Chapter 12
Seeing and Believing: Social Influences on Language Processing

David W. Vinson, Rick Dale, Maryam Tabatabaeian and Nicholas D. Duran

12.1 Introduction and Motivation

Most readers have likely experienced writing peacefully at a cafe, and being suddenly foisted into an uninvited conversation with a loose acquaintance, or even a stranger, who happens to be there. Approached while intimately nested within a paper or new analysis, some of us might try carefully to tailor our language to convince the interlocutor that we are not desperately trying to end the interaction so we can get back to our work. This subtle process is all but simple. A whole range of social factors from low-level visual processes such as observing the other's gaze, to higher-level processes such as knowledge of their belief states, unfolds simultaneously, and probably mostly implicitly. Put simply, language processing in these natural situations is undergirded by many sources of information: gaze, gesture, tone of voice, lexical to syntactic levels, topics of conversation, not-so-gently-executed dialogue moves and so on. This complex array of information *simultaneously* shapes our own language towards specific goals. These variables involved in language processing *in context* do not enjoy the benefits of laboratory distillation. They together incrementally guide how we process the dialogue and how we contribute to it, as we may nevertheless try to get out of it.

D.W. Vinson · R. Dale (✉) · M. Tabatabaeian
Cognitive and Information Sciences, University of California,
5200 N. Lake Rd, Merced, CA 95343, USA
e-mail: rdale@ucmerced.edu

D.W. Vinson
e-mail: dvinson@ucmerced.edu

N.D. Duran
Arizona State University, Tempe, USA

R.K. Mishra et al. (eds.), *Attention and Vision in Language Processing*, DOI 10.1007/978-81-322-2443-3_12

This example, and many other instances of language use in context, cast language as an active and social process, intimately involving both speaker and listener (Tanenhaus and Brown-Schmidt 2008). This may be obvious to many readers, yet much research on language processing focuses on phenomena closely related to what may be termed "monologue", such as decontextualized word recognition or sentence processing. Indeed, much of this research has investigated language production and comprehension independent of one another (Pickering and Garrod 2004). From this standpoint, language happens *to* the individual at a single moment and not *between* two or more individuals over time (see Kreysa and Pickering 2011 for review). Alternatively, the past decade or two has revealed a rapidly growing trend to shift the field's focus to mechanisms involved in what might be the most basic and natural form of communication: dialogue (Clark 1992, 1996; Garrod and Pickering 2004). As Clark (1996) and colleagues have famously argued, well before this recent trend began, instances of language behaviour are difficult to explain or understand without making reference to participants in the dynamic, incremental process of dialogue. This resurgent trend looks to dialogue with an eye to understanding the *mechanisms and processes* that underlie it. The work of Clark (1996) and others certainly formed part of the bedrock of this endeavour, primarily the manner in which coordination shapes everything from lexical choice (Brennan and Clark 1996) to the organisation of overt behaviours during interaction (Clark 2005). Yet, despite this intensive work, there still remain a number of important questions about the mechanisms driving dialogue (Garrod and Pickering 2004).

Our aim is to survey this trend by looking to *social factors*, such as social gaze, and how they influence language processing in context. One impactful and inherently social process is gaze following, or attending to the spatial location visually attended to by another (Friesen and Kingstone 1998). Specifically, observing another's gaze, under the right conditions, can affect the development of common ground (Bard et al. 2007): the mutually understood and shared content between interlocutors (Clark and Marshall 1981). Yet, as in our natural language situation described above, gaze following—and visual attention in general—operate alongside a range of other cognitive processes that also influence language comprehension. Other higher-level social processes naturally bear influence. For example, Brown-Schmidt and colleagues (2009b, 2011) found that taking another's perspective impacts the ability to disambiguate linguistic referents from their visual counterparts. Although observing another's gaze and understanding another's point of view seem like disparate processes, the language processing system has the potential to be influenced by both while comprehending in context. At the same time, effective social interaction can rely heavily on visual processes. Consider even the relatively lower-level process of phoneme perception, which can be highly affected by visual content. The McGurk effect demonstrates this powerfully (McGurk and MacDonald 1976; see also Johnson et al. 1999). Low-level processes such as gaze and gaze tracking, and higher-level processes such as knowledge and beliefs about another while taking their perspective, are not independent

of each other; they mutually constrain each other during unfolding dialogue (see Brown-Schmidt and Hanna 2011 for review).

The above studies provide a brief introduction to the primary themes of this chapter: How do low- and high-level social processes impact language processing, and what role does vision play? First, we will review studies geared towards understanding how low-level social processes, such as observing another's gaze, impact language comprehension (Sect. 12.2). We then address higher-level social contexts where interlocutors must become highly attuned to the knowledge shared by their conversation partners (Sect. 12.3). This will be followed by a discussion of how many levels of processing work together during language usage in social context, using conversational deception as a case study (Sect. 12.4). We end by drawing on a dynamic approach to these processes, describing human interaction as a multilayered complex system that establishes particular strategies of operating (Dale et al. 2013; Fusaroli et al. 2013). By using this theoretical framework, we describe the future directions for mechanistic exploration (Sect. 12.5). As we argue below, the influence of such social factors is pervasive. Low-level and high-level processes, from seeing what another sees, to knowing what another knows, can sharply influence language comprehension. We begin with the importance of the eyes for language processing in context.

12.2 Seeing, and Seeing Seeing

Successful communication is sometimes mediated by the observation of another's gaze (Hanna and Brennan 2007) and by where their gaze is fixated (Gallup et al. 2012). Gaze behaviour has become a key variable in studies geared towards understanding the role of social factors in language comprehension. Readers of this volume are no doubt aware of these paradigms, many of which use eye-tracking technologies to capture gaze behaviour. Its use in research on language processing is now rather pervasive, including research on linguistically-mediated visual attention (Huettig et al. 2012), visual-world effects in language processing (Tanenhaus et al. 1995; Farmer et al. in press), disambiguation (Eberhard et al. 1995; Allopenna et al. 1998), knowledge states of an interaction partner (Brennan et al. 2010; Brown-Schmidt 2009a) and even the appearance or apparent goals of one's interlocutors (Laidlaw et al. 2011)—among many other linguistic variables (see Kreysa and Pickering 2011 for review). This massive array of findings shows that visual attention is crucial to multiple levels of language processing. At one level, the same phrase may result in entirely different gaze fixations dependent on access to specific visual content (Tanenhaus et al. 1995; cf. Cooper 1974). At the word level, gaze reveals when, in an auditory sentence stimulus, competition between potential referents occurs and when that ambiguity is resolved (Eberhard et al. 1995). Several of the chapters of this volume review this important research,

revealing the intrinsic role visual attention plays in language processing. This opens up language comprehension to a battery of possible influences, the existence of which may only be possible through vision.[1] In our review here, we focus on the impact of these findings in the contexts of natural language usage, such as dialogue.

Consider, for example, the coupling between eye-movement patterns during interaction. Research on interaction has shown that successful communication is associated with the coupling of both posture (Shockley et al. 2003) and eye movements (Richardson and Dale 2005). For example, Richardson and Dale (2005) had a speaker view a grid of images depicting characters of a popular television series (e.g., *Friends*). Recording both speech and eye movements of the speaker, they were asked to explain the relationships between the characters. Separate participants were asked to listen to one speaker's recording while also viewing the same series of images. Importantly, listeners were not given access to the speaker's eye movements. Yet, eye movements of both speaker and listener were highly coupled. More importantly, listeners whose gaze pattern was more tightly coupled, closer in time to the gaze pattern of the speaker, performed better on a comprehension test. A follow-up experiment showed that when visual attention was drawn towards the image a speaker was addressing in real time, by flashing the corresponding image, faster responses were given on a comprehension test. In a following study, Richardson et al. (2007) showed synchronous gaze coupling to occur in real-time interactions, especially if dyads shared the same common ground about their topic of conversation.

These studies suggest the coupling between gaze patterns may draw interlocutors towards useful visual information, influencing dialogue more broadly. One may argue, however, that this research simply reveals how conversational structure constrains visual attention, but not as compellingly how visual attention feeds back and can influence language processing. Strong evidence for this may come from research on language acquisition, in word learning, where studies have shown the observation of another's gaze increases later comprehension. Yu et al. (2005) had English-speaking individuals listen to a story told in Mandarin Chinese depicting a child's picture book. Participants followed along with the corresponding images while some others were presented with an additional crosshair indicating the speaker's gaze fixations. Participants who observed the crosshair performed better on a Mandarin Chinese comprehension test. This shows more

[1]A reviewer of this chapter suggested that we consider the possibility that vision presents some unique sources of information not present in other modalities. This would certainly be a controversial thesis, though there is inarguably a unique benefit to visual attention and gaze that other modalities may not have. For example, gaze may reveal the knowledge of a task partner that may only be made explicit or implicit through an overt linguistic act such as a reference (Yu et al. 2005). Gaze fixations to objects or their presence in a visual array serves as potentially "cheap but efficient" information about the task and a task partner's knowledge (Bard 2007; Brown-Schmidt 2009b). These properties may indeed give vision some unique characteristics relative to other modalities—they may have lower thresholds to achieving the shared knowledge than (say) overt speech or gesture.

directly how low-level social factors such as gaze following can directly affect comprehension. Importantly, exact timing of the occurrence of speech and eye movements was not as important as the *coupling* between speech and eye movements. As long as both auditory and visual content were experienced as coupled—manipulated via lag between speech sounds and oculomotor movements—greater comprehension occurred (Yu et al. 2005). If substantially displaced, however, gaze had no impact. Therefore, attending to the corresponding visual information while engaged in conversation may be important for effective comprehension. Accurate comprehension is highly influenced by one's own visual attention and observation of another's gaze fixations. Further, visual and auditory coupling in the process of connecting visual referents with acoustic signals may also be crucial (with children, see examples in Yu and Smith 2012).

Perhaps when considering the pervasive influence of social variables in basic cognitive processing, these results should *not* come as a surprise. It is possible that attending to the spatial region attended to by another may be a reflexive process (Friesen and Kingstone 1998; Shepherd 2010; Kuhn and Kingstone 2009). For example, Friesen and Kingstone (1998) found that reaction times decreased in response to objects presented on the screen when a cartoon face in the centre of the screen gazed towards the region where the object was to appear. Further, reaction times did not decrease when cues were presented by non-social images (e.g., a garbled face stimulus). Gaze following is inherently social and impacts visual attention more so than other types of content that might occur simultaneously within our visual field. Gaze following can be used to orient our attention towards potentially relevant information while avoiding potentially irrelevant information (Gallup et al. 2012). Importantly, what captures visual attention may be the result of our social context. This means that the shared knowledge between two or more people, common ground (Clark and Marshall 1981), may be established more implicitly, emerging out of constraints imposed by one's social context and less by the active process of explicitly deciding what information will be shared.

This process of quick integration of another's gaze may indeed help in dialogue. Gaze following can be used to disambiguate visual referents within established common ground well before clarifying linguistic information is presented. Hanna and Brennan (2007) seated participants at a table across from one another. Each half of the table contained the same objects. Though aware of having the same objects, participants unable to view the other's objects were given access to their partner's eyes. On critical trials, one participant directed the other to pick up an object similar to another object (e.g., a blue circle with five dots opposed to a blue circle with six dots). On these trials, participants used the gaze of the director to disambiguate the correct referent well before linguistic disambiguation occurred. Using the director's gaze to locate the proper referent occurred even when the order of the director's display did not match the actor's display. Clearly social information acquired through vision is used to predict what actions should be taken before clarifying linguistic information is presented. Indeed, knowing the proper referents allows participants to act in accordance with a shared common ground and even predict what linguistic information will be presented next.

Despite these intriguing results, gaze is not a social factor that is without constraint from factors such as cognitive cost. If the conditions are not right, specifically when following another's gaze is not of benefit to completing an established goal, gaze following may not occur. According to Bard et al. (2007) one assumption made by most researchers is that common ground is established only when interlocutors are able to "[model] another's knowledge while maintaining his or her own" (p. 617). Alternatively, establishing common ground may be mediated by the goals of specific individuals especially when time may *not* permit one to model another's knowledge. In one experiment, Bard and colleagues had participants navigate a map while under a time constraint. Another participant provided instructions and feedback while a crosshair on a map supposedly indicated where, on a corresponding map, the instructing participant's gaze was located. The crosshair was manipulated at a crucial point to either "look" at a relevant or irrelevant landmark. Findings show participants remained unaffected by the supposed gaze of another, indicated by the position of the crosshair, when the crosshair landed on irrelevant stimuli. This suggests that modelling another's knowledge may be limited to only that which is relevant while under a time constraint. Furthermore, when under time pressure, integrating all possible knowledge of another into one's model is costly and simply not feasible.

This may seem to throw into question the *automatic* nature of gaze following, instead supporting the notion that high-level processes such as one's goals may mediate low-level social processes such as gaze following. More likely, social gaze and what it implies about an interaction partner are rarely independent of each other, and despite the seeming simplicity of the former, the latter can be considerably more complicated. So when gaze may reflect cognitively costly social modelling, it may simply not have the same quick effects as when it merely implies spatial orientation (Friesen and Kingstone 1998).

Throughout this section, we have reviewed studies showing the impact of basic, relatively lower-level, social content on language processing. These studies show that one interlocutor's access to the other's gaze influences task competency regardless of the roles played by each participant (Bard et al. 2007; Richardson et al. 2007). In some cases low-level social information can be "ignored" depending on associated information, such as relatively more complex information like goal-oriented behaviour. But results do suggest that lower-level processes such as observing another's gaze *can strongly influence* the structuring of higher-level goals such as decision-making (Friesen and Kingstone 1998), novel language comprehension (Yu et al. 2005), task performance (Brown-Schmidt 2009b), and their outcomes. In this sense, social content may be supplanted in place of more individual goals. Many of these goals are based on social content provided by one's visual context, and the belief about others' access to the same content within their own visual context. Throughout the next section we will emphasise the influence of social content as it permeates linguistic processing at the level of maintaining specific goals.

12.3 Vision and "Visual Belief"

The previous section addressed how linguistic processes are affected by lower-level social factors and how these factors are tightly coupled to visual processes. Yet these factors were shown to influence language only in specific contexts. When attending to the location of another's gaze, recognition of the appearance of objects in that location occurs sooner (Friesen and Kingstone 1998). However, these cues may be completely ignored when they serve no purpose of completing one's goals, or it is too cognitively costly (Bard et al. 2007). In fact, it could be argued that gaze following only occurs when it fits within the constraints of one's current goals. Other research on visual attention aims to uncover what and how much visual information goes unnoticed when nested within a goal-oriented context (see Simons and Chabris, 1999 for a first look into this literature). Here we review a few social contexts that show how lower-level visual factors are constrained and, in turn, linguistic processing.

There is always the risk of being misunderstood given the inherent ambiguity of language. Despite this risk, conversational partners are able to rapidly converge on a shared understanding, where changes in how sentences are structured and ideas expressed are seamlessly adapted in the fast moving context of conversation. To do so, people must be able to make predictions about what another is likely to understand given subtle sources of social information (cf. Pickering and Garrod 2013). One source that rapidly guides language processing is being aware of another's location in space. Simply knowing another's location has powerful effects of disambiguating referents in visual space. For instance, Hanna and Tanenhaus (2004) asked a speaker to read aloud instructions to a partner on how to prepare a cake, where sometimes the speaker asked for a package of cake mix. Crucially, one package of mix was within the speaker's reach while the other was not. By tracking eye movements of the addressee, gaze fixations of the addressee revealed they immediately considered the cake mix that was outside the speaker's reach. The addressee naturally considered the speaker to be asking for the package they could not reach on their own, showing no signs of referent competition. Gaze fixations reveal that linguistic processing is constrained by the possibilities for action implied by another's spatial location. Such consideration, based on visual cues of proximity and location, is but one example of how language use is employed against a backdrop of other knowledge, in this case of what others can see and do.

There are many other studies that attest to the complexity and pervasiveness of visual information in language processing, where simple visual cues provide the basis for assessment of another's knowledge or mental states. These include adjustments in linguistic processing based on whether one appears to be a child (Newman-Norlund et al. 2009), member of the same social category (Isaacs and Clark 1987), friend or stranger (Savitsky et al. 2011), or male or female (Senay and Keysar 2009). Moreover, adjustments are extended across a number of linguistic behaviours, from grammatical choice (Balcetis and Dale 2005), to pronunciation and prosody (Kraljic et al. 2008a, b), to spatial language (Galati et al. 2012; Schober 1993).

The influence of visual information on language processes is not limited to actual visual information. Alternatively, the *mere belief* that one's partner does not share the same visual information can greatly sway language production. For example, in a study by Duran et al. (2011), participants were led to believe that a virtual communication partner did not have access to their location in a shared task space. Despite it having minimal impact on successful task completion, participants were more likely to spontaneously adopt their partner's perspective in their language use, even when doing so was cognitively more difficult. In an earlier study, Richardson et al. (2009) had pairs of participants watch a video, on separate screens, of multiple actors providing their political opinions. Participants then discussed these opinions with their partners while a still image of the actors was or was not displayed on their video screen. Critical to the task was whether participants *believed* their partner's display contained the still image of actors or not. When participants were presented with the image but believed their partner was not, they engaged in strategic message formation to compensate for the perceived mismatch in visual context. In turn, this led to greater coordination of eye movements on each screen respectively. In comparison, this coordination was not as strong when both believed the other to be looking at a blank screen. Interestingly, even when looking at a blank screen, while one participant believed the other to be viewing the image, greater coordination of eye movements occurred. Both studies reveal that a shared visual context can greatly shape communication, even when that shared visual context exists only as a belief. It is worth mentioning, though less relevant to the scope of this chapter, that beliefs about one's conversational partner also provides a basis for studying Human Computer Interaction (Branigan and Pearson 2006; Branigan et al. 2010).

One pressing theoretical issue guiding debate in this literature is *how and when* social information comes online during processing. There is strong evidence that the initial moments of sentence processing are unaffected by the mental states of others (Keysar et al. 1998, 2003). Yet a growing body of research indeed shows rapid social constraints on processing *under some conditions*. For instance, when a speaker asks for an object visible only to their partner, evidence from eye tracking shows participants initially direct their gaze to visually shared objects *and then* direct their gaze to unshared objects. Given this context, this effect implies a rapid reduction of the referential domain to the intended target (Hanna et al. 2003; Nadig and Sedivy 2002). One account for how people are able to do this so well is by treating social information as a probabilistic but simple constraint in guiding language processing (Brennan et al. 2010). When the saliency or reliability of this information is weak, such as in studies that are devoid of real social interaction or even the pretense of interaction, people may fail to act on the social information that might be available (Kronmüller and Barr 2007). In addition, when the cognitive mechanisms underlying the use of this information are also taxed—for example, when limitations on working memory are imposed—we have the tendency to revert to egocentric processing (Brown-Schmidt 2009b). Conversely, when the structure of a communicative task allows better allocation of memory resources, people more readily adapt their utterances to the beliefs of another (Horton and Gerrig 2005; recently Horton and Slaten 2012).

12.4 Where It All Counts, and the Case of Conversational Deception

We started this chapter with a description of a complex language situation, in which a variety of cognitive processes, such as visual attention, converge during one coherent linguistic performance. In a sense, this is "where it all counts", where the cognitive processes that we distill in the laboratory converge in bringing about dynamic interaction between two or more persons. So while, in some reductive sense, basic experimental work provides us with the fundamental processes that underlie language, there is a sense in which the discourse context also has a kind of epistemological priority: Understanding how these processes operate together in natural language usage is a non-trivial and extremely important endeavour (see review in Dale et al. 2013). In this final portion of our review, we take an unusual tack, *conversational deception*, as a particularly interesting communicative context that reveals the convergence of these processes. Visual processing and behavioural cues co-occur with a variety of other high-level factors as described earlier in this chapter, from situational knowledge to cognitive demands. How these cues merge during natural language contexts is non-obvious and remains wide open to future investigation, especially with regard to underlying mechanisms and dynamics.

Despite the long history of research in the field of deception, it is only recently that researchers have taken the interactive and dynamic nature of deception into account (Buller and Burgoon 1996; Burgoon 2006; Duran et al. 2013, Grammer et al. 2002; White and Burgoon 2006). It has been common practice to study this complex form of human communication in strictly information processing terms, where meaningful messages are encoded as discrete units of information and then transmitted to a receiver who must decode the message for understanding. This characterisation of communication assumes that verbal and non-verbal signals are easily observed, acknowledged as having a common meaning, and transmitted back and forth in sequential fashion (Akmajian et al. 1987; Grammer 1990; Grammer et al. 2002). Even though this approach provides an intuitive framework for carving out bounded behaviours of conversational intent, the resulting categorisation schemes are not able to capture more realistic and complex forms of communication, such as situations where peoples' true intents are meant to be obscured and hidden. Here, in the deceptive context, there is no easily discernable one-to-one correspondence between signal and intention, nor the passive reception of information. Deceivers must be vigilant of what another knows, or is likely to know, to avoid contradiction, while also maintaining the sincere appearance of conversational cooperation and believability. At the same time, the deceiver must also monitor the behavioural expressions of their partner for covert signals of suspicion, and if detected, be able to immediately adjust their own behaviours appropriately (Buller and Burgoon 2006; Duran and Dale 2012). Of course, this tightly coupled interplay of real-time reaction and action varies across the countless situational contexts in which deception can occur. It is one thing to lie about a friend's desirable attributes in a social setting where cognitive demands are minimal, but

an entirely different beast when attempting to conceal valuable information in a high-risk interrogation.

Although the goals of conversational deception may vary, what is common throughout all contexts is that the high-level factors involved have a direct impact on the low-level behaviours that shape communication. In a study by De Paulo and Bell (1996), researchers examined how possessing knowledge about another's beliefs can lead to deception, which in turn, can alter various properties of language use. Participants were instructed to discuss their ideas about paintings that they had liked or disliked with people who sometimes introduced themselves as the artist and who also expressed various levels of investment with each painting (e.g., by stating, "This is one of my favorites", vs. "This is one that I did".). When investment by the artist was potentially high, but the participant did not particularly care for the work, participants were more likely to be dishonest. In doing so, they tended to delay or avoid clear answers, provide misleading information, and use language that exaggerated their liking for the painting. Thus, the participants here were able to rapidly adjust their linguistic and communicative behaviour based on inferences about another's beliefs.

The subtle social information that must be managed when lying to another can also be in the form of ongoing actions generated by the person being lied to. These actions are not independent of how the liar responds, but are very much shaped, intentionally or not, by the liar during the interaction. In the intentional case, Burgoon and colleagues (Burgoon et al. 1999) have shown that when deceivers interact with conversational partners who do not appear to be involved with the ongoing discourse, such as by avoiding eye contact, leaning backward, or turning their bodies to distance themselves, deceivers are more likely to engage in compensatory movement behaviours to increase their partner's involvement. These particular behaviours, such as increasing proximity and the number of gaze fixations on the other's face, are also accompanied by greater verbal involvement. Importantly, the non-verbal and verbal patterns expressed here are elicited by a situational context where there is a perceived need to mitigate suspicion, a belief brought about by visually attending to the low-level changes in the partner's behaviour. When situational factors change, such as when the threat of detection is less severe, deceivers may express distinct and opposite patterns of behaviour. For example, using eye-tracking techniques, Pak and Zhou (2013) have found that deceivers fixate the faces of interlocutors less often, and that averting gaze seems to increase in frequency during the deceptive act (cf. Vrij and Semin 1996).

The above studies also raise an interesting possibility that visual information may feedback into the dialogue structure itself. Along these lines, Doherty-Sneddon et al. (1997) found that co-presence (visibility through video) modulated the entire discourse structure of an interaction. So while visual attention to non-verbal behaviours both influences and is influenced by discourse, the deceptive case suggests that this may be a more complex functional relationship when processing the language and behaviour of one's interlocutor. The cues can sometimes compete, and give way to more or less effective social evaluations depending on how they are focused upon.

The consideration of social information in deceptive communication also extends beyond the deceiver to the recipient or observer of a lie. In a study conducted by Boltz et al. (2010), participants were instructed to listen to a conversation between a man and a woman whose speech rates and response latencies were varied as they answered a number of questions. Participants were then asked to guess who was lying and when. A correlation was found between responses labelled as lies and how long it took the man or woman to start answering each question. Participants associated short and on-time latencies with honesty and mostly took long latencies as a cue for deception. But the tendency to do so depended on the gender of the speaker and their perceived motivation to lie. When the content of a response made someone else appear more favourable, and was spoken by a female, participants were more likely to selectively overlook verbal cues of deception. On the other hand, when the content of a response made the speaker look more favourable, and was spoken by a male, the verbal cues were more strictly applied. These results suggest that generalisations based on speaker attributes can dramatically alter how language is processed when assessing deceptive intent.

The very complexity of this functional relationship—how high-level discourse demands and low-level perceptual and cognitive demands interrelate in natural language use—has not enjoyed as systematic a theoretical development as is greatly needed. One way of pursuing this systematic exploration, as we describe in our concluding section, is to treat the cognitive system as multiply constrained and adaptive, more like a complex web of interdependencies, rather than a system of many independent controllers or processes.

12.5 Self-organisation of Interaction

We have offered some discussion and review of how language comprehension is fundamentally shaped by social factors. We have focused in several places on the lower-level process of visual attention and gaze as both a social factor and cognitive process that shapes the comprehension of language. We have also shown that higher-level social factors, including the knowledge and beliefs about an interlocutor, sharply impact comprehension, including back onto the process of attention itself. The resulting view of the cognitive system is one of a complex, multilayered system that involves a variety of *interdependencies*—systems that interact actively during the process of language comprehension. It seems unlikely that there is a central computational executive which is simultaneously "computing the positions and velocities" of all of these bits and pieces of human interaction (Dale et al. 2013). Instead, there must be active flows of information continually and mutually interacting with one other.

In order to make tractable this abundance of multimodal and dynamic structures, it seems fruitful to consider a process of *self-organisation* as driving complex language processes such as comprehension in context. Self-organisation is

based on the idea that a coherent performance, such as naturalistic language processing, is not "controlled centrally", but develops through a distributed process of mutual influence among the parts of a system. Such influence can cut across all levels. For example, if we learn a new fact about a conversation partner, it might shift our attention both in how we sample the visual array, but also in terms of what is to be said or interpreted. Here is where the concept of self-organisation becomes important: Two people interacting in a joint task come to form their behaviours through compensatory, complementary behaviours. These behaviours influence one another locally and incrementally, making the whole conversational performance itself a kind of self-organising "synergy" (Fusaroli et al. 2013). It is "self-organizing" in the sense that there is no one central system dictating how the interaction should unfold. Its fate is driven instead by the interdependencies among the parts as they function *together*.

Dale et al. (2013) discuss this problem as the "centipede's dilemma" of interaction research: Understanding how the various processes at play during language come to coordinate and work together. The famous children's poem by Craster has a toad ask a centipede, "Pray, which leg moves after which?" The centipede ponders this effortfully, attempting awareness of this coordination, only to find that she disrupts her very ability to move. The same happens if we do this during conversation. The cognitive mechanisms involved in a conversational performance probably outnumber a centipede's legs, especially if we counted mechanisms unavailable to conscious report. So how do we coordinate everything? The standard approach of distilling channels and exploring its behaviour piece by piece may suffer the same consequence as the centipede. This is not to say such distillation is not required, as it probably is. It is to say that recourse to language in context, and exploring the interdependencies among many channels, should also be a central part of the explanatory agenda.

In this paper, we identified some "flows of information" across cognitive processes that provide clues. When *belief about the partner* is influenced by the context of conversation, this simple piece of information may serve to highlight or amplify particular expectations or processes (Brennan et al. 2010; Brown-Schmidt 2009a, b). This relatively high-level process "sets the stage" for particular organised patterns of attention and memory at a lower level. Surely constraints of accessibility and ease of processing will influence this at all levels (Shintel and Keysar 2009), but this is not to say that the *results* of social factors do not shift the overall strategic organisation of conversational performance at a longer timescale (Duran et al. 2011). The functional timescale of conversational performance is not on the scale of "immediate or initial access", as is sometimes implied by these minimalist theories. It is instead at the timescale of hundreds of milliseconds if not seconds. At this timescale we find language comprehension processes weaved together with larger-scale pragmatic interpretation, discourse structure, social expectations and belief, and so on. We cannot consider even the lower-level processes of attention as separate from these higher-level social factors.

But how do these lower-level processes constrain each other, and act together? Akin to the centipede's dilemma, rather than understanding the interaction "leg

by leg by leg," we would suggest the idea of a "synergy" operating within and between people during coherent conversational performance. The behaviours that can mutually influence each other are quite numerous: turn-taking and rhythms, prosody or pauses, use of particular words or phrases, gesture and other bodily variables, facial expression, distribution of eye gaze, and so on. These flow into larger structures that are also numerous: adjacency pairs, topics of conversation and so on. Because these processes are often studied independently (or in small clusters), many theories tend to assume our cognitive system is composed of modules uniquely evolved or developed for each such process (Dale et al. 2013). But these can also be seen as an array of levels that are mutually constraining, and dynamically evolving, as two people come to form, in an important way, a "unit of analysis", and the interaction itself a stable, if temporary, synergy itself.

This is the notion of a synergy: a functional reduction of variability, where processes do not simply align, but can complement and compensate for each other. These different processes get coupled and constrained, moving the system into a lower-dimensional functional unit, and smaller number of stable categories—perhaps surprisingly simpler than what would be anticipated from the multidimensionality of the system itself (Shockley et al. 2009). For example, perhaps at the coarsest level of description in human interaction, one could see stable modes in the form of *arguing* (Paxton and Dale, 2013), or *flirting* (Grammer et al. 1998), or *joint decision-making* (Fusaroli and Tylén 2012), or *giving-directions* (Cassell et al. 2007). These have sometimes been referred to as "oral genres" (e.g., Busch 2007).

What is still lacking is a systematic agenda to uncover how these various processes work together to bring about *multimodal coordination* between two interacting people. We have argued in this chapter that visual attention is not independent of a range of other information sources. Social factors from gaze of another, or belief about another, can modulate the dynamics of one's attentional processes. So vision and attention are a key component, figuring into a heterogeneous assemblage of experimental techniques and observational analyses, and an associated array of diverse theoretical mechanisms that have yet to be integrated. But this array of mechanisms described above does not *merely* interact. They weave processes together into coherent interactive "structures". And it is a powerful force, much the way we experience that awkward cafe conversation that started this chapter. The great difficulty *many* of us seem to have with tailoring such short interactions is perhaps reflective of the rapid integration of diverse cognitive processes that underlie it.

References

Akmajian, A., Demers, R. A., & Harnish, R. M. (1987). *Linguistics: An introduction to language and communication* (2nd ed.). Cambridge, MA: MIT Press.

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.

Balcetis, E., & Dale, R. (2005). An exploration of social modulation of syntactic priming. In *Proceedings of the 27th Annual Meeting of the Cognitive Science Society* (pp. 184–189).

Bard, E. G., Anderson, A. H., Chen, Y., Nicholson, H. B., Havard, C., & Dalzel-Job, S. (2007). Let's you do that: Sharing the cognitive burdens of dialogue. *Journal of Memory and Language, 57*(4), 616–641.

Boltz, M. G., Dyer, R. L., & Miller, A. R. (2010). Jo are you lying to me? Temporal cues for deception. *Journal of Language and Social Psychology, 29*(4), 458–466.

Branigan, H., & Pearson, J. (2006). Alignment in human-computer interaction. In K. Fischer (Eds.), *How people talk to computers, robots, and other artificial communication partners*, (pp. 140–156).

Branigan, H. P., Pickering, M. J., Pearson, J., & McLean, J. F. (2010). Linguistic alignment between people and computers. *Journal of Pragmatics*, *42*(9), 2355–2368.

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.

Brennan, Susan E., & Clark, H. H. (1996). Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 22*(6), 1482.

Brown-Schmidt, S. (2009a). Partner-specific interpretation of maintained referential precedents during interactive dialog. *Journal of Memory and Language, 61*(2), 171–190.

Brown-Schmidt, S. (2009b). The role of executive function in perspective taking during online language comprehension. *Psychonomic Bulletin & Review, 16*(5), 893–900.

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

Buller, D. B., & Burgoon, J. K. (1996). Interpersonal deception theory. *Communication Theory, 6*, 203–242.

Burgoon, J. K. (2006). The dynamic nature of deceptive verbal communication. *Journal of Language and Social Psychology, 25*(1), 76–96.

Burgoon, J. K., Buller, D. B., White, C. H., Afifi, W., & Buslig, A. L. (1999). The role of conversational involvement in deceptive interpersonal interactions. *Personality and Social Psychology Bulletin, 25*(6), 669–686.

Busch, M. W. (2007). *Task-based pedagogical activities as oral genres: A systemic functional linguistic analysis*. ProQuest.

Cassell, J., Kopp, S., Tepper, P., Ferriman, K., & Striegnitz, K. (2007). Trading spaces: How humans and humanoids use speech and gesture to give directions. In T. Nishida (Ed.), *Conversational informatics* (pp. 133–160). New York: Wiley.

Clark, H. H. (n.d.), & C. P. Marshall. (1981). Definite reference and mutual knowledge. A. K. Joshi, B. L. Webber, & I. A. Sag (Eds.), *Elements of discourse understanding*. Cambridge: Cambridge University Press.

Clark, H H. (1992). *Arenas of language use*. Chicago: University of Chicago Press.

Clark, H. H. (1996). *Using language* (Vol. 4). Cambridge: Cambridge University Press.

Clark, H. H. (2005). Coordinating with each other in a material world. *Discourse studies, 7*(4–5), 507–525.

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. D., & Richardson, D. C. (2013). The self-organization of human interaction. In B. Ross (Ed.), *Psychology of learning and motivation* (Vol. 59, pp. 43–95). Elsevier Inc: Academic Press.

De Paulo, B. M., & Bell, K. L. (1996). Truth and investment: Lies are told to those who care. *Journal of Personality and Social Psychology, 71*(4), 703.

Doherty-Sneddon, G., Anderson, A., O'Malley, C., Langton, S., Garrod, S., & Bruce, V. (1997). Face-to-face and video-mediated communication: A comparison of dialogue structure and task performance. *Journal of Experimental Psychology: Applied, 3*(2), 105–125. doi:10.1037/1076-898X.3.2.105.

Duran, N. D. & Dale, R. (2012). Increased vigilance in monitoring others' mental states during deception. In N. Miyake, D. Peebles, & R. P. Cooper (Eds.), *Proceedings of the 34th annual conference of the cognitive science society* (pp. 1518–1523). Austin: TX: Cognitive Science Society.

Duran, N. D., & Dale, R., Kello, C., Street, C. N. H., & Richardson, D. C. (2013). Exploring the movement dynamics of deception. Frontiers in Cognitive Science. Gilbert, D. T. (1991). How mental systems believe. *American Psychologist*, *46*, 107–119.

Duran, N. D., Dale, R., & Kreuz, R. J. (2011). Listeners invest in an assumed other's perspective despite cognitive cost. *Cognition, 121*(1), 22–40.

Eberhard, K. M., Spivey-Knowlton, M. J., Sedivy, J. C., & Tanenhaus, M. K. (1995). Eye movements as a window into real-time spoken language comprehension in natural contexts. *Journal of Psycholinguistic Research, 24*(6), 409–436.

Farmer, T. A., Anderson, S. E., Freeman, J. B., & Dale, R. (in press). Coordinating motor actions and language. In P. Pyykkonen & M. Crocker (Eds.), *Visually situated language comprehension*. Amsterdam: John Benjamins.

Friesen, C. K., & Kingstone, A. (1998). The eyes have it! Reflexive orienting is triggered by nonpredictive gaze. *Psychonomic Bulletin & Review, 5*(3), 490–495.

Fusaroli, R., & Tylén, K. (2012). Carving language for social interaction: A dynamic approach. *Interaction studies, 13*, 103–123.

Fusaroli, R., Rączaszek-Leonardi, J., & Tylén, K. (2013). Dialog as interpersonal synergy. *New Ideas in Psychology, 32*, 1–11.

Galati, A., Michael, C., Mello, C., Greenauer, N. M., & Avraamides, M. N. (2012). The conversational partner's perspective affects spatial memory and descriptions. *Journal of Memory and Language*. Retrieved from http://www.sciencedirect.com/science/article/pii/S0749596X12001040.

Gallup, A. C., Hale, J. J., Sumpter, D. J., Garnier, S., Kacelnik, A., Krebs, J. R., & Couzin, I. D. (2012). Visual attention and the acquisition of information in human crowds. *Proceedings of the National Academy of Sciences, 109*(19), 7245–7250.

Garrod, S., & Pickering, M. J. (2004). Why is conversation so easy? *Trends in cognitive sciences, 8*(1), 8–11.

Grammer, K. (1990). Strangers meet: Laughter and nonverbal signs of interest in opposite-sex encounters. *Journal of Nonverbal Behavior, 14*(4), 209–236.

Grammer, K., Fink, B., & Renninger, L. (2002). Dynamic systems and inferential information processing in human communication. *Neuro Endocrinology Letters, 23*(Suppl 4), 15–22.

Grammer, K., Kruck, K. B., & Magnusson, M. S. (1998). The courtship dance: Patterns of nonverbal synchronization in opposite-sex encounters. *Journal of Nonverbal Behavior, 22*(1), 3–29.

Hanna, J. E., & Brennan, S. E. (2007). Speakers' eye gaze disambiguates referring expressions early during face-to-face conversation. *Journal of Memory and Language, 57*(4), 596–615.

Hanna, J. E., & Tanenhaus, M. K. (2004). Pragmatic effects on reference resolution in a collaborative task: Evidence from eye movements. *Cognitive Science, 28*(1), 105–115.

Hanna, J. E., Tanenhaus, M. K., & Trueswell, J. C. (2003). The effects of common ground and perspective on domains of referential interpretation. *Journal of Memory and Language, 49*(1), 43–61. doi:10.1016/S0749-596X(03)00022-6.

Horton, W. S., & Gerrig, R. J. (2005). The impact of memory demands on audience design during language production. *Cognition, 96*(2), 127–142.

Horton, W. S., & Slaten, D. G. (2012). Anticipating who will say what: The influence of speaker-specific memory associations on reference resolution. *Memory & Cognition, 40*(1), 113–126.

Huettig, F., Mishra, R. K., & Olivers, C. N. (2012). Mechanisms and representations of language-mediated visual attention. *Frontiers in Psychology, 2*, 394.

Isaacs, E. A., & Clark, H. H. (1987). References in conversation between experts and novices. *Journal of Experimental Psychology: General, 116*(1), 26–37.

Johnson, K., Strand, E. A., & D'Imperio, M. (1999). Auditory–visual integration of talker gender in vowel perception. *Journal of Phonetics, 27*(4), 359–384.

Keysar, B., Barr, D. J., & Horton, W. S. (1998). The egocentric basis of language use: Insights from a processing approach. *Current Directions in Psychological Sciences*, *7*, 46–50.

Keysar, B., Lin, S., & Barr, D. J. (2003). Limits on theory of mind use in adults. *Cognition, 89*(1), 25–41.
Kraljic, T., Brennan, S. E., & Samuel, A. G. (2008a). Accommodating variation: Dialects, idiolects, and speech processing. *Cognition, 107*(1), 54–81.
Kraljic, T., Samuel, A. G., & Brennan, S. E. (2008b). First impressions and last resorts how listeners adjust to speaker variability. *Psychological Science, 19*(4), 332–338.
Kreysa, H., & Pickering, M. J. (2011). Eye movements and dialogue. Retrieved from http://pub.uni-bielefeld.de/publication/2028210.
Kronmüller, E., & Barr, D. J. (2007). Perspective-free pragmatics: Broken precedents and the recovery-from-preemption hypothesis. *Journal of Memory and Language, 56*(3), 436–455.
Kuhn, G., & Kingstone, A. (2009). Look away! Eyes and arrows engage oculomotor responses automatically. *Attention Perception and Psychophysics, 71*, 314–327.
Laidlaw, K. E. W., Foulsham, T., Kuhn, G., & Kingstone, A. (2011). Potential social interactions are important to social attention. *Proceedings of the National Academy of Sciences, 108*(14), 5548–5553.
McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. Retrieved from http://www.nature.com/nature/journal/v264/n5588/abs/264746a0.html.
Nadig, A. S., & Sedivy, J. C. (2002). Evidence of perspective-taking constraints in children's online reference resolution. *Psychological Science, 13*(4), 329–336.
Newman-Norlund, S. E., Noordzij, M. L., Newman-Norlund, R. D., Volman, I. A., de Ruiter, J. P., Hagoort, P., & Toni, I. (2009). Recipient design in tacit communication. *Cognition, 111*(1), 46–54.
Pak, J., & Zhou, L. (2013). Eye gazing behaviors in online deception. *AMCIS 2013 Proceedings*. Retrieved from http://aisel.aisnet.org/amcis2013/ISSecurity/RoundTablePresentations/3.
Paxton, A., & Dale, R. (2013). Argument disrupts interpersonal synchrony. *Quarterly Journal of Experimental Psychology*, *66* (11), 2092–2102.
Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences, 27*(2), 169–189.
Pickering, M. J., & Garrod, S. (2013). Forward models and their implications for production, comprehension, and dialogue. *Behavioral and Brain Sciences*, *78*, 49–64.
Richardson, D. C., & Dale, R. (2005). Looking to understand: The coupling between speakers' and listeners' eye movements and its relationship to discourse comprehension. *Cognitive Science, 29*(6), 1045–1060.
Richardson, D. C., Dale, R., & Kirkham, N. Z. (2007). The art of conversation is coordination common ground and the coupling of eye movements during dialogue. *Psychological Science, 18*(5), 407–413.
Richardson, D. C., Dale, R., & Tomlinson, J. M. (2009). Conversation, gaze coordination, and beliefs about visual context. *Cognitive Science, 33*(8), 1468–1482.
Savitsky, K., Keysar, B., Epley, N., Carter, T., & Swanson, A. (2011). The closeness- communication bias: Increased egocentrism among friends versus strangers. *Journal of Experimental Social Psychology, 47*, 269–277.
Schober, M. F. (1993). Spatial perspective-taking in conversation. *Cognition, 47*(1), 1–24.
Senay, I., & Keysar, B. (2009). Keeping track of speaker's perspective: The role of social identity. *Discourse Processes, 46*(5), 401–425.
Shepherd, S. V. (2010). Following gaze: Gaze-following behavior as a window into social cognition. *Frontiers in Integrative Neuroscience*, *4*. Retrieved from http://www.ncbi.nlm.nih.gov/pmc/articles/pmc2859805/.
Shintel, H., & Keysar, B. (2009). Less is more: A minimalist account of joint action in communication. *Topics in Cognitive Science, 35*(2), 281–322.
Shockley, K., Richardson, D. C., & Dale, R. (2009). Conversation and coordinative structures. *Topics in Cognitive Science, 1*(2), 305–319.

Shockley, Kevin, 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.

Simons, D. J., & Chabris, C. F. (1999). Gorillas in our midst: Sustained inattentional blindness for dynamic events. *Perception*, *28*(9), 1059–1074.

Tanenhaus, M. K., & Brown-Schmidt, S. (2008). Language processing in the natural world. *Philosophical Transactions of the Royal Society B: Biological Sciences, 363*(1493), 1105.

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*(5217), 1632–1634.

Vrij, A., & Semin, G. R. (1996). Lie experts' beliefs about nonverbal indicators of deception. *Journal of Nonverbal Behavior, 20*(1), 65–80.

White, C., & Burgoon, J. (2006). Adaptation and communicative design. *Human Communication Research, 27*(1), 9–37.

Yu, C., Ballard, D. H., & Aslin, R. N. (2005). The role of embodied intention in early lexical acquisition. *Cognitive Science, 29*(6), 961–1005.

Yu, C., & Smith, L. B. (2012). Embodied attention and word learning by toddlers. *Cognition, 125*(2), 244–262. doi:10.1016/j.cognition.2012.06.016.