The following full text is a publisher's version.

For additional information about this publication click this link.
http://hdl.handle.net/2066/213579

Please be advised that this information was generated on 2021-07-01 and may be subject to change.
**Article 25fa pilot End User Agreement**

This publication is distributed under the terms of Article 25fa of the Dutch Copyright Act (Auteurswet) with explicit consent by the author. Dutch law entitles the maker of a short scientific work funded either wholly or partially by Dutch public funds to make that work publicly available for no consideration following a reasonable period of time after the work was first published, provided that clear reference is made to the source of the first publication of the work.

This publication is distributed under The Association of Universities in the Netherlands (VSNU) ‘Article 25fa implementation’ pilot project. In this pilot research outputs of researchers employed by Dutch Universities that comply with the legal requirements of Article 25fa of the Dutch Copyright Act are distributed online and free of cost or other barriers in institutional repositories. Research outputs are distributed six months after their first online publication in the original published version and with proper attribution to the source of the original publication.

You are permitted to download and use the publication for personal purposes. All rights remain with the author(s) and/or copyrights owner(s) of this work. Any use of the publication other than authorised under this licence or copyright law is prohibited.

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please contact the Library through email: copyright@ubn.ru.nl, or send a letter to:

University Library  
Radboud University  
Copyright Information Point  
PO Box 9100  
6500 HA Nijmegen

You will be contacted as soon as possible.
Conceptual Alignment as a Neurocognitive Mechanism for Human Communicative Interactions

IVAN TONI AND ARJEN STOLK

In the past few years there has been renewed interest in using empirical investigations to understand the principles and the mechanisms that allow humans to communicate (de Ruiter et al., 2010; Fusaroli & Tylén, 2016; Galantucci, 2005; Selten & Warglien, 2007; Scott-Phillips, Kirby, & Ritchie, 2009). The theoretical framework of this research field, loosely grouped under the label of experimental semiotics (Galantucci & Garrod, 2012), is broadly grounded in the view of "language use as joint action" (Clark, 1996). The novelty of this research field consists in using experimentally controlled approaches, in a genuinely interactive context, to capture generative elements of human communication. Those generative elements are at the root of our ability to share thoughts with other people during our daily social interactions (Levinson, 2006). Yet, paradoxically, interactive contexts as well as generative communicative elements have often remained outside the focus of experimentally oriented linguistic studies. For instance, traditional psycholinguistics has largely focused on encoding–decoding of linguistic material as implemented by individual agents, away from the context of interaction with other interlocutors. Generative linguistics has also been largely uninterested in the actual use of language, claiming that internal structural dependencies cannot be understood through contingent externalizations (Everaert, Huybregts, Chomsky, Berwick, & Bolhuis, 2015). Experimental semiotics takes interactive contexts and generative elements seriously, while deepening our understanding of human communication beyond purely linguistic means. There is a theoretical reason supporting this approach. Language is but one of several ways in which human communicative acts can be manifested. Understanding the conditions necessary for the communicative use of language avoids the parochial yet widespread stance of considering only language-based communication as worthwhile. There is also an empirical reason for pursuing this approach to study communication. In humans, it is hard to gain experimental access to core generative processes supporting our communicative abilities, given the presence of a large set of preestablished symbols (e.g., our daily spoken idiom). The presence of those preestablished conventional symbols makes it hard to distinguish the generation of new communicative elements from the retrieval of previous communicative elements. Starting from the premise of those two reasons, experimental semiotics has proven effective in bringing generative elements of human communication under the lens of quantitative and reproducible experimental procedures, without sacrificing truly interactive aspects. The general approach is to examine communicative behaviors evoked by procedurally well-defined problems where conventional communicative tools, such as spoken language, are not available. This approach allows for quantifying the generation of novel communicative solutions and understanding the conditions under which communication arises, both within as well as across iterated communicative turns (Galantucci & Garrod, 2011).

Besides broadening the study of human communication beyond language, this new wave of empirical work on human communication has been characterized by using experimental situations that capture its inherent interpersonal nature. This is an important issue to consider when interpreting the scope and the relevance of findings from this field of inquiry. To date, the bulk of research in cognitive neuroscience and language sciences has usually neglected language use and rarely paid attention to the neurocognitive consequences of
having agents producing and comprehending communicative behaviors embedded in a genuinely interactive setting. In fact, laboratory-based studies of human communication have rarely been approached in terms of mutual understanding (Hari, Henriksson, Malinen, & Parkkonen, 2015). The issue of the triadic representation of a referent, a central issue in human communication, provides a case in point (Grice, 1957; Tomasello, 2008). It has long been acknowledged that humans regulate most communicative interactions triadically, that is, by directing the mental states of an addressee to a mutually known referent (Tomasello, Carpenter, Call, Behne, & Moll, 2005). Put differently, most human communicative behaviors do not rely on stereotyped signal–response associations as those often used across several animal taxa (Dyer, 2002; Mather, 2004; Seyfarth, Cheney, & Marler, 1980). In fact, most human communicative behaviors are structurally ambiguous, that is, the relation between behavior and informative intentions is many-to-many (Levinson, 2006). One way in which that myriad behavior–meaning mappings can be disambiguated relies on constraints imposed on the interlocutors by what they commonly presume is known and believed by the other under the current circumstances (common ground; Brennan & Clark, 1996). Studying individual agents producing scripted utterances or processing isolated sentences is hardly conducive to capture those constraints, that is, the propositions jointly taken for granted as well as the likely joint goals, norms, and affordances embedded in a communicative event. The challenge is to understand how interlocutors define those constraints precisely enough to select and understand their communicative behaviors, despite differences in their private perspectives and background knowledge of the situation at hand.

The emphasis of experimental semiotics on respecting the interpersonal and generative nature of human communication has led to several studies that have captured fundamental features of the problems solved during human communication (de Ruiter et al., 2010; Fay, Arbib, & Garrod, 2013; Fusaroli & Tylen, 2016; Galantucci, 2005; Garrod, Fay, Lee, Oberlander, & MacLeod, 2007; Healey, Sloboda, Umata, & King, 2007; Selten & Warglien, 2007; Scott-Phillips et al., 2009; Roberts, Lewandowski, & Galantucci, 2015). These features will be the focus of the next section. Here we would like to emphasize that the approach used in experimental semiotics is clearly different from other recent attempts to characterize human social cognition, largely based on either passive observation of human behaviors (Rizzolatti & Sinigaglia, 2016), dynamic interactions between multiple agents lacking a referent (Noy, Dekel, & Alon, 2011; Sebanz, Bekkering, & Knoblich, 2006), or strictly choreographed interactions where participants need to retrieve extensively trained communicative behaviors, rather than generating new ones (Silbert, Honey, Simony, Poeppe, & Hasson, 2014). Experimental semiotics has also been instrumental in focusing the research community toward an important and defining aspect of human communication, often missed by earlier work on action observation and joint action. Namely, experimental semiotics has not sacrificed the open-ended nature of human communication for the sake of experimental control. Many studies on action observation, and even some studies on the emergence of novel communicative systems (Selten & Warglien, 2007) have relied on experimentally constrained sets of stimulus–meaning mappings. This approach removes one of the most fundamental and characteristic features of the human communicative faculty, namely its ability to disambiguate a referent from a potentially infinite set of referents. Other studies have also inadvertently simplified the communicative problem by removing the need (for both communicator and addressee) to disambiguate communicative behaviors from instrumental behaviors (i.e., the problem of "signaling signalhood"; Scott-Phillips et al., 2009). This situation arises when experimenters prespecify the signals available to participants (Kirby, Griffiths, & Smith, 2014; Selten & Warglien; Smith, 2004). Experimental semiotics has also been careful to avoid confounding the generation of communicative behaviors with collateral by-products of a rigid experimental setting. An example of those experimental by-products are the cognitive control demands often imposed by apparently trivial task instructions. For instance, instead of verbally imposing a task constraint that needs to be remembered and continuously controlled (e.g., “do not speak”), experimental semiotics studies would rather nudge participants toward generating novel communicative signals by removing the possibility of using speech through computer-mediated interactions (de Ruiter et al., 2010; Galantucci, 2005). Similarly, instructing participants through task contingencies, rather than verbal instructions, is functionally closer to the natural ethology of human communication, promoting the spontaneous deployment of pragmatic inferences and shared goals (Oosterwijk et al., 2017).

2. Cognitive Mechanisms of Conceptual Alignment

Experimental semiotics uses the generation of novel shared symbols as a privileged window into the mechanisms supporting the negotiation of meaning across interacting agents, contingent on the interaction dynamics. Given that novel symbols lack a preexisting shared
representation, jointly establishing their meaning relies on converging on a common ground of knowledge and beliefs across communicators, more so than the meaning of already known words and gestures. Accordingly, the generation of novel shared symbols requires a mechanism that allows us to rapidly converge on a shared meaning, constraining a potentially infinite cognitive search space of mappings between symbols and their possible interpretations. In order to alter an interlocutor’s mental state according to the content of the communicative behavior, that mechanism should involve predictions based on presumed knowledge and beliefs of the interlocutor, conceptual knowledge that needs to be continuously updated and sharpened according to the shared history of the interaction (Brennan, Galati, & Kuhlen, 2010; Clark, 1996). For instance, we can direct the attention of an addressee to a remote referent by means of a proximate sign, as when a diner points to an empty glass to request a new bottle of wine from the waiter. The motoric simplicity of the pointing movement might easily lead one to underestimate the computational complexity of selecting that particular behavioral vehicle for conveying the intended meaning. Part of the complexity arises from the fact that, when the diner points to her empty glass, she uses a considerable scaffold of conceptual information to minimize the referential ambiguity of the movement, and she expects the waiter to use a similar scaffold to interpret the movement. The diner might use knowledge that wine can be ordered from waiters, but not from cleaners; knowledge that waiters know that customers may like a refill when their glass is empty; knowledge that the waiter saw the pointing being directed toward a red-wine glass, and that the waiter can differentiate red- from white-wine glasses; and so on. That knowledge is not immediately available in the environment. Furthermore, it is not obvious how a communicator can rapidly navigate through her general world knowledge to converge on elements that are relevant for that particular communicative behavior and to do that on first encounter with that unique set of circumstances (see also de Ruiter et al., 2010; Galantucci, 2005). This core phenomenon of human communication cannot be easily accommodated by traditional reinforcement learning theories (Sutton & Barto, 1998), game theories (Pinker, Nowak, & Lee, 2008), fast and frugal heuristics (Gigerenzer, 2008), or Bayesian models (Tenenbaum, Griffiths, & Kemp, 2006). Those models, theories, and heuristics require either a priori internal models of all possible novel signals (i.e., the models have meanings of symbols built in as conventions) or an unrealistically large number of training trials (Steels, 2003). We have recently suggested a possible computational path that does not seem to suffer from these problems: structure mapping theory, or analogical reasoning (Blokpoel, 2015; Gentner, 1983). In analogical reasoning, one uses representations of the relational structure of concepts to find analogical matches between different concepts (e.g., the atom is like a solar system, with electrons circling the nucleus in much the same way as planets circle the sun). Based on such matches, one can then transfer knowledge from a base concept to a target, generating new concepts (e.g., perhaps the revolving of electrons is caused by the attraction of the nucleus like the revolving of the planets is caused by attraction of the sun). This kind of reasoning might have the combination of generative power and conceptual sensitivity adequate to support the computational requirements for fast, even one-trial, convergence on a communicative behavior that can plausibly convey an intended content to an addressee. However, even assuming that this (untested) suggestion is viable, it remains unclear how a communicator can finesse those inferences over the presumed knowledge landscape of the waiter (Blokpoel et al., 2011), as it often happens when communicative signals are adjusted to the presumed characteristics of an addressee (audience design; Clark & Carlson, 1982; Clark & Murphy, 1982; Newman-Norlund et al., 2009).

We have recently suggested that a further set of constraints used by interlocutors to rapidly infer the meaning of communicative signals might be found by embedding those symbols in a conceptual space whose activation predates in time the processing of the signals themselves and that consider multiple counterfactual communicative scenarios (Stolk, Noordzij, Verhagen, et al., 2014; Stolk, Verhagen, & Toni, 2016). The suggestion is that, even during a simple conversation, we continuously update and sharpen our (conceptual) priors according to the recent history of the communicative interaction, generating possible-world scenarios based on the presumed conceptual space shared with the interlocutor. In fact, several possible-world contexts might need to be continuously assessed to achieve the flexibility characteristic of human interactions (e.g., in the pointing example, when the waiter does not take the order, we might rapidly adjust to the possibility that the restaurant is closing, or that the presumed waiter is in fact a customer). This suggestion implies that alternative possible-world scenarios, including those not currently guiding the interpretation of the communicative behavior, are concurrently evaluated. Besides monitoring the effectiveness of the conceptual framework currently guiding behavior, evidence in favor of several counterfactual conceptual frameworks might need to be concurrently evaluated to enable an individual to adaptively and rapidly adjust to the fleeting

CONCEPTUAL ALIGNMENT AS A NEUROCOGNITIVE MECHANISM 251
communicative setting. The effectiveness of the ongoing conceptual framework is evaluated based on its reliability, defined as the likelihood of actual action outcomes given the current situation. The reliability of the ongoing framework thus indicates whether the situation has changed and the current conceptual framework should be adjusted accordingly (Donoso, Collins, & Koechlin, 2014; Koechlin, 2016). To optimally infer when to switch, and to which alternative regulatory strategy, the reliability of multiple previously used strategies should be concurrently monitored in an inferential buffer (Donoso et al., 2014). As the interaction unfolds, at each utterance, communicators continuously update their conceptual spaces, probe counterfactual spaces, and possibly generate additional conceptual spaces for interpreting the existing set of communicative behaviors. This conceptual alignment process builds on semantic structures operating over multiple timescales, from the fleeting idiosyncrasies of an ongoing communicative interaction to the long-term verbal and nonverbal semantic regularities acquired throughout life (Bozeat, Lambon Ralph, Patterson, Garrard, & Hodges, 2000; Regier & Kay, 2009).

3. Neural Mechanisms of Conceptual Alignment

Building on the approach detailed in the previous sections, experimental semiotics has made it possible to empirically interrogate, at the implementation level, long-standing theoretical issues (Clark, 1996; Grice, 1957; Levinson, 2006; Wilson & Sperber, 2004) and to provide new insight into the neural mechanisms of human communication. Recent reviews have summarized several important findings (Hasson & Frith, 2016; Stolk et al., 2016). Here we will focus on recent work particularly relevant to the issue of conceptual alignment, providing empirical evidence for three major insights. First, it has been shown that both production and comprehension of novel communicative behaviors is supported by a right-lateralized frontotemporal network (Noordzij et al., 2009; Stolk et al., 2013), known to be necessary for implementing the flexible conceptual processes supporting pragmatic and mental-state inferences, embedding utterances into their conversational context (Beeman, 1993; Bašnáková, Weber, Petersson, van Berkum, & Hagoort, 2013; Han et al., 2013; Sabbagh, 1999; Stolk, D’Imperio, di Pellegrino, & Toni, 2015). Those observations converge in linking a right-lateralized frontotemporal network to the processing of conceptual knowledge attributed to other human agents, dynamically adjusting that knowledge to the current situation. Second, neural activity in this right-lateralized frontotemporal network predates in time the production and comprehension of the communicative behaviors (Stolk et al., 2013; Stolk, Noordzij, Verhagen, et al., 2014). This observation fits with the notion that communicative behaviors are disambiguated in relation to a dynamic conceptual space defined by the ongoing communicative interaction. Interference studies of anterior frontal and posterior temporal nodes of that right-lateralized frontotemporal network confirm and qualify the contribution of those cortical regions. Ventromedial prefrontal lesion patients, known for their social conduct deficits (Beer, John, Scabini, & Knight, 2006), were found lacking the specific ability to constrain their ongoing communication with knowledge implied by the circumstances of an interaction, including the presumed abilities of an interlocutor (Stolk et al., 2015). Transient disruption of neural function in the right posterior superior temporal sulcus diminishes the ability to constrain the interpretation of communicative behaviors using knowledge abstracted from the recent communicative history (Stolk, Noordzij, Volman, et al., 2014). These causal observations emphasize the constraining role of contingently updated conceptual knowledge in social interaction. Third, neural activity in the same right-lateralized frontotemporal network is also modulated by experimental manipulations of the common ground shared by interlocutors.

This evidence has been gathered thanks to the novel and controlled environments used in experimental semiotics. Those environments offer the possibility of having interlocutors encounter situations for which they have previously converged on a shared communicative solution and contrast those situations with others where a shared communicative solution has not been developed yet (Stolk, Noordzij, Verhagen, et al., 2014). Communicating in those circumstances had behavioral and neural consequences. Behaviorally, interlocutors understood one another better, as indexed by the number of communicative problems jointly completed. Neuroically, the same frontotemporal network identified earlier was more involved, both during production and comprehension of a communicative behavior. Given the ability of these experimental settings to minimize access to existing communicative conventions, it becomes possible to understand how and when interlocutors define the individual constraints that disambiguate each other’s communicative behaviors. One subregion of that network, the superior temporal gyrus, showed neural dynamics matched to the behavioral dynamics of mutual understanding. A neural dynamics coupled across interlocutors was selectively evoked during communicative interactions in which the signal–meaning mappings were still open for negotiation. Crucially, the neurodynamic coupling between
interlocutors occurred over temporal scales independent of the occurrence of the communicative behaviors themselves. These observations define a neural basis for the notion that we continuously update and sharpen our conceptual spaces according to the recent history of the communicative interaction, in line with the work we have described. Moreover, the simultaneous neural changes in superior temporal gyrus during meaning negotiation across interlocutors suggest that we mutually coordinate and adjust our individual knowledge spaces.

4. Outstanding Issues

To date, most research efforts have been directed at characterizing the cognitive complexity of human communicative abilities (Clark, 1996; Tomasello, 2008; Wilson & Sperber, 2004). Yet, any communicative system, irrespectively of its complexity, would be useless without a motivational system that makes an organism devote cognitive resources toward producing and interpreting communicative behaviors. What are the mechanisms that make us invest resources into changing the mental state of another person and liberally share goals with other agents, an apparently insurmountable obstacle for several other primates (Herrmann, Call, Hernandez-Lloreda, Hare, & Tomasello, 2007; Melis, Hare, & Tomasello, 2006)? Can the cognitive exploration of conceptual spaces relevant to the current communicative situation be controlled by relatively simple hormonal factors, as shown to be the case for the cognitive processes supporting economic transactions (Kosfeld, Heinrichs, Zak, Fischbacher, & Fehr, 2005)? We have recently started to explore whether oxytocin, a neuropeptide known to influence how humans share material resources (Meyer-Lindenberg, Domes, Kirsch, & Heinrichs, 2011; Shamay-Tsoory & Abu-Akel, 2016), is also involved in modulating how we share knowledge during a communicative exchange (de Boer et al., 2017), and how those modulations are neuronally implemented (Stolk, Kokal, de Boer, Oostenveild, & Toni, 2018). We found that oxytocin administration (intranasally, placebo-controlled) drives communicators to generate signals of higher referential quality, that is, signals that disambiguate more communicative problems, and to more rapidly adjust those communicative signals to what the addressee understands. The combined effects of oxytocin on referential quality and audience design fit with the notion that oxytocin administration leads communicators to explore more pervasively behaviors that can convey their intention and to consider more deeply alternative models of their addressees. These findings suggest that, besides affecting prosocial drive and salience of social cues (Shamay-Tsoory & Abu-Akel, 2016; Zak, Stanton, & Ahmadi, 2007), oxytocin influences how we share knowledge by promoting cognitive exploration (Ye, Stolk, Toni, & Hagoort, 2016). We also found that those modulatory effects of oxytocin on human communicative behavior have a neural counterpart in the same right-lateralized frontotemporal network identified earlier as critical for processing conceptual knowledge and adjusting it to the current communicative context (Stolk, Kokal, et al., 2018). More precisely, individuals with stronger activity in this network adjusted more readily their communicative behaviors to what their interlocutor actually understood during their interactions and were less biased by their prior stereotypical beliefs about the abilities of their interlocutor. This finding points to a neurophysiological mechanism through which oxytocin influences how we keep our thoughts aligned to the current situation and to each other, driving us away from perseverating in exploiting the same semantic structures. This research line illustrates how relatively simple motivational factors, as those mobilized and modulated by hormones, might have substantial cognitive consequences, possibly by boosting preexisting social tendencies through a reduction in social anxiety (Chang, Barter, Ebitz, Watson, & Platt, 2012; Radke et al., 2015).

The crucial effect is that reduced social anxiety can then release the expression of cognitive competences that would be otherwise inhibited by competitive social dynamics (Burkart, Hrdy, & van Schaik, 2009; Hare, 2017; Melis et al., 2006), driving individuals to take more risky cognitive foraging decisions by enhancing their exploratory tendencies.

The neurocognitive mechanisms underlying communicative alterations is another research field that, in our opinion, will provide important insights when addressed through the approach and tools of experimental semiotics. For instance, there is evidence that neural degeneration of frontotemporal structures interferes with the ability to hold coherent conversations (Mates, Mikesell, & Smith, 2010) and to exhibit socially appropriate behavior (Josephs et al., 2009), which is possibly a reflection of an inability to cope with the cognitive alignment demands emphasized earlier. Ongoing behavioral work using tools from experimental semiotics might be able to precisely quantify communicative alterations in patients with frontotemporal dementia. By exploiting the natural variability in the location and extent of damaged brain tissue in this pathology, it might become possible to triangulate the relation between brain tissue loss and deficits in cognitive alignment, as the pathology progresses. Ultimately, it will be important to implement systematic...
comparisons between patient groups with communicative alterations. This pathophysiological comparative approach might isolate cognitive phenotypes that unify apparently disparate disorders (e.g., schizophrenia, autism spectrum disorder, frontotemporal dementia). For instance, it might emerge that patients with either one of those disorders have problems in creating and maintaining a shared conceptual space with an interlocutor. Preliminary evidence from our group suggests that this alteration is present in people with high-functioning autism spectrum disorder (ASD), despite relatively intact abilities and motivation to generate otherwise intelligible solutions to novel communicative challenges (Wadge, Brewer, Bird, Toni, & Stolk, 2019). By the same token, the comparative approach might pinpoint cognitive phenotypes exclusive to either one of those disorders. For instance, it might become possible to delineate the biological boundaries of the communication impairments that define one of the subdimensions of ASD (Bishop, Havdahl, Huerta, & Lord, 2016; Pina-Camacho et al., 2012). Given the aberrant functional and structural connectivity found in ASD throughout the frontotemporal network (Pina-Camacho et al., 2012; Voineagu et al., 2011), it might emerge that ASD patients have specific impairments in parsing apparently continuous behaviors into communicative and instrumental elements, preventing them from focusing on the communicatively relevant elements of a social interaction (e.g., in the diner example, the customer might communicatively point with the right hand while adjusting her hair with the left hand).

Ultimately, we believe that neurological patients will be crucial for understanding the neural mechanisms implementing conceptual alignment. Namely, patients implanted with high-density multielectrode arrays for the treatment of drug-resistant epilepsy (electrocorticography) will provide fundamental insights into the neuronal mechanisms supporting cognitive alignment. The main reason is that, despite the procedural control afforded by experimental semiotics protocols, it remains extremely hard to reliably capture neuronal traces associated with transient and abstract events such as conceptual alignment across interlocutors. Those events, being contingent on the shared history of interaction during a communicative exchange, cannot be simply averaged across multiple repeated instances, as is standard and necessary practice in most of the cognitive neurosciences. Electrocorticography can record neuronal populations with high signal-to-noise, opening up the possibility of investigating trial-unique events such as history-dependent episodes of conceptual alignment, analogous to how humans establish mutual understanding during social interaction.

5. Conclusion

We have outlined how experimental semiotics has opened the way for approaching human communication in terms of mutual understanding and offered a theoretically and empirically grounded framework that provides a neurocognitive account of our ability to disambiguate everyday communicative behaviors. In contrast to existing accounts of our communicative abilities that have focused on the information content of the symbols we use to understand each other during communicative interactions, the conceptual alignment framework emphasizes the dynamic conceptual space human communicators mutually coordinate, informed by what they commonly presume is known and believed by the other under the current circumstances. The empirical evidence suggests we continuously update and sharpen our conceptual spaces as interaction unfolds, providing an elusive yet critical backdrop for disambiguating a referent from a potentially infinite set of referents and for generating even completely novel symbols.

REFERENCES


CONCEPTUAL ALIGNMENT AS A NEUROCOGNITIVE MECHANISM


