw preliminary links among a wide-ranging set of research findings. We do so in a narrative style, eschewing formal definitions, with the intention of providing a brief and accessible overview.[1]

In Section 4, we turn our attention to the potential of models for understanding perspective-taking processes. We argue that progress can be made by building models (a) that implement dynamical principles computationally; and (b) that do so by simulating existing experimental paradigms. We focus primarily on the assumption that behavior and cognition are subject to subtle variables that can radically alter the system's behavior—variables that are present in the social environment and produce outcomes that dynamically adapt in time.[2]

## 2. Cues and constraints across embedded timescales

How do social contexts influence peoples' interpretation of what their conversational partners say or do? Consider the everyday scenario of parting ways with your money. Whether buying a coffee from your local barista, or haggling over the price of a new car, unique situational demands shape how perspectives are taken and meaning understood. The same behavior in one context might be taken as a whimsical display, and in another, a cause for concern. The barista's wink is a playful gesture, but coming from the car dealer, deception. Although these interpretations may be driven by active monitoring of another's knowledge and intentions, an adaptive account also opens up the possibility that such "high-level" demands are supported by, and certainly working in concert with, a

host of additional factors that are more or less implicit and operate across a range of timescales.

For example, when conversing with that barista, you may not notice that your bodies are swaying in similar ways (Richardson, Dale, & Shockley, 2008), or that your voice rates are beginning to subtly align (Manson, Bryant, Gervais, & Kline, 2013). Meanwhile, you may engage in complementary turn-taking patterns, where your contributions are unknowingly timed at regular, oscillatory intervals (Wilson & Wilson, 2005). Even the phrases being used can become conceptually aligned without conscious attempt, reflecting a shared understanding that may not be readily understood by a non-participant listening from across the room (Mills, 2014; Schober & Clark, 1989). What is happening during this interaction is an integration of perception and action, expressed as an implicit anticipation and convergence across behavioral and linguistic channels, that all unfolds over time (Pickering & Garrod, 2013).

Although the functional consequences of such "low-level" phenomena on social and cognitive processes are a focus of ongoing research, one promising account is that it serves language comprehension and common ground processing (Richardson et al., 2008). Much like skilled dancers or improvisational musicians, language users are highly attuned to each other's understanding and perspective. Such accommodation has been argued as being central to interpersonal communication, and it has recently been described in terms of *synergistic coupling* (Fusaroli, Rączaszek-Leonardi, & Tylén, 2014). In dynamical systems parlance, synergies occur when the degrees of freedom between separate behavioral and processing systems become linked through interaction, resulting in rapid and compensatory adjustments of behavior (Riley, Richardson, Shockley, & Ramenzoni, 2011). This capacity for immediate responsiveness suggests that the efforts entailed in perspective taking are distributed across social agents, and that when disparities in understanding do exist, they can be quickly recognized and resolved in a collaborative manner (Brennan et al., 2010).

Another focus within the perspective-taking literature has been the perceptual and information-based cues that provide opportunities for social responding (Brennan et al., 2010). These include the physical characteristics and action capabilities of others, people's location and relationship with other people and objects in space, and even basic "one-bit" informational units, such as having knowledge of what another is likely to know or see (Galati & Brennan, 2010). For example, returning to our barista introduced earlier, upon detecting a foreign accent in his or her voice, you may spontaneously alter the way you speak to ensure mutual understanding (Costa, Pickering, & Sorace, 2008). Or perhaps, having been explicitly told that this person is Dutch, you mention how "The Orange almost had it in 2014," and provide clarification only when a look of confusion appears, quickly understanding that he or she is apparently not a sports fan. In another interaction, your barista might hear you ambiguously ask, as you fumble with your wallet, for "the cup" despite two identical cups being present, one in front of you and the other further away. Your barista, however, does not waver, immediately looking to the farther cup and handing it to you, assuming this is the one you meant given your limitation and spatial configuration (Hanna & Tanenhaus, 2004). Or you may even spontaneously take

the visual perspective of the barista to ease his or her understanding. Standing across from each other at a display case, you ask, "Could you grab me the biscotti on your left," despite requiring a mental rotation to do so (Duran, Dale, & Kreuz, 2011).

These findings point to a perspective-taking process that is probabilistically guided and driven by factors that are forged at longer timescales, integrating histories of social learning and situational expectations, as well as immediate demands that may arise to divert attention or tax other cognitive resources. According to our adaptive account, multiple interacting components come together at any moment to guide possible interpretations and behavior, and thus no single component will have causal priority.

These observations also open up the possibility that there are many instances of successful communication where it is uncertain whether disparities in common ground are actually acted upon and, instead, egocentric processes primarily hold. As noted previously, flexible adaptivity arises from a system that is responsive to environmental and social constraints, encompassing intrinsic biases of the system, previous histories of experience and learning, and even genetic predispositions. Because these various forces can activate both other- and egocentric responses, their competition and resolution during language use is yet another nested and interactive time course to be explored. Such dynamics suggest a mechanism of integration whereby people can *simultaneously* be other- *and* egocentric, and where, even in similar contexts, simple cues can have a huge consequence on spontaneous perspective-taking behavior (see Duran & Dale, 2014 for a detailed account).

To further explore the role of perceptual and information-based cues on perspective-taking abilities, we target next a set of existing studies that involve spatial and visual tracking during interpersonal interaction. Here, the constraints and affordances of our bodies necessitate that we invariably occupy distinct spatial viewpoints from our conversational partners. As a consequence, we have to consider spatial perspectives that are distinct from our own. How we resolve this competition requires the integration of multiple sources of information across time, and it appears to be supported by, but not entirely beholden to, how information is initially remembered.

## 3. An example domain: The integrative adaptiveness of spatial perspective taking

Across a variety of non-social tasks, it appears that people consider a number of contextual cues when selecting the perspective from which to organize spatial information in memory (e.g., McNamara, 2003). Although the viewer's egocentric viewpoint is often used as the organizing direction of spatial information (Shelton & McNamara, 2001), other contextual and environmental cues, when available, can influence the selection of that preferred orientation. These cues include the symmetry of the spatial configuration (Mou & McNamara, 2002), functional features of the constituent objects (Taylor & Tversky, 1992), and the geometry of the environment in which the configuration is embedded (Shelton & McNamara, 2001).

In collaborative tasks, social cues, such as a conversational partner's viewpoint, have also been shown to influence how spatial information is organized in memory (Galati,

Michael, Mello, Greenauer, & Avraamides, 2013), how it is described (e.g., Schober, 1993), and how it is interpreted (Duran et al., 2011). Moreover, attributions about the partner's ability to contribute to the task, including whether the partner is believed to be real (vs. simulated, Duran et al., 2011), or familiarity with the environment (Hölscher, Tenbrink, & Wiener, 2011), also influence whether an egocentric or other-centric perspective is adopted and what interpretive strategies are used.

Nevertheless, most investigations of spatial perspective have focused on the contribution of single contextual or social factors. Recent work by Galati and Avraamides (2015) has shown that perspective selection is instead guided by multiple, converging factors. These findings are compatible with our adaptive account, which predicts that multiple cues (egocentric, other centric, and contextual) will be integrated simultaneously. To demonstrate, Galati and Avraamides (2014) asked participants ("directors") to study a spatial configuration of different objects with the goal of describing it later from memory to a partner ("matcher"). This partner would then later reconstruct the configuration at his or her own workstation. At study, directors either knew the matcher's viewpoint (the partner was co-present in the room) or they did not (the partner was absent). Moreover, the participants' position was manipulated as to be aligned or not with an intrinsic configuration of the object display (objects were organized around a bilateral axis of symmetry; see Fig. 1 below). Directors were either aligned alone, matchers aligned alone (assuming they were present), or neither was aligned (see Fig. 1).

Memory tests preceding the description phase revealed that directors organized spatial relations in memory according to the convergence of cues (e.g., their own and partner's

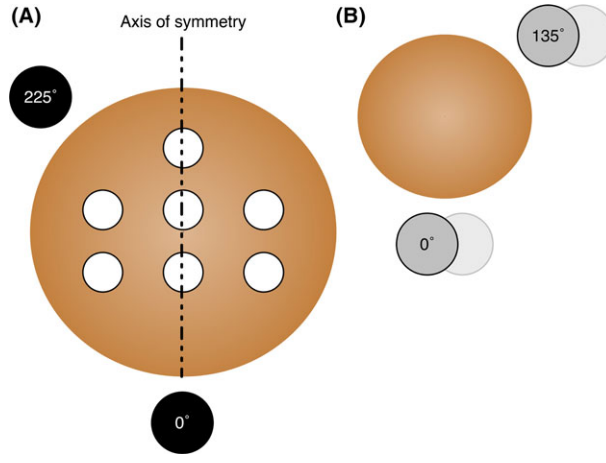


Fig. 1. (A) The setup of study phase with seven-object array organized around a bilateral axis of symmetry, based on Galati and Avraamides (2014). Director (black circles) was either at 0° or offset at 225°. (B) The matcher was either present during the study and description phases at 0° or 135° (dark gray circles), or present during description phase alone at 0° or 135° (light gray circles). (A) and (B) During the description phase (and during study when the matcher was present), director was either at 0° and the matcher at 135°, the matcher at 0° and the director at 225°, or director at 225° and matcher at 135°. *Note*: during description phase there were no objects on the table for the director.

position, visibility of partner during study, orientation of configuration). For example, when directors drew the object configuration based on memory, those who had studied the configuration while aligned with its intrinsic orientation (0°) always drew them from that viewpoint—the intrinsic orientation of the array reinforced their egocentric viewpoint as the organizing direction. For those directors who occupied a misaligned orientation (225°), the convergence of social and contextual cues now influenced the orientation of their drawings. When directors at 225° knew in advance that their partner would be aligned with the intrinsic orientation of the configuration (0°), they were more likely to organize their drawings along that canonical axis of the configuration. When directors at 225° did not know in advance their partner's viewpoint, they were more likely to use their own viewpoint as the preferred orientation of their drawings. And when they knew in advance that their partner would also be misaligned with the intrinsic orientation of the configuration (at 135°), they were equally likely to draw arrays from their own viewpoint and from the configuration's intrinsic axis (which was perhaps made more salient upon considering the oblique viewpoint of the partner).

The convergence of cues available at the description phase also predicted the perspective from which directors described the spatial configurations to their partner. When the matcher was aligned with the intrinsic orientation of the configuration, directors used more other-centric spatial expressions (e.g., to your left) than egocentric expressions (e.g., to my right), and when directors were the ones aligned with the intrinsic orientation they used more (numerically though not reliably) egocentric than other-centric expressions. Moreover, directors were able to integrate cues at the description phase, even if the relationship between these cues was not known at the time of study. For example, as already noted, directors positioned at 225°, whose partner was not present at the time of study, were more likely to organize spatial information egocentrically in memory. But when describing the configuration to a partner who was now present and positioned at 0°, the directors adopted the other's perspective while referring to the objects. This is critical because it shows directors do not simply rely on the preferred direction of their initial encoding in spatial memory but are able to flexibly adapt to changing circumstances and needs.

These findings provide compelling evidence that multiple sources of information converge and interact over time. Such integration occurs in the initial encoding of spatial organization and in subsequent communicative planning and interpretation. Rather than ascribing precedence to single social cues, egocentric biases, or environmental structure, perspective taking appears to be better captured by a process of *multicausality*, whereby multiple factors are brought together in a single moment.

## 4. Need for integration: Exemplary models

To reiterate, the preceding sections suggest that multiple cognitive processes are functioning during real-time interaction and perspective taking. Low-level and high-level cognitive processes operate together to support coherent and often informationally complex interaction. The resulting inferences—perhaps subtle and implicit, or other times explicit

and strategic—operate over an already robust egocentric frame that we employ when navigating the physical and social world. The empirical results we described above suggest that these processes are integrative and can sometimes produce very different perspectival outcomes depending on whether information is present or noticed.

But how can we further mitigate the existing debates about the primacy of ego- versus other-centric processes? How can we develop a more integrative framework? In the past we have referred to this as a kind of "centipede's dilemma" problem (Dale, Fusaroli, Duran, & Richardson, 2013). The cognitive science of interaction includes numerous specific paradigms and measures that tend to focus on particular situations or behaviors of interest. The centipede's dilemma describes the difficulty in achieving progress with a strategy of this kind. There is not yet an integrative mechanistic account that overcomes this in a more synthetic sort of analysis or modeling framework. One way forward, which we take a first step toward in this final section, is to consider the kinds of computational models that would support systematizing our understanding of interaction and perspective taking in a mechanistic framework. In sum, the preceding discussion suggests the following desiderata for a computational framework:

1 The framework should be capable of *integrating multiple simple sources of probabilistic information*.
2 It should be capable of *non-monotonic transformation*, allowing sometimes opposite outcomes from only slightly changed input.
3 Similarly, it should *nonlinearly depend on small but important fluctuations* such that one output or another may be dependent on individual task factors.
4 It should be flexible enough to explore processing and learning in a way that allows *rapid prototyping and exploration of information combination*.

In previous work we have argued that complex dynamical systems offer a suitable theoretical domain to think about the problems of interaction and perspective taking (Dale et al., 2013; Duran & Dale, 2014). However, we also noted that complex dynamical systems are still significantly limited in their ability to flexibly build extensive cognitive models to which dynamic principles can easily apply (see Dale & Duran, 2013). We argue that a fruitful way forward would be to consider parallel distributed processing (PDP) models of the theoretical sort—used traditionally as "theoretical prototype" models or existence proofs—as satisfying each of the desiderata above (see McClelland, 2009; for some discussion). Indeed, recent discussions on dynamical systems theory and PDP argue that these frameworks ought to be integrated and are at root little different from each other (Spencer, Thomas, & McClelland, 2009). It has long been known that neural networks instantiate dynamical systems of various kinds, and that differences among theorists and modelers come primarily in the form of computational or theoretical detail, such as ontological commitments over a network's input or output space, learning algorithm, and so on.

A natural next challenge is to determine the type of PDP model to develop. There are numerous possibilities, but we consider two obvious ones and develop a straightforward prototype for each to demonstrate that the above 1–4 desiderata can be easily accommodated.

### 4.1. Normalized recurrence network for real-time processing

The normalized recurrence network was devised by Spivey and Tanenhaus (1998; see also Spivey & Dale, 2004) as a means of implementing a dynamic processing network for exploring how a parallel and probabilistic system mitigates potentially divergent sources of information. It was inspired by classic TRACE connectionist models (McClelland & Elman, 1986) and was initially put to the service of investigating ambiguity resolution in sentence or word processing. It has been used recently by McMurray, Horst, Toscano, and Samuelson (2009) to study word learning and processing, and by Dale (2007) to study lexical categorization.

It is easy to devise a perspective-taking normalized recurrence system for demonstration. For example, consider Fig. 2A that is based on an experimental paradigm developed by Duran et al. (2011). In this paradigm, listeners saw identical objects on a table and received verbal instructions from a conversational partner to grab the "object on the left" or "object on the right," and to then hand it over. Sometimes speakers were oriented in such a way that they were next to the listeners (at 0°), and the perspectives of both conversational partners were aligned. In a more ambiguous situation, speakers might be oriented opposite the listeners (180°), and the instruction "grab the object on the left" can now be interpreted as meaning the speakers' left. If so, listeners would select the object on *their* right, an other-centric response. It was found that particular selections were guided by very simple social beliefs held by listeners. For example, if listeners believed speakers could not see them, they were more likely to act other centrically (presumably because speakers were issuing instructions from the only perspective available to them—their own).

Now, imagine a simple normalized recurrence system that similarly integrates input from social and task parameters to determine the output of an ego- or other-centric "left" or "right" response. In such a model, as depicted in Fig. 2B, an "integration layer" would correspond to object selection possibilities (shown in double-lined circles) and would receive numerous inputs, including information about a task partner's orientation (shaded circles), simple belief information about a social partner (square nodes), and the "verbal" instructions from a partner (single-lined circles). We can also build in egocentric biases that might be reinforced when conversational partners' orientations are aligned at 0°, or that are merely present a priori. To do so, we increase the initial activations of the links originating from the egocentric response possibilities in the integration layer.

In processing these sources of information to provide a "decision," the object integration layer receives combined activation input from the various input layers, such that for each node $i$ of each layer,

$$\text{object}_i = \text{instruction}_i + \text{orientation}_i + \text{belief}_i$$

and once integrated, the input layers themselves are updated by receiving input from the object integration layer,

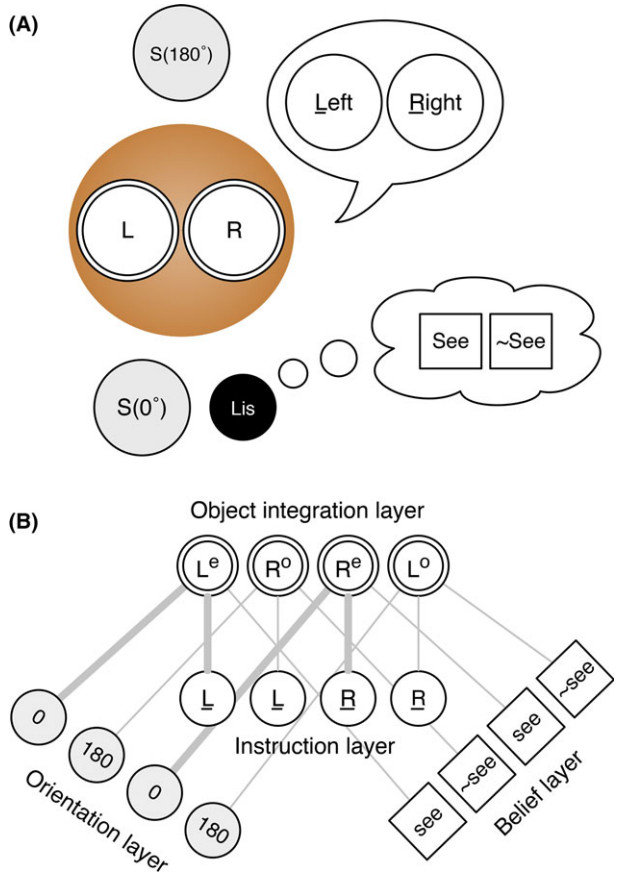


Fig. 2. Normalized recurrence network used to simulate perspective-taking behavior. (A) The architecture reflects the task arrangement in Duran et al. (2011, 2014): Two instructions (left, right), two objects (egocentrically on the left, right), and social information that can be activated (position of "speaker" ("S") at 0° or 180°, and whether speaker can see or not). *Note*: the network "listener" ("Lis") is always positioned at 0°. (B) These sources of information feed into an integration layer of the neural network architecture. This layer includes the possible selection of the right object (double-lined $R^o$, where superscript "o" corresponds to an other-centric response, and "e" corresponds to an egocentric response) when hearing the left instruction. *Note*: the thickness of the lines near 0° and egocentric left/right nodes are slightly increased to reflect egocentric biases.

$$\text{instruct}_i = \text{instruction}_i + \text{object}_i \times \text{instruction}_i$$

$$\text{orientation}_i = \text{orientation}_i + \text{object}_i \times \text{orientation}_i$$

$$\text{belief}_i = \text{belief}_i + \text{object}_i \times \text{belief}_i$$

This integrative feedback continues iteratively in a dynamic sense, where current inputs are the outputs of previous time steps, until the system stabilizes or achieves some activation threshold with one of the "left" or "right" object nodes (see Fig. 3).[3] In this

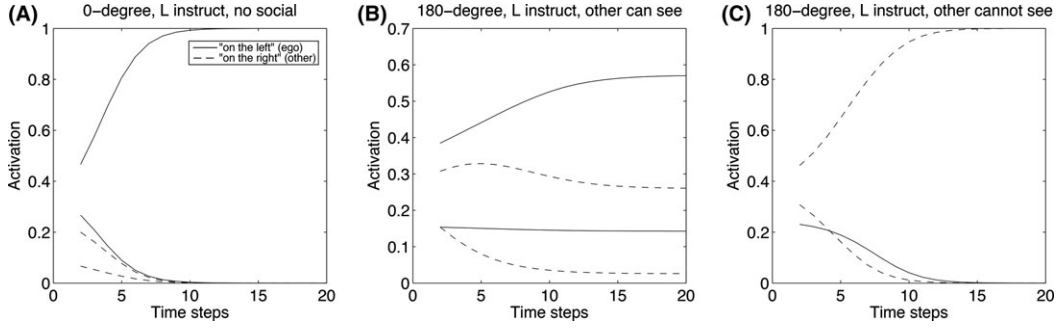


Fig. 3. Iterative activation of integration layer nodes corresponding to objects on the left (solid line) or on the right (dashed lines), where objects are "selected" based on highest activation stabilization. (A) When we activate input layer nodes corresponding to 0° orientation, left instructions, but do not activate social information, the network rapidly selects the egocentric left object (solid line). (B) If we instead activate the 180° orientation along with the left instructions, but also activate the "can see" social information, there is more competition among objects, but the egocentric left response (the left object relative to the "participant" network) is selected. (C) If we now activate the "cannot see" network, the response is precisely the opposite; there is a relatively quick response to the opposite object (right), although the dynamics are slightly more drawn out than the egocentric response seen in panel (A).

way, the network's decision is shaped by a dynamically updating activation space. Given space restrictions, we cannot present detailed specifications, but code and further descriptions are provided at www.github.com/nickduran.

With this model, we can begin to manipulate the input to demonstrate how the contributions of multiple cues interact over time, producing the same perspective-taking dynamics observed in previous experimental findings. Fig. 3A–C shows the results of some of these critical manipulations. Importantly, input layer nodes are "turned on" by giving them greater starting activation values (either 0.25 or 0.5, relative to a value of 0), akin to a flexible one-bit either/or memory instantiation (Galati & Brennan, 2010). For example, in Fig. 3B, the nodes for 180° orientation (0.5), left instructions (0.5), and "can see" (0.5) are turned on, with the model rapidly settling on the egocentric "left" object node. However, when the "cannot see" node is now turned on, a social belief previously found to facilitate other-centric responding, the dynamics of the model converge on a similar decision, selecting the other-centric "right" object (Fig. 3C). Interestingly, this decision is more drawn out as it approaches threshold, approximating greater processing costs and similar dynamics in human response movements (Duran & Dale, 2014).

### *4.2. Multilayer perceptron that performs spatial transformation and learning*

A downside of the model described in Section 4.1 is that its connections are hand coded. It abstracts over the complexity of learning and memory. To overcome some of these limitations, we seek insight from perhaps the best-known PDP framework: the multilayer perceptron that learns by error backpropagation at each time step. It has been applied to a number of domains (for a review and introduction, see McLeod, Plunkett, &

Rolls, 1998) and has famously been extended in various ways to include sequential processes (Elman, 1990). This architecture provides a highly flexible domain to combine information sources and develop task parameters.[4]

Given this flexibility, we implemented a more extensive array of task-based nodes that correspond to the full repertoire of social and environmental cues as used in Duran et al. (2011) (Fig. 4A). Doing so allows objects to occur at four unique locations, and it allows speakers' instructions to describe objects as being "above" or "below." Moreover, the speaker could also be positioned at 90° and additional social belief information is available. As shown in Fig. 4B, these cues served as additional weighted input nodes to a hidden layer. As is standard in these models, input is nonlinearly transformed and used to predict an outcome (e.g., egocentric/other-centric objects 1–4), and any error in this

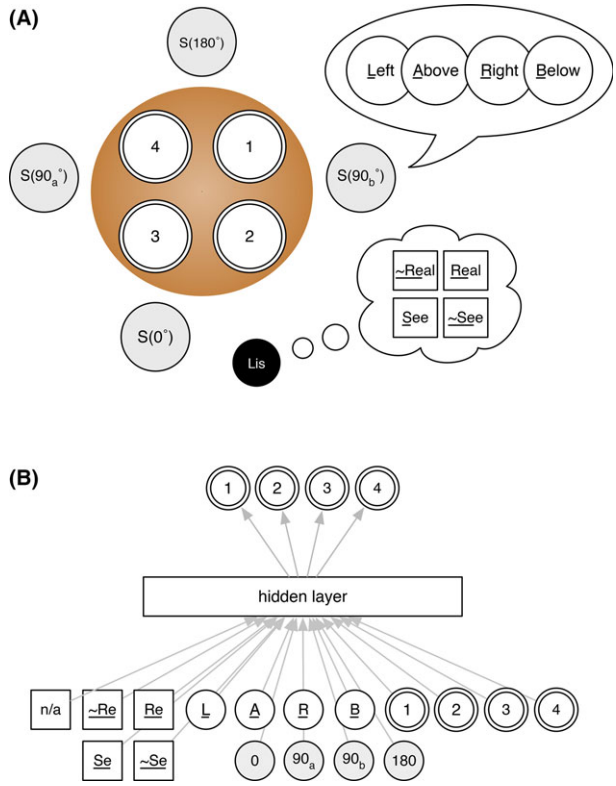


Fig. 4. The multilayer perceptron. (A) A more flexible architecture allows a wider array of task-based input nodes. In addition to those described in Section 4.1, we add a more complex configuration of object positions, which in turn allows for more complex instructions. We also add additional orientations in which the "speaker" partner ("S") can be positioned (90° to the left relative to the "listener" ("Lis"), marked with subscript "a," or to the right, marked with subscript "b"), and an additional social belief (a belief that the partner is a real or simulated agent, where believing the partner to be simulated has been shown to engender greater other-centric responding). (B) These sources of information feed into a hidden layer to predict possible outcomes (objects 1–4) where, during a training phase, errors are corrected through backpropagation and weights across links and nodes dynamically updated.

prediction is used to update connection weights via backpropogation on a trial-by-trial basis. More information can be found in the online supplementary material.

Importantly, we have to design a set of learning trials to set the network's weights for testing. For example, we can initially bias the weight space to favor egocentric responses by having the model expect egocentric objects when presented with combinations of instruction types. Such learning (reduction in error toward 0) can be seen in Fig. 5A for about the first 2,500 trials. After instilling this ego bias, we then expose the network to the alternative non-egocentric response possibility. At this point, the network has to reorganize its weight space to accommodate as the error spikes in the face of these new ambiguous trials. Following this training (about 10,000 trials), the network must then learn how to socially transform its weight space by "changing perspective" in response to social belief input. That is, when an other-centric belief such as "partner is not a real person"[5] ("~Re" node in Fig. 4B) is given, with the left instruction and $90°_a$ orientation nodes, the model should select "object 1" opposed to "object 3." Again, despite a brief but substantial spike in error, the model appears to efficiently learn over the remaining trials.

To test how the network might respond to single trials, as we did with the normalized recurrence network above, we use an approach akin to the cascade model initiated by McClelland (1979). We begin by activating a set of conditions for one trial, for example, a left instruction with 0° orientation (as shown in Fig. 5B), and pass activation one time through the network. We then use the network's output activations to update the input activations. This effectively allows the network to factor in immediate prior expectations with new activations, doing so across a weight space that has previously been shaped by learning, establishing a kind of long-term memory. An object node is eventually selected when it hits threshold stabilization.[6]

Fig. 5B shows an egocentric response convergence across 10 iterative time steps. Importantly, the network is also capable of flipping its response when we activate social

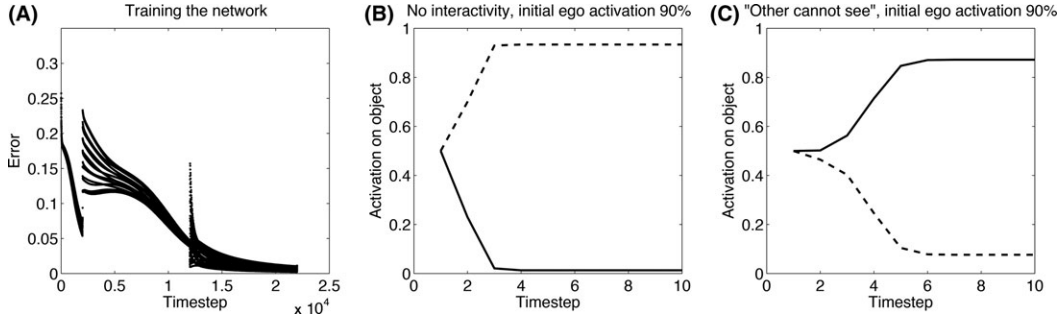


Fig. 5. Learning and response behavior for the multilayer perceptron network. (A) Error reduction in network being trained to appropriately respond to input combinations that lead to egocentric and other-centric expectations. (B) Referent corresponding to an egocentric response (dashed line) is converged upon when no social information is given and initial ego activation is set at 90%. (C) The referent corresponding to an other-centric response (solid line) is converged upon with initial ego activation still set at 90%, but social information now provided that overcomes egocentric bias.

and orientation information (e.g., partner is located at 90° and "cannot see"). The network does this even when, in both cases, initial egocentric activation is set at 90%—we give the network a strongly biased 0-degree "ego-centric" node activation, and it is still able to radically alter the output object that reaches threshold. Moreover, these other-centric responses are probabilistic as a result of the long-term training. Some networks still fall into an egocentric strategy, other networks become even more other centric. That is, in keeping with the language of the desiderata outlined above, nonlinear fluctuations across multiple sources of probabilistic information allow for non-monotonic transformation of perspective-taking outcomes.

The dynamic output from both models may permit data fitting of the kind seen in the normalized recurrence model in other studies, such as Spivey and Tanenhaus (1998). Indeed, such fixation profiles, over time, are what often distinguish ego- vs. other-centric processes in observed data (e.g., Brown-Schmidt, 2009; Wu, Barr, Gann, & Keysar, 2013; e.g., Duran et al., 2011, with computer-mouse tracking). Rigorous statistical analysis of the data often separates interpretations. The approach here would be different. Space restricts our presentation only to initial demonstrations, but we hope that flexible neural modeling would permit *generative* models, to see which constraint conditions bring about different behaviors in time. This framework would indeed permit such explorations, as we further elaborate below.

## 5. Discussion

In the preceding sections we reviewed research suggesting that perspective taking and interaction are highly adaptive, involving the integration of many diverse cues. We gave two examples of neural network models that have important properties needed to develop a mechanistic integration of perspective taking. Our models suggest that adaptive outcomes are possible through the nonlinear and simultaneous competition of multiple constraints over time. These outcomes are possible even with basic assumptions about the nature of common ground information. Here, information was available as simple visual cues or beliefs directly available from context. This framework may serve as a computational instantiation of the "one-bit" account of Brennan et al. (2010). Within the multilayer perceptron model, this information was associated with certain perspective-taking orientations established through repeated exposure and error correction, akin to typical development and learning. These associations were "remembered," in a sense, within the distributed weight space of the hidden layer, where processing constraints were also imposed by the size of the hidden layer.

Although simple, these assumptions have plausible experimental grounding and could provide modeling support for the notion that a great deal of interpersonal interaction might involve minimal burdens on cognitive processing. A major advantage of these models is that such claims can be tested by manipulating the nature of the input and parameters (task related, cognitive, or social), thus exploring these simple generalizations in a range of interactional domains. By modifying the architecture, input and output

representation space, the training regime, and so on, one could explore the role of social memories (Horton & Gerrig, 2005), executive-control processes (Brown-Schmidt, 2009), and more.

This line of theoretical development can also be pursued in a complementary fashion with other modeling approaches. For instance, recent accounts of how perspective-taking information is used have taken a "constraint-based" view that prioritizes the probabilistic weightings of available social cues (Brown-Schmidt & Hanna, 2011). The confluence of these weightings can guide hypotheses for interpretation, a view that is highly compatible with Bayesian models (Barr, 2014). What these models emphasize is thus the strength of contextual "priors" and rational combinations of which to produce "posterior" updating of perspective choice. In our account, these priors can also be implemented as initial system constraints—as was done in these preliminary modeling demonstrations. Some variant of the multilayer perceptron (MLP) model, for example, could be seen as implementing dynamic updating of posteriors in the face of new input (cf. Richard & Lippmann, 1991).

The major difference in our theoretical account, however, is the explicit focus on the dynamic process in which these choices are resolved. Understanding these interactions, in time, may reveal the mechanisms underlying reference-frame resolution, and neural network models are well suited and easily adapted for these goals. We would argue that models that emphasize time as a key unit of analysis are crucial for resolving the inconsistencies seen in the empirical data, too—the variability in ego- versus other-centric responses in a wide variety of tasks requires dynamic models to help us understand what constraints bring about one dynamic profile or another. As noted above in the simulations, eye movement data, for example, show divergent fixation profiles when ego- versus other-centric processes hold sway. Though space restricts exploring this here, dynamic models of this kind may permit a direct map onto such dynamic data (see, e.g., Duran & Dale, 2014).

### *5.1. A further challenge: Integrating memory and common ground*

The interconnections between memory and common ground information are fast becoming a central issue in understanding communicative perspective taking. Dynamical systems have sometimes avoided, or completely dismissed, the memory capacities that underlie perspective-taking abilities. In some cases this is for good reason, as perceptuomotor constraints in the environment can account for a great deal of complex behavior without resorting to internal representations (see Barrett, 2011, for an excellent introduction). But in light of the previous discussion, it seems worthy to consider the possibility that for some behaviors in some contexts, various memory representations regulate communicative behavior.

So what role does memory and common ground play in our account? We argue that, to make progress in this domain, each should be viewed as embedded in a highly adaptive cognitive/behavioral/environmental system. Their contribution to perspective taking can then be understood in terms of interactions within a larger ecology of high- and low-level constraints that are organized across multiple timescales. It is true that this account, at least in its current form, does not directly inform the representational nature of

common ground information or how it is retrieved from memory. But it does suggest that answers will be shaped by the nature of the interactional dynamics themselves, and further appreciation of the unique communicative contexts in which they occur. Insofar as the purpose of memory "representations," however conceived, is to act on a sometimes predictable, sometimes unstable, social world, their retrieval and deployment must in turn be flexible and probabilistic.

In conclusion, we hope the reader is intrigued by the promise of building more integrative dynamic models for these important and complex social processes. This line of theoretical development, we feel, might help systematize our understanding of adaptive perspective taking. Whether more egocentric in some contexts, or other centric in others, integrative models may help us understand how an adaptive and context-sensitive process can bring about both.

## Notes

1. See Richardson, Dale, and Marsh (2014) for a more formal treatment of dynamical systems within the social sciences.
2. It should be noted that cognitive scientists who adopt a dynamical perspective generally seek to understand systems as they are structured in time, although this perspective has a number of flavors. For example, some forgo representations, and wish only to see formal specification of observed behavior through mathematical models (see Chemero, 2008 for a summary), whereas others may be more inclined to adopt computational models that invoke some form of internal representation (e.g., Spivey & Dale, 2006). In all these accounts, it is generally assumed that behavior and cognition are subject to subtle variables that can radically alter the system's behavior—in other words, systems dynamically adapt in time. This is the general theoretical feature that guides the current exploration.
3. Note that the entire vector of activations, for each layer, is normalized to 0–1 before the next pass of activations.
4. Some have complained that this flexibility is a weakness as a theoretical framework (Marcus, 2001), a critique now leveled at Bayes, too (Marcus & Davis, 2013). However, any productive cognitive modeling framework has the same extreme flexibility (including classical ones). The critique is an empty one when one regards models as ever-nascent conceptual/quantitative explorations of some task/process rather than a rigid mathematical theory which, even in the "purest" case in physics, is subject to vibrant debate about excessive flexibility and philosophical implications (Smolin, 2006).
5. Greater other centricity is consistent with the principle of least collaborative effort. Because the simulated partner is unable to take perspective, the listener is instead willing to put in greater effort to do so (for greater detail, see Duran et al., 2011).
6. To avoid settling in permanently ambiguous output values, we square the output activations $\text{output}_i = \text{output}_i^2$ then renormalize $\text{output}_i = \text{output}_i / \sum_j \text{output}_j$.

## References


Barr, D. J. (2014). Perspective taking and its impostors in language use: Four patterns of deception. In T. M. Holtgraves (Ed.), *The Oxford handbook of language and social psychology* (pp. 98–110). London: Oxford University Press.

Barrett, L. (2011). *Beyond the brain: How body and environment shape animal and human minds*. Princeton, NJ: Princeton University Press.

Beckner, C., Blythe, R., Bybee, J., Christiansen, M. H., Croft, W., Ellis, N. C., Holland, J., Ke, J., Larsen-Freeman, D., & Schoenemann, T. (2009). Language is a complex adaptive system: Position paper. *Language Learning*, *59*, 1–26.

Brennan, S. E., Galati, A., & Kuhlen, A. K. (2010). Two minds, one dialog: Coordinating speaking and understanding. *Psychology of Learning and Motivation*, *53*, 301–344.

Brown-Schmidt, S. (2009). The role of executive function in perspective taking during online language comprehension. *Psychonomic Bulletin & Review*, *16*, 893–900.

Brown-Schmidt, S., & Hanna, J. E. (2011). Talking in another person's shoes: Incremental perspective-taking in language processing. *Dialogue & Discourse*, *2*, 11–33.

Chemero, A. (2008). Self-organization, writ large. *Ecological Psychology*, *20*, 257–269.

Clark, H. H. (1996). *Using language*. Cambridge, UK: Cambridge University Press.

Costa, A., Pickering, M. J., & Sorace, A. (2008). Alignment in second language dialogue. *Language and Cognitive Processes*, *23*, 528–556.

Dale, R. (2007). The relationship between decision and action: Simulating response dynamics in categorization. In D. S. McNamara & J. G. Trafton (Eds.), *Proceedings of the 29th Annual Meeting of the Cognitive Science Society* (pp. 911–916). Mahwah, NJ: Lawrence Erlbaum.

Dale, R., & Duran, N. D. (2013). Dealing with complexity differently: From interaction-dominant dynamics to theoretical plurality. *Ecological Psychology*, *25*, 248–255.

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

Duran, N. D., & Dale, R. (2014). Perspective-taking in dialogue as self-organization under social constraints. *New Ideas in Psychology*, *32*, 131–146.

Duran, N. D., Dale, R., & Kreuz, R. J. (2011). Listeners invest in an assumed other's perspective despite cognitive cost. *Cognition*, *121*, 22–40.

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

Fusaroli, R., Rączaszek-Leonardi, J., & Tylén, K. (2014). Dialog as interpersonal synergy. *New Ideas in Psychology*, *32*, 147–157.

Galati, A., & Avraamides, M. N. (2014). Social and representational cues jointly influence spatial perspective-taking. *Cognitive Science 39*, 739–765.

Galati, A., & Brennan, S. E. (2010). Attenuating information in spoken communication: For the speaker, or for the addressee? *Journal of Memory and Language*, *62*, 35–51.

Galati, A., Michael, C., Mello, C., Greenauer, N. M., & Avraamides, M. N. (2013). The conversational partner's perspective affects spatial memory and descriptions. *Journal of Memory and Language*, *68*, 140–159.

Gibbs, R. W., & Van Orden, G. (2012). Pragmatic choice in conversation. *Topics in Cognitive Science*, *4*, 7–20.

Hanna, J. E., & Tanenhaus, M. K. (2004). Pragmatic effects on reference resolution in a collaborative task: Evidence from eye movements. *Cognitive Science*, *28*, 105–115.

Hölscher, C., Tenbrink, T., & Wiener, J. M. (2011). Would you follow your own route description? Cognitive strategies in urban route planning. *Cognition*, *121*, 228–247.

Horton, W. S., & Gerrig, R. J. (2005). The impact of memory demands on audience design during language production. *Cognition*, *96*, 127–142.

Keysar, B., Lin, S., & Barr, D. J. (2003). Limits on theory of mind use in adults. *Cognition*, *89*, 25–41.

Manson, J. H., Bryant, G. A., Gervais, M. M., & Kline, M. A. (2013). Convergence of speech rate in conversation predicts cooperation. *Evolution and Human Behavior*, *34*, 419–426.

Marcus, G. (2001). *The algebraic mind: Integrating connectionism and cognitive science*. Cambridge, MA: MIT Press.

Marcus, G. F., & Davis, E. (2013). How robust are probabilistic models of higher-level cognition?. *Psychological Science*, *24*, 2351–2360.

McClelland, J. L. (1979). On the time relations of mental processes: An examination of systems of processes in cascade. *Psychological Review*, *86*, 287–330.

McClelland, J. L. (2009). The place of modeling in cognitive science. *Topics in Cognitive Science*, *1*, 11–38.

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

McLeod, P., Plunkett, K., & Rolls, E. T. (1998). *Introduction to connectionist modelling of cognitive processes*. Oxford, England: Oxford University Press.

McMurray, B., Horst, J., Toscano, J., & Samuelson, L. (2009). Towards an integration of connectionist learning and dynamical systems processing: Case studies in speech and lexical development. In J. Spencer, M. Thomas, & J. McClelland (Eds.), *Toward a unified theory of development: Connectionism and dynamic systems theory re-considered* (pp. 218–249). London: Oxford University Press.

McNamara, T. P. (2003). How are the locations of objects in the environment represented in memory? In C. Freksa, W. Brauer, C. Habel, & K. F. Wender (Eds.), *Lecture notes in artificial intelligence: Spatial cognition III* (pp. 174–191). Berlin: Springer-Verlag.

Mills, G. J. (2014). Dialogue in joint activity: Complementarity, convergence and conventionalization. *New Ideas in Psychology*, *32*, 158–173.

Mou, W., & McNamara, T. P. (2002). Intrinsic frames of reference in spatial memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*, 162.

Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences*, *36*, 329–347.

Richard, M. D., & Lippmann, R. P. (1991). Neural network classifiers estimate Bayesian a posteriori probabilities. *Neural Computation*, *3*, 461–483.

Richardson, M. J., Dale, R., & Marsh, K. L. (2014). Complex dynamical systems in social and personality psychology: Theory, modeling and Analysis. In H. T. Reis, & C. M. Judd (Eds.), *Handbook of research methods in social and personality psychology* (2nd ed). New York: Cambridge University Press.

Richardson, D. C., Dale, R., & Shockley, K. (2008). Synchrony and swing in conversation: coordination, temporal dynamics and communication. In I. Wachsmuth, M. Lenzen, & G. Knoblich (Eds.), *Embodied communication* (pp. 75–93). London: Oxford University Press.

Riley, M. A., Richardson, M. J., Shockley, K., & Ramenzoni, V. C. (2011). Interpersonal synergies. *Frontiers in Psychology*, *2*, 38.

Schober, M. F. (1993). Spatial perspective-taking in conversation. *Cognition*, *47*, 1–24.

Schober, M. F., & Brennan, S. E. (2003). Processes of interactive spoken discourse: The role of the partner. In A. Graesser, M. Gernsbacher, & S. Goldman (Eds.), *Handbook of discourse processes* (pp. 123–164). Mahwah, NJ: Lawrence Erlbaum Associates.

Schober, M. F., & Clark, H. H. (1989). Understanding by addressees and overhearers. *Cognitive Psychology*, *21*, 211–232.

Shelton, A. L., & McNamara, T. P. (2001). Visual memories from nonvisual experiences. *Psychological Science*, *12*, 343–347.

Smolin, L. (2006). *The trouble with physics: The rise of string theory, the fall of a science and what comes next*. New York: Mariner Books.

Spencer, J. P., Thomas, M. S., & McClelland, J. L. (Eds.) (2009). *Toward a unified theory of development: Connectionism and dynamic systems theory re-considered*. New York: Oxford University Press.

Spivey, M. J., & Dale, R. (2004). On the continuity of mind: Toward a dynamical account of cognition. *Psychology of Learning and Motivation*, *45*, 87–142.

Spivey, M. J., & Dale, R. (2006). Continuous dynamics in real-time cognition. *Current Directions in Psychological Science*, *15*, 207–211.

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.

Taylor, H. A., & Tversky, B. (1992). Descriptions and depictions of environments. *Memory & Cognition*, *20*, 483–496.

Tomasello, M. (2008). *Origins of human communication*. Cambridge, MA: MIT Press.

Wilson, M., & Wilson, T. P. (2005). An oscillator model of the timing of turn-taking. *Psychonomic Bulletin & Review*, *12*, 957–968.

Wu, S., Barr, D. J., Gann, T. M., & Keysar, B. (2013). How culture influences perspective taking: Differences in correction, not integration. *Frontiers in Human Neuroscience*, *7*, 822.