Consciousness and Cognition 36 (2015) 376–389

Contents lists available at ScienceDirect

Consciousness and Cognition

journal homepage: www.elsevier.com/locate/concog

The felt presence of other minds: Predictive processing, counterfactual predictions, and mentalising in autism

Colin J. Palmer [a,*], Anil K. Seth [b], Jakob Hohwy [a]

[a] Cognition and Philosophy Lab, Philosophy Department, Monash University, Clayton, Victoria 3800, Australia
[b] Sackler Centre for Consciousness Science, School of Engineering and Informatics, University of Sussex, Brighton BN1 9QJ, UK

ARTICLE INFO

Article history:
Received 31 October 2014
Revised 28 March 2015
Accepted 14 April 2015
Available online 28 April 2015

Keywords:
Predictive coding
Autism
Mentalising
Theory of mind
Social cognition
Presence
Veridicality
Counterfactuals
Active inference

ABSTRACT

The mental states of other people are components of the external world that modulate the activity of our sensory epithelia. Recent probabilistic frameworks that cast perception as unconscious inference on the external causes of sensory input can thus be expanded to enfold the brain's representation of others' mental states. This paper examines this subject in the context of the debate concerning the extent to which we have perceptual awareness of other minds. In particular, we suggest that the notion of *perceptual presence* helps to refine this debate: are others' mental states experienced as veridical qualities of the perceptual world around us? This experiential aspect of social cognition may be central to conditions such as autism spectrum disorder, where representations of others' mental states seem to be selectively compromised. Importantly, recent work ties perceptual presence to the counterfactual predictions of hierarchical generative models that are suggested to perform unconscious inference in the brain. This enables a characterisation of mental state representations in terms of their associated counterfactual predictions, allowing a distinction between spontaneous and explicit forms of mentalising within the framework of predictive processing. This leads to a hypothesis that social cognition in autism spectrum disorder is characterised by a diminished set of counterfactual predictions and the reduced perceptual presence of others' mental states.

1. Introduction

There is an intuitive difficulty in how to characterise our experience of other people's mental states. At times we engage in reasoning about what might be going on in another person's mind, but it is also common for our awareness of others' mental states to appear more automatic and direct than explicit thought. For example, as our friend smiles at us we have an immediate sense of her happiness. As she glances at her drink we have an immediate understanding of her intention to reach out and grasp it. There is an ongoing contemporary debate regarding how 'direct' social cognition is (reviewed in Michael & De Bruin, 2015); the question is, to what extent do we have *perceptual* awareness of others' mental states? This subject is important, in part, because clinical conditions, including autism spectrum disorder (ASD) and schizophrenia, are associated with atypicalities in social function that elude circumscribed descriptions in terms of perceptual and cognitive functions related to the representation of others' mental states. The neural and functional bases of symptoms in ASD (and schizophrenia) may be better understood as our conception of mentalising itself is further developed.

* Corresponding author.
E-mail addresses: Colin.Palmer@monash.edu (C.J. Palmer), A.K.Seth@sussex.ac.uk (A.K. Seth), Jakob.Hohwy@monash.edu (J. Hohwy).

http://dx.doi.org/10.1016/j.concog.2015.04.007

In this paper, we outline an account of how mental states come to be represented in the brain – this being mentalising – via the same process as perceptual representation. Namely, this is the unconscious process of hierarchical Bayesian inference on the (hidden) causes of sensory input. Just as non-social objects in our environment are causes of our visual input, the mental states of other people are a part of the physical structure of the world that produces the stream of sensory impressions that our brains receive. In this view, mentalising occurs implicitly and shares a fundamental similarity with the representation of non-social objects: each is a natural result of the brain's endeavour to best explain its sensory input.

Against this background, we argue that the perceptibility of mental states can be explicated in terms of whether they are experienced as having *perceptual presence*. This is the sense of veridicality attached to the content of our perception – the feeling that the objects that we perceive *really exist* in the environment around us. A recent development of unconscious inference points to specific functional aspects of sensory processing that arguably underlie this sense of presence. On this view, perceptual presence rests on the hierarchical depth and – innovatively – the counterfactual richness of the generative models supporting unconscious inference, where counterfactual richness refers to the degree to which generative models encode predictions not only about the hidden causes of current sensory inputs, but also about the causes of prospective sensory inputs conditioned on possible but unexecuted actions. Generalising these ideas to social perception furnishes a concept of 'mental presence', which rests on counterfactual predictions about the mental-state causes of sensory impressions. Consequently, we examine how difficulties in social cognition in ASD may relate to problems with the implicit predictive modelling of the sensory consequences of actions that we can perform to interact with others' mental states.

In the next section, we cast mentalising as a process of causal inference and discuss how the representation of mental states can be situated relative to perception in the predictive processing framework of the brain. In Sections 3 and 4, we show how the directness of social perception can be explicated in terms of the perceptual presence of mental states, and that this coheres with recent predictive processing accounts of this phenomenon that call on the counterfactual richness of hierarchical generative models. Finally, in Section 5, we draw out the implications for our understanding of clinical conditions associated with differences in mentalising: is the perceptual presence of others' mental states etiolated in ASD?

2. Social cognition as causal inference

2.1. The world in the brain

To understand the relationship between perception and social cognition, we should first consider how the brain comes to represent the world beyond the skull. To this end, we will briefly introduce the *predictive processing* framework of brain function (Clark, 2013; Friston, 2005; Hohwy, 2013). This framework provides a powerful and compelling account of how the brain responds to sensory stimulation to embody information about its environment.

Predictive processing builds upon the notion of *unconscious perceptual inference* developed by Helmholtz (1860), Gregory (1980) and others, most recently in the form of Bayesian models of brain function (Kersten, Mamassian, & Yuille, 2004; Knill & Pouget, 2004; Vilares & Kording, 2011). This is the idea that what we are perceptually aware of, at any given moment, is the state of the world that is calculated as being *most likely* to be causing the sensory input that our brain receives, given prior beliefs about these causes that are furnished by previous experiences, development, and evolution. This idea emerges as a response to a fundamental problem, namely that through various causal interactions, physical states external to the brain (but including the body) conspire to produce an ambiguous stream of activity at our peripheral sensory receptors. Specifically, the relationship between states of the environment and sensory input is not one-to-one, as multiple causes interact non-linearly to produce the received input. The brain is thus faced with an inverse problem: the sensory effects of this process are directly accessible but their external causes are not. Yet it is only by controlling the causes of sensory input – which are physical features of the body and the world – that the brain can ensure adaptive behaviour and ultimately its own continued existence.

How then are the causes of stimulation represented in the brain? Predictive processing suggests that the problem is resolved by *hierarchical generative models* instantiated across the cerebral cortex (Friston, 2005). A generative model is a probabilistic mapping from worldly (and bodily) causes to sensory data; in other words, it specifies what sensory input *would* be received *if* a certain set of causes existed in the world. By feeding a generative model some expected (hypothetical) causes, we get *predictions* about the sensory data, which when mapped to cortical architecture are mediated by backwards connections between distinct levels in a hierarchy of cortical levels. These predictions are compared, continuously, to the *actual* sensory states at each level, with mismatch between the two constituting the *prediction error*. Prediction errors are propagated to higher cortical levels to provide information on how well the generative model and its input (the hypothesised causes) are replicating the sensory data. *Recognition* is the process of determining what causes are most likely to produce the observed sensory input; this is unconscious perceptual inference, and occurs by implicitly 'inverting' the generative model by finding the hypothetical causes that minimise prediction error.

A crucial feature of this framework is that prediction-error minimisation occurs simultaneously at multiple levels of representation across the cortex (Friston, 2005; Shipp, Adams, & Friston, 2013). Lower levels of the hierarchy encode hypothesised causes that act over smaller spatial and temporal scales (e.g., the perspective-dependent edges and colours that compose facial features). Higher levels represent expected causes operating over greater spatial and temporal scales (e.g., non-perspectival face recognition; Shipp et al., 2013). Expectations regarding higher-level sensory attributes generate