Available online at www.sciencedirect.com
ScienceDirect
Journal homepage: www.elsevier.com/locate/cortex

Special issue: Research report

Active inference, communication and hermeneutics☆

Karl J. Friston * *and Christopher D. Frith*

The Wellcome Trust Centre for Neuroimaging, Institute of Neurology, UCL, United Kingdom

ARTICLE INFO

Article history:
Received 28 September 2014
Reviewed 28 November 2014
Revised 6 December 2014
Accepted 27 March 2015
Published online xxx

Keywords:
Communication
Neuronal
Hermeneutics
Theory of mind
Active inference
Predictive coding
Bayesian
Synchronisation of chaos

ABSTRACT

Hermeneutics refers to interpretation and translation of text (typically ancient scriptures) but also applies to verbal and non-verbal communication. In a psychological setting it nicely frames the problem of inferring the intended content of a communication. In this paper, we offer a solution to the problem of *neural hermeneutics* based upon *active inference*. In active inference, action fulfils predictions about how we will behave (e.g., predicting we will speak). Crucially, these predictions can be used to predict both self and others – during speaking and listening respectively. Active inference mandates the suppression of prediction errors by updating an internal model that generates predictions – both at fast timescales (through *perceptual inference*) and slower timescales (through *perceptual learning*). If two agents adopt the same model, then – in principle – they can predict each other and minimise their mutual prediction errors. Heuristically, this ensures they are singing from the same hymn sheet. This paper builds upon recent work on active inference and communication to illustrate perceptual learning using simulated birdsongs. Our focus here is the neural hermeneutics implicit in learning, where communication facilitates long-term changes in generative models that are trying to predict each other. In other words, communication induces perceptual learning and enables others to (literally) change our minds and *vice versa*.

1. Introduction

The term hermeneutics refers to the art of interpreting written texts such as holy scriptures. The key problem for hermeneutics rests on developing criteria for deciding when an interpretation is correct. This problem is not restricted to the interpretation of ancient texts. When talking to you, I cannot access your mind to check whether my interpretation

☆ "Much of the introductory material (and text) in the current paper is taken from a companion paper (Friston and Frith, 2015) that establishes the basic model of birdsong used in this paper. The companion paper establishes the notion of generalised synchrony in the setting of communication, focusing on sensory attenuation and its role in turn taking (assuming a shared generative model). The current paper addresses the next fundamental issue; namely the (joint) acquisition or learning of generative models that underlie communication and dynamic coordination (Kelso 2012)."Kelso JA. (2012). Multistability and metastability: understanding dynamic coordination in the brain. Philos Trans R Soc Lond B Biol Sci. Apr 5; 367(1591):906–18.

* *Corresponding author*. The Wellcome Trust Centre for Neuroimaging, Institute of Neurology, UCL WC1N 3BG, United Kingdom.
E-mail address: k.friston@ucl.ac.uk (K.J. Friston).

http://dx.doi.org/10.1016/j.cortex.2015.03.025
0010-9452/

of what you have just said corresponds to what you meant. I can invent a coherent story or narrative, but I can never independently verify my interpretations (Frith & Wentzer, 2013). Nevertheless, people seem to understand each other most the time. How is this achieved? In this paper, we suggest the criteria for evaluating and updating my interpretation of your behaviour are exactly the same criteria that underlie action and perception in general; namely, the minimisation of prediction error or (variational) free energy.

In a companion paper (Friston and Frith, 2015), we considered communication in terms of inference about others, based on the notion that we model and predict our sensations – sensations that are generated by other agents like ourselves. This leads to a view of communication based on a generative model or narrative that is shared by agents who exchange sensory signals. Given a shared narrative, communication can then be cast as *turn taking* (Wilson & Wilson, 2005), by selectively attending and attenuating sensory information. Attending to exteroceptive sensations enables the shared narrative to predict the sensory input generated by another (while listening). Conversely, attenuating exteroceptive input enables one to articulate the narrative by realising proprioceptive predictions (while speaking). Using simulations, we demonstrated this turn taking by assuming that both agents possessed the same generative model. In this paper, we consider how and why generative models learned by agents – who exchange sensory signals – become the same (shared) model.

Our underlying premise is that we are trying to model the causes of our sensations – and adjust those models to maximise Bayesian model evidence or, equivalently, minimise surprise (Brown & Brün, 2012; Kilner, Friston, & Frith, 2007). This perspective on action and perception has broad explanatory power in several areas of cognitive neuroscience – and enjoys support from several lines of neuroanatomical and neurophysiological evidence (Egner & Summerfield, 2013; Rao & Ballard, 1999; Srinivasan, Laughlin, & Dubs, 1982). In communication and the interpretation of intent, the very notion of *theory of mind* speaks directly to inference, in the sense that theories make predictions that have to be tested against (sensory) data. Imagine two brains, each mandated to model the (external) states of the world causing sensory input. Now imagine that sensations can only be caused by (the action of) one brain on the other. This means that the first brain has to model the second. However, the second brain is modelling the first, which means the first brain must have a model of the second brain, which includes a model of the first – and so on *ad infinitum*. At first glance, the implicit infinite regress appears to preclude a veridical modelling of another's brain. However, this infinite regress dissolves if each brain models the sensations caused by itself and the other as being generated in the same way. In other words, if there is a shared narrative or dynamic that both brains subscribe to, then they can predict each other exactly – at least for short periods of time. This is the basic idea that we pursue in the context of active inference and predictive coding.

In our previous paper, we focused on the dynamical phenomena that emerge when two dynamical systems try to predict each other. Mathematically, this dynamical coupling is called *generalised synchrony* (aka synchronisation of chaos) (Barreto, Josic, Morales, Sander, & So, 2003; Hunt, Ott, & Yorke, 1997). Generalised synchrony was famously observed by Huygens in his studies of pendulum clocks – that synchronized themselves through the imperceptible motion of beams from which they were suspended (Huygens, 1673). This nicely illustrates the *action at a distance* among coupled dynamical systems. Put simply, generalised synchronisation means that knowing the state of one system (e.g., neuronal activity in the brain) means one can predict the another system (e.g., another's brain).

We will consider a special case of generalized synchronization; namely, *identical synchronization*, in which there is a one-to-one relationship between the states of two systems. Identical synchronisation emerges when the systems that are coupled are the same. In the context of active inference, this means the two generative models are identical. But why should two agents have the same generative model? The answer is rather obvious – when they share the same generative model they can predict each other more accurately and minimise their prediction errors or surprise. The key point here is that the same principle that leads to generalised synchrony also applies to the selection or learning of the model generating predictions. This learning is the focus of the current paper, which provides an illustrative proof of principle that the hermeneutic cycle can be closed by simply updating generative models and their predictions to minimise prediction errors. Crucially, these prediction errors can be computed without ever knowing the true state of another; thereby solving the problem of hermeneutics (see Fig. 1).

The treatment of communication in this paper is rather abstract and borrows mathematical concepts from dynamical systems theory. Although we will use birdsong as a vehicle to illustrate the ideas, we do not pretend this is a meaningful model of linguistic communication (or indeed songbirds). Rather, we try to understand the dynamic coordination of richly structured behaviours, such as singing and dancing, without ascribing any (semantic) meaning or syntax to sensory exchanges. Having said this, there is growing interest in applying the principles of predictive coding to language: e.g., (Arnal & Giraud, 2012; Hickok, 2013; Pickering & Clark, 2014; Wang, Mathalon, et al., 2014) – and understanding the algebra of dynamical systems in terms of communication; e.g., (Scott-Phillips & Blythe, 2013). Furthermore, predictive coding is starting to shed light on spectral asymmetries – in coupling within the auditory hierarchy – evident in electrophysiological studies of speech processing (Arnal, Wyart, & Giraud, 2011).

This paper comprises five sections. The first sections reprise the material in (Friston and Frith, 2015), which provides a brief review of active inference and predictive coding in communication. In the second section, we described the particular (birdsong) model used to illustrate communicative inference. This model has been used previously to illustrate several phenomena in perception; such as perceptual learning, repetition suppression, and the recognition of stimulus streams with deep hierarchical structure (Friston & Kiebel, 2009; Kiebel, Daunizeau, & Friston, 2008). In the third section, we provide a simple illustration of omission related responses – that are ubiquitous in neurophysiology and