CellPress

Opinion

Conceptual Alignment: How Brains Achieve Mutual Understanding

Arjen Stolk,[1,*] Lennart Verhagen,[2] and Ivan Toni[3]

We share our thoughts with other minds, but we do not understand how. Having a common language certainly helps, but infants' and tourists' communicative success clearly illustrates that sharing thoughts does not require signals with a pre-assigned meaning. In fact, human communicators jointly build a fleeting conceptual space in which signals are a means to seek and provide evidence for mutual understanding. Recent work has started to capture the neural mechanisms supporting those fleeting conceptual alignments. The evidence suggests that communicators and addressees achieve mutual understanding by using the same computational procedures, implemented in the same neuronal substrate, and operating over temporal scales independent from the signals' occurrences.

Why Doesn't My Phone Understand Me, Yet?

Human communication is often framed in terms of signal transmission [1,2]. This framework presupposes that communicators already share the same set of coding–decoding rules (e.g., a common language, body emblems). That intuition seems plausible until we try to build artificial cognitive agents that can deal with human communicators beyond the boundaries of precisely choreographed interactions (http://www.newyorker.com/tech/elements/why-cant-my-computer-understand-me). Consider human interactions with common artificial agents such as Apple's Siri, Microsoft's Cortana, or Google Now. Dramatic increases in computing power, data set availability, and sophistication of machine-learning algorithms have made those artificial agents extremely useful within our daily lives. However, we would question the understanding of an English-speaking interlocutor that, in all seriousness, guides us to the nearest casino after being told 'I have a gambling problem', as Siri does. Even an embodied humanoid agent designed to physically and communicatively interact with us fails to consider that someone might want to raise their hand for reasons other than asking a question (http://www.bbc.com/news/technology-23196867 and Movie S1 in the supplemental information online). Is there a fundamental reason for those communicative failures? Robotics has long struggled with perceptual and sensorimotor restrictions, but modern artificial cognitive agents have access to powerful feature detectors and associative procedures such as hierarchically organized convolutional neural networks and reinforcement learning algorithms [3–5]. Those algorithms can reliably extract and categorize a signal's features, given the contextual background of statistical regularities present in a set of training exemplars, and link them to adaptive action selection procedures [4,6]. Here we argue that adding more feature detectors, sensorimotor associations, or processing speed is unlikely to solve those communicative failures. Indeed, those communicative failures might be engineered, in the sense that those artificial agents have been focused on information transfer (in Shannon's sense [1]) rather than on the computational problem solved in human communication (in Marr's sense [7]). Building on previous suggestions [8–14], we argue

Trends

State-of-the-art artificial agents such as the virtual assistants on our phones are powered by associative deep-learning algorithms. Yet those agents often make communicative errors that, if made by real people, would lead us to question their mental capacities.

We argue that these communicative errors are a consequence of focusing on the statistics of the signals we use to understand each other during communicative interactions.

Recent empirical work aimed at understanding our communicative abilities is showing that human communicators share concepts, not signals.

The evidence shows that communicators and addressees achieve mutual understanding by using the same computational procedures, implemented in the same neuronal substrate, and operating over temporal scales independent from the signals' occurrences.

[1]Helen Wills Neuroscience Institute, University of California, Berkeley, CA 94720, USA
[2]Department of Experimental Psychology, University of Oxford, Oxford OX1 3UD, UK
[3]Donders Institute for Brain, Cognition, and Behaviour, Radboud University, 6500 HB Nijmegen, The Netherlands

*Correspondence: astolk@berkeley.edu (A. Stolk).

Trends in Cognitive Sciences

CellPress

that creating a shared conceptual space across communicators might be a more fruitful characterization of the computational problem solved in human communication.

Considering human communication as a signal coding–decoding problem continues to entice empirically oriented scholarly fields. For instance, many communication studies in cognitive neuroscience build on the assumption that retrieval and selection of a signal's meaning is bound to and triggered by the signal occurrence [2,15–20]. Priming, neural synchrony, and shared sensorimotor associations are among the mechanisms that have been suggested to implement information transfer [2,15–20]. Here we discuss a mechanism geared to implement a different computational function of human communication. This mechanism has emerged because it has become feasible to empirically interrogate, at the implementation level, longstanding theoretical perspectives of human communication [8–14]. We review evidence showing that the main load of human communication is carried by a neuronal mechanism continuously aligning our conceptual structures with those of another agent. We argue that this dynamic alignment provides a conceptual frame necessary for interpreting intrinsically ambiguous communicative signals. We discuss neuronal mechanisms supporting the integration of a communicative signal within those dynamically adjusted conceptual structures.

Dynamic Conceptual Alignments

Across multiple generations, natural selection can drive organisms toward shared coding–decoding rules [21,22], and artificial agents can share novel symbols by trial and error [23–25]. However, establishing those symbols without the presence of pre-existing common knowledge requires artificial agents to use thousands of pair-wise interactions [23] or several iterations across independent populations of communicators [22]. This learning dynamic is not remotely related to how humans converge on a shared meaning and disambiguate situations lacking predefined coding–decoding schemes, an ability that often operates on the basis of a single trial [26–30]. Rapid resolution of novel or ambiguous symbols is the daily bread and butter of human communication and is crucial for understanding the problem solved during **mutual understanding** (see Glossary) [31]. Dealing with ambiguities by converging on shared meanings is not an exceptional situation that requires exceptional learning times. We promptly resolve multiple communicative ambiguities in everyday conversations, as well as when we learn a language as an infant [10,32]. Even commonly used words do not contain fixed meanings – they may provide us with clues to a communicative meaning [33–35] – but are coordinated through an interactive process by which people seek and provide evidence that they understand one another [36,37]. Accordingly, the notion of 'communicative meaning' encompasses both the ability to recognize the communicator's 'communicative intention' and the content of the behavior manifesting that intention ('informative intention' [9]). These naturalistic observations are supported by recent empirical observations on how human communicators colonize a semiotically virgin territory [38,39]. If the computational problem of communicators is to align their conceptual spaces, signals should differentiate contingently to the local interactional conditions, leading to rapid semiotic diversity across communicative groups. Conversely, if human communication is about optimizing signal coding–decoding [1,15,16] equal communicative demands and initial conditions across communicators (e.g., background knowledge) should lead to similar signals across communicative groups. Empirical and natural observations clearly favor the former scenario [26,29,39–43] (Box 1).

Given the multiple semantic ambiguities present in an everyday utterance, how can we quickly and reliably identify signals adequate to focus the mind of an addressee on our communicative intention or infer the communicative intention suggested by our interlocutor? For instance, a customer might solicit a drink in a bar with a propositionally inconsequential and semantically ambiguous statement ('A cup of Joe'; Figure 1A). A bartender might reply with a logically unrelated and similarly ambiguous statement ('We're closing in five'). Yet, an eavesdropper at the

Glossary

Conceptual alignment: condition in which individuals' mental representations have become aligned, or sufficiently compatible, despite those individuals' idiosyncratic experiences and knowledge structures. People can, for instance, refer to the same coffee cup in a kitchen although they may have different (visual) perspectives or background knowledge of the object. The alignment is conceptual because it occurs at a level abstracted away from actual experience, such that the same people may refer to the same object in a different moment or situation (see mutual understanding).

Mutual understanding: when different minds mutually infer they agree on an understanding of an object, person, place, event, or idea (see conceptual alignment).
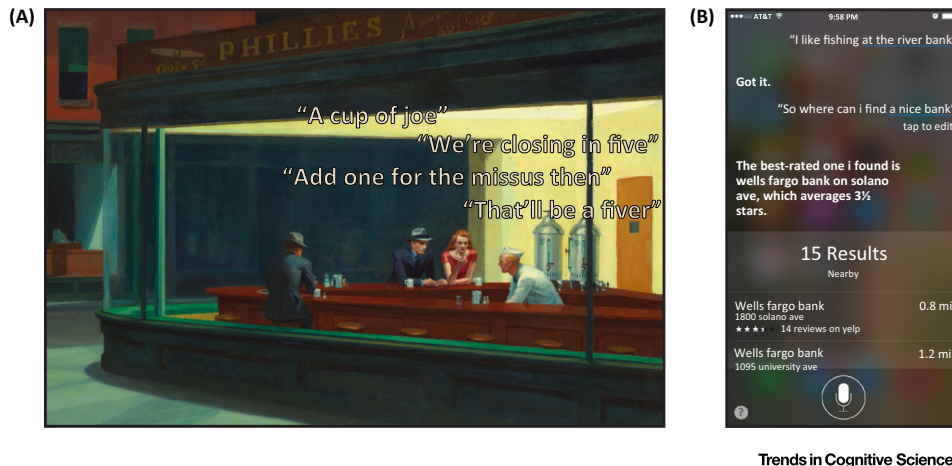
Trends in Cognitive Sciences

CellPress

Figure 1. Prototypical Human Communicative Interaction with Another Human (A), and with an Artificial Agent (B). Artwork courtesy of the Art Institute of Chicago. Screenshot of Apple Siri interaction.

counter would not be surprised when the interlocutors quickly resolve the pragmatic implications of those semantic ambiguities. The customer would also have no difficulty in interpreting the bartender's reply ('That'll be a fiver') in the context of his opening statement, rather than limiting its search for a relevant context to temporally closer or arithmetically more consistent events (e.g., 'Add one...'; Figure 1A) in the turn-taking sequence.

This vignette illustrates that, in human communication, an action can be a response to a signal occurring at any time along the interaction's trajectory, irrespective of linear order or syntactic regularities. These hierarchically embedded and temporally irregular conceptual dependencies between utterances cannot be easily resolved by neural and computational processes exclusively focused on statistically predominant features within signal sequences [2,44]; that is, the regularities captured by convolutional neural networks and reinforcement learning algorithms [3].

Box 1. Capturing Communication in the Laboratory

Understanding how humans communicate requires experimental protocols that capture mutual understanding. Yet, paradoxically, cognitive neuroscientists have often approached the study of human communication by minimizing mutual understanding; for example, by studying an individual agent producing scripted utterances or processing isolated sentences [80–82]. Recently, empirical studies have started to pay attention to the communicative context in which those signals are embedded. One approach particularly effective for capturing communicators' shared communicative history involves people communicating in a novel medium [24,26,29,39,42,95,96]. The novel medium minimizes participants' access to a number of pre-existing conventions exploited in everyday communication (e.g., a common language, body emblems, facial expressions). Consequently, the generation and comprehension of communicative signals become strongly conditional on the conceptual space idiosyncratically defined by the ongoing interaction. For instance, when pairs of players are asked to communicate by moving geometric shapes on a digital board (Figure IA), the same signal can be used by different pairs to coordinate different meanings. The same signal can even have different meanings in different trials of the same pair and vice versa (for examples see movies in [29] and [72]). There are no *a priori* correct solutions to this communicative task nor a limited set of options from which the players can choose. Several pieces of evidence indicate that the players jointly and dynamically establish an agreement (also known as a 'conceptual pact' [97]) on the meaning of a signal. For instance, changes in the communicator's movement characteristics after a misinterpretation of the addressee are dependent on the nature of the 'error' made by the addressee [98], suggesting that communicators take into account how addressees interpret their signals and adjust them accordingly. Thus, as in everyday dialogue, effective communication in this game arises only when players align their conceptual solutions. Yet, communicative difficulty and communicators' shared cognitive history can fairly easily be manipulated by varying the complexity of the spatial goal configurations (for examples see [98]) and having pairs encounter problems for which they previously have established a joint solution (see [72]), respectively. Accordingly, this computer game offers an experimental platform with the possibility to manipulate and quantify communicators' fleeting conceptual alignments over a series of interactions and isolate the fundamental mechanisms of human social interaction. See Figure IB for a comparison of the communication game with everyday dialogue.
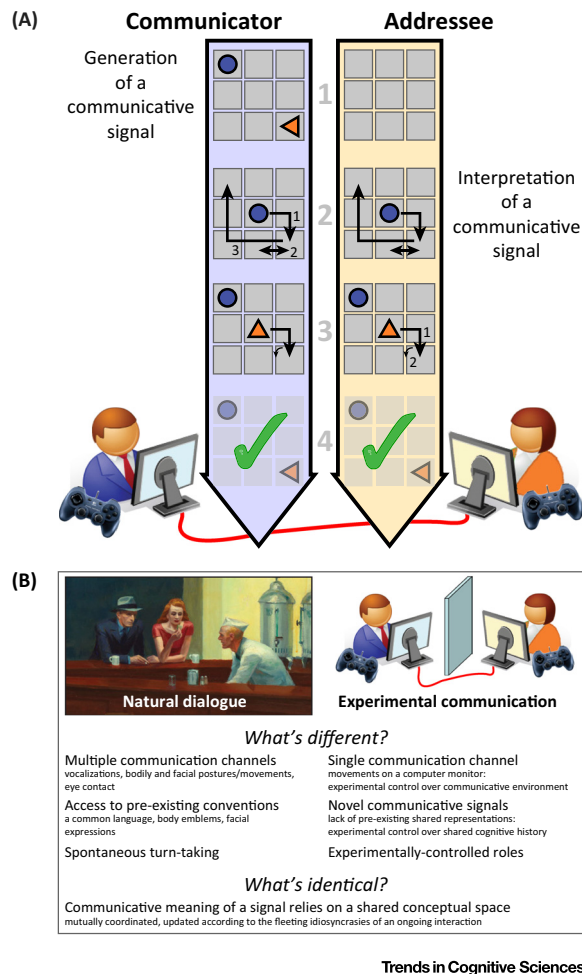
Trends in Cognitive Sciences

Figure I. Experimentally Controlled Communicative Interaction. (A) The two-player game setup is computer programmed and presented individually on two separate monitors. The players control their shape movements (horizontal and vertical translations and 90° rotations) using hand-held controllers. In this example the player in blue, labeled as the 'Communicator', tries to make clear to the player in orange, labeled as the 'Addressee', that her triangular token should be positioned in the bottom right square, pointing left. At onset of each interaction a shape is assigned to each player, followed by presentation of the goal configuration to the Communicator (epoch 1). During this epoch he can plan for as long as needed, but he has only 5 s to execute his movements in the next. After pressing a start button, the Communicator's shape will appear in the center of the grid. He now can execute his actions, visible to the Addressee who needs to infer the Communicator's intentions from his movements (epoch 2); for instance, by first going to her target location, ostensibly 'pause' to indicate the relevance of that location (number 1 action), then 'wiggle' to indicate her shape's orientation (number 2 action), and then completing his own target configuration (number 3 action). Please note that this is only one among a series of possible solutions. For instance, some participants converge on using the number of subsequent 'wiggles' to mark the number of clockwise rotations that the Addressee needs to make to achieve the target orientation of her shape, while others do not use the 'wiggle' but leave the triangle location along the direction to which the triangle needs to point. Following the Communicator's movements, the Addressee can plan and execute her actions to complete the goal configuration (epoch 3). Finally, feedback on their communicative success is presented to both players in the form of a green tick or a red cross (epoch 4). Adapted from [29]. (B) Comparison of the communication game with everyday social interaction.

The vignette also illustrates how communication requires more than pruning a decision tree of possible options stored in memory [45,46]. The concepts used in everyday interaction are not well-defined structures with definite values [47]. Indeed, a communicator often needs to generate novel candidate conceptualizations of a signal, as when a customer would hear 'fiver' for the first time.

CellPress

Resting on theoretical accounts that have highlighted the social implicatures intrinsic to a given communicative signal [8,9,11], we propose that the customer generates a set of possible-world scenarios based on his shared conceptual space with the bartender and plans how to probe his inference to the best explanation in his next utterance or move. This generative ability would require a mechanism for finding consistent relationships between abstract features of communicative signals, a mechanism that spins a conceptual web across potentially unrelated elements of the communicative interaction (e.g., a link between 'Joe', hypotheses about 'fiver', a large body of background knowledge about bartenders) [14,37,48–50]. This mechanism would kick in even before the first signal in the vignette is produced; for example, when the customer assumes that the person behind the bar is familiar with the intended meaning of 'Joe' and is willing to sell drinks. Several possible-world contexts might need to be prepared to achieve the flexibility characteristic of human interactions (e.g., the bar might be already closed, the presumed bartender is a cleaner). Thus, humans need to be able to efficiently explore large search spaces and establish connections between different conceptual structures [29]. Although it remains unclear how to make this operation computationally tractable [5,51,52], similar exploration-exploitation trade-offs might be used during domain-general foraging decisions in unstable environments [45,53–55]. As the interaction unfolds, communicators continuously update their conceptual spaces, keeping their mutually inferred thoughts aligned to the current situation and to each other. This alignment process builds on semantic structures operating over multiple timescales, from the fleeting idiosyncrasies of an ongoing communicative interaction to long-term semantic memories and regularities acquired throughout development [56].

The **conceptual alignment** account can be contrasted with signal-centered frameworks of communication in which retrieval and selection of a signal's meaning is bound to and triggered by the signal occurrence [1,2,15–20]. For instance, the scripted responses of current artificial agents might be informed by long-term statistical regularities but fail to consider the importance of the ongoing communicative dynamics for building mutual understanding (Figure 1B). This fundamental aspect of human communication can be easily overlooked when the focus is on the signal itself rather than on the conceptual space evolving between communicators. In the conceptual alignment framework, meaning is not a property of a signal but a property of a mutually inferred conceptual space in which signals are merely a means to probe and bias that conceptual space (Figure 2, Key Figure). By embedding signals in a conceptual space defined by the ongoing interaction, communicators can flexibly infer the meanings of those signals and overcome the 'curse of dimensionality' (i.e., a combinatorial explosion with a number of features) intrinsic in signal-centered approaches to human communicative situations [2,12,57]. The conceptual alignment framework provides an account of our ability to disambiguate everyday symbols and to reliably zoom onto relevant features of multimodal communicative signals [8], and more generally for the evolutionary anomaly of extreme referential flexibility in human signals [10,58,59].

Neural Evidence for a Shared Conceptual Space Supporting Human Communication

According to the conceptual alignment framework, the communicative meaning of a signal relies on a fleeting conceptual space defined by the ongoing interaction. This notion leads to four predictions of the characteristics of neural activity supporting human communication. First, achieving mutual understanding should evoke neural activity reflecting flexible conceptual processes [60–62] rather than sensorimotor operations with limited generalization potential [2,16,18,63–65]. Second, there should be shared patterns of neural activity during the generation and interpretation of communicative signals, given that these processes relate to the same conversational context [66,67]. Third, the timing of this shared neural pattern should lead, not follow, the occurrence of a communicative signal, given that the conceptual space that gives

CellPress

Key Figure

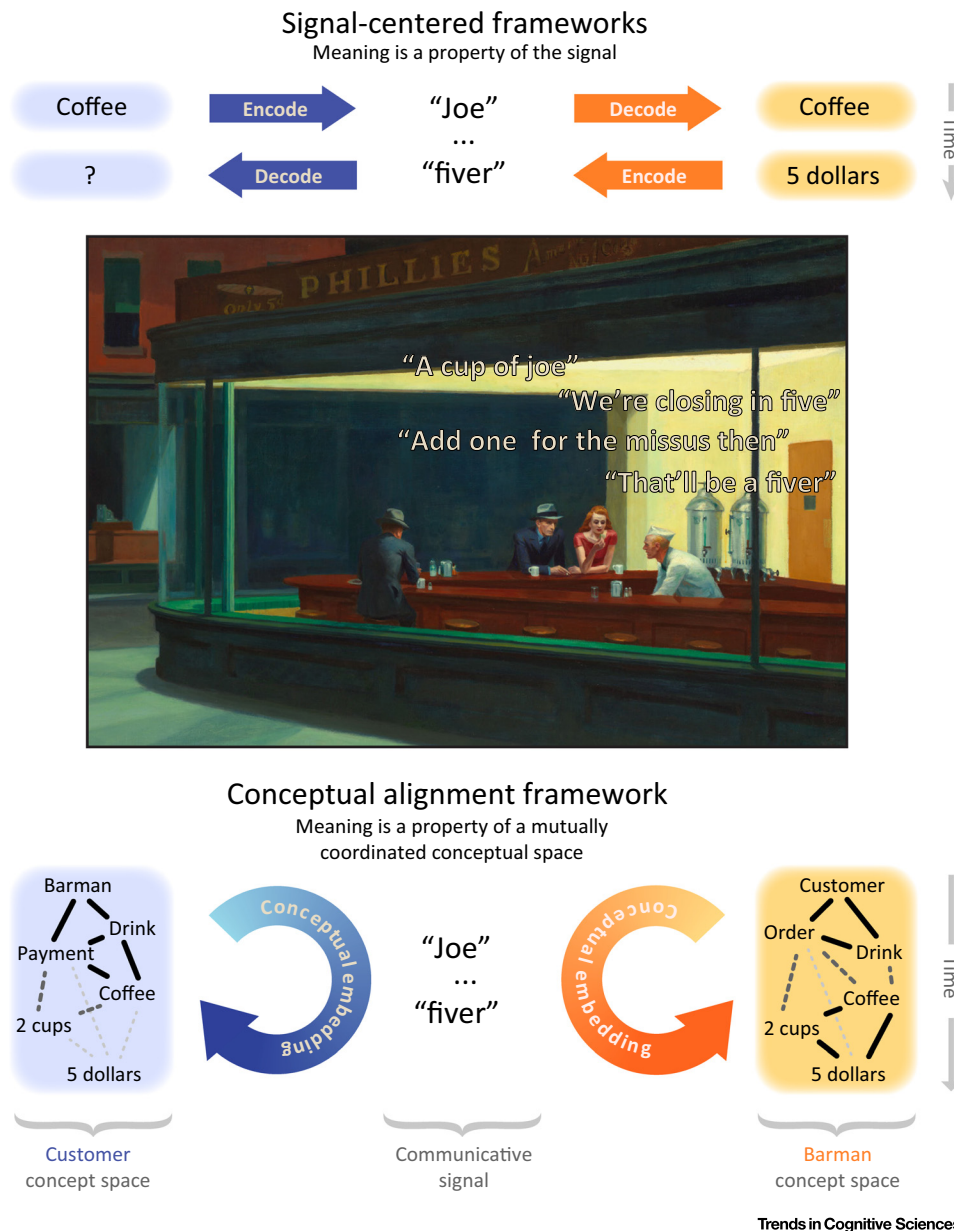## A Shift in Our Conceptualization of Human Mutual Understanding

Figure 2. To date, several accounts of our communicative abilities have focused on the information content of the signals we use to understand each other during communicative interactions. These accounts implicitly assume that statistical regularities in these signals could coordinate the production and comprehension of their meanings across communicators. By contrast, according to the conceptual alignment framework, human communicators mutually coordinate a fleeting conceptual space in which signals are merely a means to seek and provide evidence for mutual understanding. By embedding communicative signals in this dynamically adjusted space (portrayed by the broken lines), communicators can flexibly resolve the ambiguities inherent in these signals (e.g., 'Joe'), often at their first occurrence (e.g., 'fiver'). Artwork courtesy of the Art Institute of Chicago.

CellPress
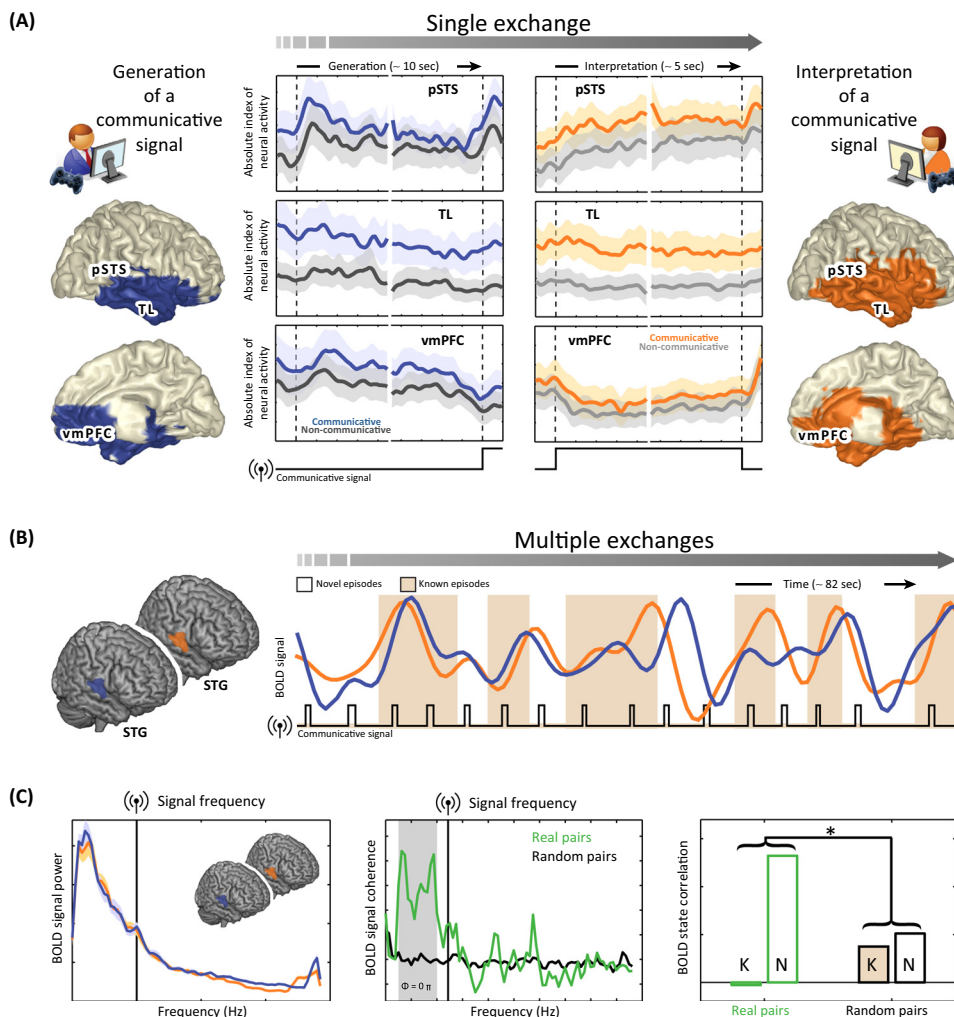
Trends in Cognitive Sciences

Figure 3. Neural Dynamics of Sharing Conceptual Spaces within a Single Communicative Exchange (A) and Over Multiple Exchanges (B). (A) Compared with non-communicative control interactions, brain regions necessary for processing conceptual knowledge are already upregulated during communicative interactions before a communicative signal is generated or interpreted (epochs 1 and 2 of Figure I in Box 1) [29]. The right temporal lobe (TL) shows tonic upregulation of neural activity during both the generation and interpretation of communicative signals, without a transient response time locked to the sensorimotor events of these epochs. These temporal dynamics indicate that neural activity in the right TL is modulated by communication over a timescale decoupled from signal occurrences. Ongoing neural activity in the ventromedial prefrontal cortex (vmPFC) is also upregulated as a function of the communicative task set, with generation evoking more computational loads than interpretation irrespective of the communicative nature of the task. This pattern fits with the recent observation that vmPFC lesion patients remain able to generate communicatively effective signals but these communicative decisions are not fine-tuned with a conceptual space defined by the ongoing interaction [106]. Differently from the right TL and the vmPFC, the posterior superior temporal sulcus (pSTS) is sensitive to computational demands that occur early during generation and rise during interpretation; that is, with the presentation of new stimulus material. Repetitive transcranial magnetic stimulation (TMS) over the right pSTS perturbs action understanding on the basis of the recent communicative history, suggesting that this region is necessary for integrating sensory material with a conceptual space defined by the ongoing interaction [107]. (B) As people converge on shared conceptual spaces over multiple communicative exchanges, the right superior temporal gyrus (STG) is increasingly involved irrespective of whether a communicative signal is being generated or interpreted (see [72]). The right STG exhibits more cerebral activity during communicative episodes in which interlocutors can rely on previously shared conceptual spaces (tan blocks depicting one or more consecutive exchanges of a type for which joint solutions have been previously established). (C) Spectrotemporal characteristics of conceptual alignment. Blood oxygen level-dependent (BOLD) signal power and coherence spectra of the right STG indicate that sharing a communicative history (real pairs, green) may result in matched, zero phase-lag cerebral dynamics across

(Figure legend continued on the bottom of the next page.)

meaning to the signal is defined by the ongoing communicative interaction rather than by the signal itself. Fourth, the temporal dynamics of the shared neural pattern should reflect the communicators' adjustments of their shared conceptual space.

The first three predictions were verified in a magnetoencephalography (MEG) study that revealed changes in neural activity sensitive to the task context rather than to the occurrence of specific communicative events [29]. Achieving mutual understanding (Box 1) upregulated neural activity in the right temporal lobe (TL) and in the ventromedial prefrontal cortex (vmPFC), two brain regions necessary for processing conceptual knowledge [68–71]. The same upregulation of neural activity was found across communicators and addressees irrespective of whether a communicative signal was being generated or interpreted (Figure 3A). It should be emphasized that conventional neuroimaging approaches, focused on changes in neural activity relative to a trial or session baseline, would not be able to detect these tonic upregulations of activity. These novel effects were isolated using dedicated time–frequency analyses focused on an absolute index of neural activity (Figure 3A). The finding indicates that these brain regions supported communication by means of a tonic upregulation of neural activity linked to task features shared across interlocutors rather than to sensorimotor events differing between interlocutors by experimental design. The overlapping neural upregulation occurred before any communicative signal was generated or interpreted and the magnitude of the neural upregulation had measurable consequences on communicative performances (see [29]). These empirical observations confirm the intuition that the meaning of a communicative signal is inferred by embedding that signal in a conceptual space whose activation precedes in time the processing of the communicative stimulus material itself (see Box 2 for a discussion of a neuronal implementation).

The fourth prediction was verified using two fMRI scanners to simultaneously record cerebral activity in pairs of communicators trying to understand each other over a series of communicative interactions (Box 1) in which shared conceptual spaces had to be coordinated de novo ('novel interactions') or retrieved from a preceding training session ('known interactions') [72]. This experimental manipulation evoked more cerebral activity in the right TL and the vmPFC during known than during novel interactions, in both communicators and addressees. Crucially, an anterior portion of the right TL followed the behavioral dynamics of mutual understanding over the course of the experiment. This region, the superior temporal gyrus (STG), was not only sensitive to the known/novel nature of communicative interactions (Figure 3B) but also became increasingly involved as people converged on shared conceptual spaces over the course of novel interactions (see [72]). Participants with a common communicative history showed synchronized intercerebral dynamics in their STG activity. Crucially, this pair-specific intercerebral coherence occurred over a timescale decoupled from signal occurrences (Figure 3C) and only when the communicators needed to mutually adjust their conceptual spaces (far-right plot in Figure 3C). This observation indicates that converging on conceptual spaces may result in cerebral coherence between communicators at temporal scales independent from signal occurrences, providing empirical evidence against accounts of human communication that emphasize priming, neural synchrony, or shared sensorimotor processes as the basic mechanism for mutual understanding [1,2,15–20,73].

---

communicators over a timescale decoupled from signal occurrences [see the black boxcar traces in (A) and (B)]. Gray surface indicates frequencies with a statistically significant difference in coherence between real and random pairs. The bar plot indicates that this pair-specific cerebral coherence occurs when pairs need to mutually adjust their conceptual spaces (N, novel episodes), but not when those pairs can fully rely on previously shared conceptual spaces that do not require equally frequent updating (K, known episodes). BOLD state in the bar plot refers to cerebral activity evoked during a known or novel episode, controlled for transient task events within each communicative episode. Random pairs (black) are combinations of participants that were engaged in the same communicative task as real pairs but without a shared communicative ground (e.g., communicator from pair 1 with addressee from pair 2). Asterisk denotes a statistically significant interaction between pair and episode type. Adapted from [29,72].

Box 2. A Neuronal Mechanism for Integrating Communicative Signals into Dynamically Adjusted Conceptual Structures

What neurophysiological mechanism allows a sustained yet adjustable influence of conceptual knowledge on transient signal production and perception? A clue is provided by the spectrotemporal nature of neural activity observed over the frontal and temporal cortices when people interact. These brain regions are tonically upregulated during communicative interactions – indeed already before the occurrence of observable events – and with measurable behavioral consequences on communicative performance [29]. These same regions also show strikingly similar phasic neural dynamics during the generation and interpretation of non-communicative events (Figure 3A). This observation is consistent with an increasing body of evidence showing that contextual demands can modulate ongoing neural activity yet retain responsiveness to event-related neural processing (e.g., [99,100]).

Second, the neural upregulation found over the frontal and temporal cortices had an extremely broad spectral profile [29]. Physiologically, broadband shifts of neural activity (a change of amplitude across all frequencies) are functionally distinct from band-limited neuronal oscillations [101] and are thought to reflect changes in the mean firing rates of neuronal populations [102,103]. These population-level firing rates have been shown to be affected by internal cortical states as much as by external stimuli [104] and are instrumental for integrating driving afferences with contextual information [74–78]. As portrayed in Figure I, ongoing contextual inputs can temporarily hold selective neurons in an excitable state. The excitable state increases the probability of these neurons propagating afferences, effectively integrating information associated with the input streams.

This neuronal mechanism, based on upregulated broadband neural activity, might provide a neural marker of fleeting knowledge spaces [29]. By contrast, phasic neural dynamics may be indicative of event-related computations. Analytical approaches focused on event-related neural activity may largely miss temporally extended computations that support conceptual processing [99].
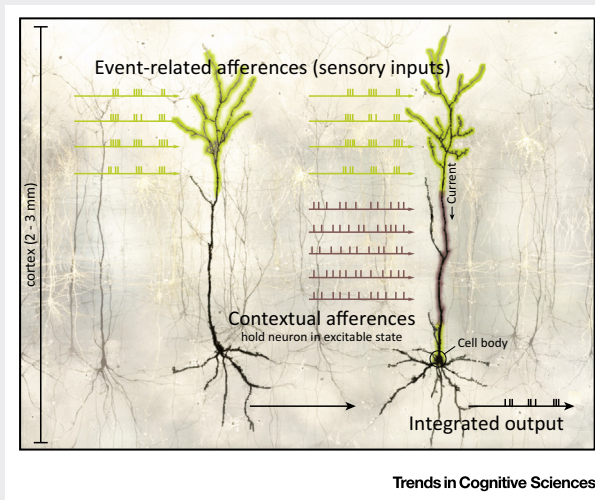

Trends in Cognitive Sciences

Figure I. Neuronal Implementation of Conceptual Spaces. Unlike axonal action potentials, dendritic currents do not propagate reliably over long distances. Propagation relies on the currents' ability to reach a threshold at the neuron's cell body and thus, in part, on their proximity to that cell body. This location dependency of a neuron's excitatory input is instrumental for integrating driving afferences with contextual information [74–78]; namely, neurons that are temporarily upregulated through ongoing inputs are more likely to propagate sensory afference than neurons that are in a less excitable state. When the summed inputs reach the cell body's threshold, a downstream action potential is generated, effectively integrating information associated with the two input streams (e.g., the situation depicted on the right). Electrophysiologically, even when dendritic currents fail to trigger axonal action potentials (e.g., the left situation), their field potentials are captured as a spatially weighted average by the electroencephalography (EEG)/magnetoencephalography (MEG)/electrocorticography (ECoG) signal [105]. Ongoing contextual inputs that hold neuronal populations near an excitability threshold may thus induce changes in the brain signal that are not temporally bound to the occurrence of observable events. The upregulation of broadband neural activity during human communicative interactions (Figure 3A) might be an instance of this contextual phenomenon [29], the extremely broad spectral profile owing to the noise-like distribution of input arrival times [102]. Artwork courtesy of Greg Dunn Design.

Concluding Remarks

This opinion article provides theoretical and empirical arguments for a shift in our conceptualization of human mutual understanding. Qualifying previous suggestions [8–14], we have shown how the neural mechanisms used during human communication are shared across communicators and addressees and how these mechanisms follow the dynamics of mutual understanding rather than the occurrences of communicative signals. We have argued that conceptual alignment (i.e., a continuous dynamic alignment of individual knowledge spaces) provides a cognitive framework suitable for resolving the ambiguities inherent in human communicative signals. Neuronally, conceptual alignments appear to be implemented through spectrally and temporally extended phenomena, namely upregulation of broadband neural activity [29] (Box 2), integrating driving afferences with contextual information [74–78].

Considering human communication as a meeting of minds rather than as transmitting signals has important implications for several academic fields (see Outstanding Questions). For instance, it becomes relevant to consider temporally extended neural integration mechanisms to understand how conceptual spaces support human semantic and pragmatic abilities, over and above the stimulus centered approaches currently predominantly used in neurocognitive studies of language and communication [79]. The study of these mechanisms could benefit from experimental protocols focused on generative communication, namely the generation of communicative behaviors from an open-ended set of possibilities [38,72], rather than on the reproduction of well-rehearsed scripts [80–82]. It becomes relevant to study how humans developmentally acquire mutual understanding within and through social interactions, over and above individualistic processes [83–85]. Cerebral alterations leading to communicative impairments like autism spectrum disorder and frontotemporal dementia [70,86,87] might benefit from being reconceptualized as deficits in creating and probing a shared conceptual space with a communicator. Finally, artificial cognitive agents might better satisfy human communicative expectations by using a cognitive architecture that continuously updates the conceptual space shared with an interlocutor, over and above rapid extraction of the statistically predominant features of a signal [3–5,88,89].

Supplemental Information

Supplemental information associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.tics.2015.11.007.

References

1. Shannon, C. (1948) A mathematical theory of communication. *Bell Syst. Tech. J.* 27, 379–423
2. Hasson, U. *et al.* (2012) Brain-to-brain coupling: a mechanism for creating and sharing a social world. *Trends Cogn. Sci.* 16, 114–121
3. LeCun, Y. *et al.* (2015) Deep learning. *Nature* 521, 436–444
4. Mnih, V. *et al.* (2015) Human-level control through deep reinforcement learning. *Nature* 518, 529–533
5. Gershman, S.J. *et al.* (2015) Computational rationality: a converging paradigm for intelligence in brains, minds, and machines. *Science* 349, 273–278
6. Frank, M.J. and Badre, D. (2012) Mechanisms of hierarchical reinforcement learning in corticostriatal circuits 1: computational analysis. *Cereb. Cortex* 22, 509–526
7. Marr, D. (1982) *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information,* W.H. Freeman
8. Sperber, D. and Wilson, D. (2001) *Relevance: Communication and Cognition,* Blackwell
9. Grice, H.P. (1969) Utterers meaning and intentions. *Philos. Rev.* 78, 147–177
10. Levinson, S.C. (2006) On the human interactional engine. In *Roots of Human Sociality* (Enfield, N. and Levinson, S., eds), pp. 39–69, Berg
11. Clark, H.H. (1996) *Using Language,* Cambridge University Press
12. Misyak, J.B. *et al.* (2014) Unwritten rules: virtual bargaining underpins social interaction, culture, and society. *Trends Cogn. Sci.* 18, 512–519

Outstanding Questions

The conceptual alignment framework implies that communicators continuously update their conceptual spaces, building on both background knowledge and the current interactional context. What biologically plausible algorithms can support the rapid exploration of these large search spaces and the generation of novel links between elements of those spaces? What neuronal mechanisms can realistically span the extremely long integration and alignment windows?

The conceptual alignment framework presupposes the continuous generation and exploration of many possible-world scenarios. What motivational mechanisms determine how deep and exhaustive that generation/exploration process should be? Could rapid turn taking, an apparently universal characteristic of human communication [90,91], modulate the exploratory process while providing a continuous motivational drive toward building a shared conceptual space?

How does mutual understanding take off when initial convergence on a shared conceptual space is severely limited? Could analogies provide a generative principle for hypothesizing possible interpretations of signals [48,49,92]? Could the combination of an analogical 'meaning hypothesizer' with a context-dependent selection mechanism lead to the generation of communicative signals interpretable by that addressee on first exposure [93]?

An influential suggestion holds that we develop mutual understanding abilities through social interaction itself [83–85], as when we increasingly include experience and beliefs about the world and others in our communicative interactions with others [94]. How do children acquire the neurocognitive mechanisms necessary to make themselves understood by others?

Can communicative alterations observed in numerous neurological and psychiatric disorders (e.g., schizophrenia, autism spectrum disorders, frontotemporal dementia) be causally unified as failures in building a shared conceptual space with an interlocutor?

13. Noveck, I.A. and Reboul, A. (2008) Experimental pragmatics: a Gricean turn in the study of language. *Trends Cogn. Sci.* 12, 425–431

14. Gärdenfors, P. (2000) *Conceptual Spaces: The Geometry of Thought,* MIT Press

15. Friston, K.J. and Frith, C.D. (2015) Active inference, communication and hermeneutics. *Cortex* 68, 129–143

16. Keysers, C. and Perrett, D.I. (2004) Demystifying social cognition: a Hebbian perspective. *Trends Cogn. Sci.* 8, 501–507

17. Pulvermuller, F. *et al.* (2014) Motor cognition–motor semantics: action perception theory of cognition and communication. *Neuropsychologia* 55, 71–84

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

19. Dumas, G. *et al.* (2014) The human dynamic clamp as a paradigm for social interaction. *Proc. Natl. Acad. Sci. U.S.A.* 111, E3726–E3734

20. Nummenmaa, L. *et al.* (2014) Mental action simulation synchronizes action–observation circuits across individuals. *J. Neurosci.* 34, 748–757

21. Danchin, E. *et al.* (2004) Public information: from nosy neighbors to cultural evolution. *Science* 305, 487–491

22. Kirby, S. *et al.* (2014) Iterated learning and the evolution of language. *Curr. Opin. Neurobiol.* 28, 108–114

23. Steels, L. (2003) Evolving grounded communication for robots. *Trends Cogn. Sci.* 7, 308–312

24. Puglisi, A. *et al.* (2008) Cultural route to the emergence of linguistic categories. *Proc. Natl. Acad. Sci. U.S.A.* 105, 7936–7940

25. Kirby, S. (2002) *Simulating the Evolution of Language,* pp. 121–147, Springer

26. Galantucci, B. (2005) An experimental study of the emergence of human communication systems. *Cogn. Sci.* 29, 737–767

27. Fusaroli, R. and Tylen, K. (2012) Carving language for social coordination: a dynamical approach. *Interact. Stud.* 13, 103–124

28. Scott-Phillips, T.C. *et al.* (2012) How do communication systems emerge? *Proc. Biol. Sci.* 279, 1943–1949

29. Stolk, A. *et al.* (2013) Neural mechanisms of communicative innovation. *Proc. Natl. Acad. Sci. U.S.A.* 110, 14574–14579

30. Tomasello, M. *et al.* (2007) A new look at infant pointing. *Child Dev.* 78, 705722

31. Tenenbaum, J.B. *et al.* (2011) How to grow a mind: statistics, structure, and abstraction. *Science* 331, 1279–1285

32. Noordzij, M.L. *et al.* (2010) Neural correlates of intentional communication. *Front. Neurosci.* 4, 188

33. Edmiston, P. and Lupyan, G. (2015) What makes words special? Words as unmotivated cues. *Cognition* 143, 93–100

34. Elman, J.L. (2004) An alternative view of the mental lexicon. *Trends Cogn. Sci.* 8, 301–306

35. Lupyan, G. and Bergen, B. (2015) How language programs the mind. *Top. Cogn. Sci.* Published online July 17, 2015. http://dx.doi.org/10.1111/tops.12155

36. Brennan, S.E. *et al.* (2010) Two minds, one dialog: coordinating speaking and understanding. *Psychol. Learn. Motiv.* 53, 301–344

37. Hofstadter, D. and Sander, E. (2013) *Surfaces and Essences: Analogy as the Fuel and Fire of Thinking,* Basic Books

38. Galantucci, B. and Garrod, S. (2011) Experimental semiotics: a review. *Front. Hum. Neurosci.* 5, 11

39. de Ruiter, J.P. *et al.* (2010) Exploring the cognitive infrastructure of communication. *Interact. Stud.* 11, 51–77

40. Fay, N. *et al.* (2010) The interactive evolution of human communication systems. *Cogn. Sci.* 34, 351–386

41. Evans, N. and Levinson, S.C. (2009) The myth of language universals: language diversity and its importance for cognitive science. *Behav. Brain Sci.* 32, 429–448 discussion 448–494

42. Scott-Phillips, T.C. *et al.* (2009) Signalling signalhood and the emergence of communication. *Cognition* 113, 226–233

43. Scott-Phillips, T. (2014) *Speaking Our Minds: Why Human Communication is Different, and How Language Evolved to Make it Special,* Palgrave Macmillan

44. Reber, A.S. (1993) *Implicit Learning and Tacit Knowledge: An Essay on the Cognitive Unconscious,* Oxford University Press

45. Donoso, M. *et al.* (2014) Human cognition Foundations of human reasoning in the prefrontal cortex. *Science* 344, 1481–1486

46. Botvinick, M. and Weinstein, A. (2014) Model-based hierarchical reinforcement learning and human action control. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* Published online November 5, 2014. http://dx.doi.org/10.1098/rstb.2013.0480

47. Grice, H.P. (1975) Logic and conversation. In *Syntax and Semantics 3: Speech Acts* (Cole, P. and Morgan, J.L., eds), pp. 41–58, Academic Press

48. Gentner, D. (2003) Why we're so smart. In *Language in Mind: Advances in the Study of Language and Thought* (Gentner, D. and Goldin-Meadow, S., eds), pp. 195–235, MIT Press

49. Goldstone, R.L. and Rogosky, B.J. (2002) Using relations within conceptual systems to translate across conceptual systems. *Cognition* 84, 295–320

50. Durso, F.T. *et al.* (1994) Graph-theoretic confirmation of restructuring during insight. *Psychol. Sci.* 5, 94–96

51. Blokpoel, M. *et al.* (2011) The computational costs of recipient design and intention recognition in communication. In *Proceedings of the 33rd Annual Conference of the Cognitive Science Society,* Cognitive Science Society

52. van Rooij, I. *et al.* (2011) Intentional communication: computationally easy or difficult? *Front. Hum. Neurosci.* 5, 52

53. Humphries, M.D. *et al.* (2012) Dopaminergic control of the exploration–exploitation trade-off via the basal ganglia. *Front. Neurosci.* 6, 9

54. Kayser, A.S. *et al.* (2015) Dopamine, locus of control, and the exploration–exploitation tradeoff. *Neuropsychopharmacology* 40, 454–462

55. Pezzulo, G. *et al.* (2014) Internally generated sequences in learning and executing goal-directed behavior. *Trends Cogn. Sci.* 18, 647–657

56. Garrod, S. *et al.* (2007) Foundations of representation: where might graphical symbol systems come from? *Cogn. Sci.* 31, 961–987

57. Gershman, S.J. and Niv, Y. (2010) Learning latent structure: carving nature at its joints. *Curr. Opin. Neurobiol.* 20, 251–256

58. Tomasello, M. (2008) *Origins of Human Communication,* MIT Press

59. Centola, D. and Baronchelli, A. (2015) The spontaneous emergence of conventions: an experimental study of cultural evolution. *Proc. Natl. Acad. Sci. U.S.A.* 112, 1989–1994

60. Kumaran, D. *et al.* (2009) Tracking the emergence of conceptual knowledge during human decision making. *Neuron* 63, 889–901

61. Siegal, M. and Varley, R. (2002) Neural systems involved in "theory of mind". *Nat. Rev. Neurosci.* 3, 463–471

62. Hoffman, P. *et al.* (2014) The anterior temporal lobes are critically involved in acquiring new conceptual knowledge: evidence for impaired feature integration in semantic dementia. *Cortex* 50, 19–31

63. Hari, R. *et al.* (2013) Synchrony of brains and bodies during implicit interpersonal interaction. *Trends Cogn. Sci.* 17, 105–106

64. Jiang, J. *et al.* (2012) Neural synchronization during face-to-face communication. *J. Neurosci.* 32, 16064–16069

65. Rizzolatti, G. and Craighero, L. (2004) The mirror-neuron system. *Annu. Rev. Neurosci.* 27, 169–192

66. Menenti, L. *et al.* (2012) Toward a neural basis of interactive alignment in conversation. *Front. Hum. Neurosci.* 6, 185

67. Kuhlen, A.K. *et al.* (2012) Content-specific coordination of listeners' to speakers' EEG during communication. *Front. Hum. Neurosci.* 6, 266

68. Lambon Ralph, M.A. *et al.* (2010) Coherent concepts are computed in the anterior temporal lobes. *Proc. Natl. Acad. Sci. U.S.A.* 107, 2717–2722

69. Milne, E. and Grafman, J. (2001) Ventromedial prefrontal cortex lesions in humans eliminate implicit gender stereotyping. *J. Neurosci.* 21, RC150

70. Sabbagh, M.A. (1999) Communicative intentions and language: evidence from right-hemisphere damage and autism. *Brain Lang.* 70, 29–69

71. Beeman, M. (1993) Semantic processing in the right hemisphere may contribute to drawing inferences from discourse. *Brain Lang.* 44, 80–120

72. Stolk, A. *et al.* (2014) Cerebral coherence between communicators marks the emergence of meaning. *Proc. Natl. Acad. Sci. U.S.A.* 111, 18183–18188

73. Stolk, A. (2014) In sync: metaphor, mechanism or marker of mutual understanding? *J. Neurosci.* 34, 5397–5398

74. Jarsky, T. *et al.* (2005) Conditional dendritic spike propagation following distal synaptic activation of hippocampal CA1 pyramidal neurons. *Nat. Neurosci.* 8, 1667–1676

75. Behabadi, B.F. *et al.* (2012) Location-dependent excitatory synaptic interactions in pyramidal neuron dendrites. *PLoS Comput. Biol.* 8, e1002599

76. Larkum, M. (2013) A cellular mechanism for cortical associations: an organizing principle for the cerebral cortex. *Trends Neurosci.* 36, 141–151

77. Smith, S.L. *et al.* (2013) Dendritic spikes enhance stimulus selectivity in cortical neurons *in vivo*. *Nature* 503, 115–120

78. Bittner, K.C. *et al.* (2015) Conjunctive input processing drives feature selectivity in hippocampal CA1 neurons. *Nat. Neurosci.* 18, 1133–1142

79. Gaskell, M.G. (ed.) (2007) *The Oxford Handbook of Psycholinguistics*, Oxford University Press

80. Silbert, L.J. *et al.* (2014) Coupled neural systems underlie the production and comprehension of naturalistic narrative speech. *Proc. Natl. Acad. Sci. U.S.A.* 111, E4687–E4696

81. Schilbach, L. *et al.* (2013) Toward a second-person neuroscience. *Behav. Brain Sci.* 36, 393–414

82. Di Cesare, G. *et al.* (2015) Expressing our internal states and understanding those of others. *Proc. Natl. Acad. Sci. U.S.A.* 112, 10331–10335

83. Hrdy, S.B. (2009) *Mothers and Others: The Evolutionary Origins of Mutual Understanding,* Harvard University Press

84. Carpendale, J.I. and Lewis, C. (2004) Constructing an understanding of mind: the development of children's social understanding within social interaction. *Behav. Brain Sci.* 27, 79–96 discussion 96-151

85. de Rosnay, M. and Hughes, C. (2006) Conversation and theory of mind: do children talk their way to socio-cognitive understanding? *Br. J. Dev. Psychol.* 24, 7–37

86. Mates, A.W. (2010) Using social deficits in frontotemporal dementia to develop a neurobiology of person reference. In *Language, Interaction and Frontotemporal Dementia: Reverse Engineering the Social Mind* (Mates, A.W. *et al.*, eds), pp. 139–166, Equinox

87. Healey, M.L. *et al.* (2015) Getting on the same page: the neural basis for social coordination deficits in behavioral variant frontotemporal degeneration. *Neuropsychologia* 69, 56–66

88. Levesque, H.J. (2014) On our best behaviour. *Artif. Intell.* 212, 27–35

89. Fung, P. (2015) Robots with heart. *Sci. Am.* 313, 60–63

90. Bornstein, M.H. *et al.* (2015) Mother–infant contingent vocalizations in 11 countries. *Psychol. Sci.* 26, 1272–1284

91. Stivers, T. *et al.* (2009) Universals and cultural variation in turn-taking in conversation. *Proc. Natl. Acad. Sci. U.S.A.* 106, 10587–10592

92. Blokpoel, M. (2015) *Understanding Understanding: A Computational-Level Perspective. Donders Graduate School for Cognitive Neuroscience Series 195,* Radboud Universiteit

93. Stolk, A. *et al.* (2015) On the generation of shared symbols. In *Cognitive Neuroscience of Natural Language Use* (Willems, R., ed.), pp. 201–227, Cambridge University Press

94. Stolk, A. *et al.* (2013) Early social experience predicts referential communicative adjustments in five-year-old children. *PLoS ONE* 8, e72667

95. Feiler, L. and Camerer, C.F. (2010) Code creation in endogenous merger experiments. *Econ. Inq.* 48, 337–352

96. Selten, R. and Warglien, M. (2007) The emergence of simple languages in an experimental coordination game. *Proc. Natl. Acad. Sci. U.S.A.* 104, 7361–7366

97. Brennan, S.E. and Clark, H.H. (1996) Conceptual pacts and lexical choice in conversation. *J. Exp. Psychol. Learn. Mem. Cogn.* 22, 1482–1493

98. Blokpoel, M. *et al.* (2012) Recipient design in human communication: simple heuristics or perspective taking? *Front. Hum. Neurosci.* 6, 253

99. Abitbol, R. *et al.* (2015) Neural mechanisms underlying contextual dependency of subjective values: converging evidence from monkeys and humans. *J. Neurosci.* 35, 2308–2320

100. Luczak, A. *et al.* (2013) Gating of sensory input by spontaneous cortical activity. *J. Neurosci.* 33, 1684–1695

101. Buzsaki, G. *et al.* (2012) The origin of extracellular fields and currents – EEG, ECoG LFP and spikes. *Nat. Rev. Neurosci.* 13, 407–420

102. Miller, K.J. (2010) Broadband spectral change: evidence for a macroscale correlate of population firing rate? *J. Neurosci.* 30, 6477–6479

103. Manning, J.R. *et al.* (2009) Broadband shifts in local field potential power spectra are correlated with single-neuron spiking in humans. *J. Neurosci.* 29, 13613–13620

104. Arieli, A. *et al.* (1996) Dynamics of ongoing activity: explanation of the large variability in evoked cortical responses. *Science* 273, 1868–1871

105. Okun, M. *et al.* (2010) The subthreshold relation between cortical local field potential and neuronal firing unveiled by intracellular recordings in awake rats. *J. Neurosci.* 30, 4440–4448

106. Stolk, A. *et al.* (2015) Altered communicative decisions following ventromedial prefrontal lesions. *Curr. Biol.* 25, 1469–1474

107. Stolk, A. *et al.* (2014) Understanding communicative actions: a repetitive TMS study. *Cortex* 51, 25–34