

**Is a modular cognitive architecture compatible with the direct perception of mental states?**

Jane Suilin Lavelle

School of Philosophy, Psychology and Language Sciences, University of Edinburgh.

[j.s.lavelle@ed.ac.uk](mailto:j.s.lavelle@ed.ac.uk)

School of PPLS, 3 Charles St., Edinburgh EH8 9AD

# **Is a modular cognitive architecture compatible with the direct perception of mental states?**

## **Abstract**

The Direct Social Perception Hypothesis maintains that we can perceive other people's psychological states. Furthermore, it claims that doing so does not require any cognitive process that is simulative or theory-like, putting it in sharp contrast with mainstream accounts of social cognition. This paper contrasts the DSPH against the modular account of mindreading as proposed by Peter Carruthers and H. Clark Barrett. It maintains that the modularity view can respond to the challenges levelled by the DSPH, and that the positions are not as distinct as they originally appear. Finally, the paper discusses the role of non-folk psychological state concepts in our perceptions of other people.

## **1. Introduction**

The Direct Social Perception Hypothesis (DSPH) maintains that we often perceive other people's psychological states and that doing so does not require any cognitive mechanisms in addition to those which facilitate perception. The hypothesis in its current form is associated with Shaun Gallagher's work (2005; 2008a; 2008b; Gallagher & Zahavi, 2008), and this paper examines a recent presentation of the view in Gallagher and Varga (2014). The DSPH was originally introduced as a counter to simulation and theory-based accounts of mindreading, claiming that such positions were committed to the view that mental states were unobservable. In the light of what has been written, particularly by psychologists working in social cognition, such a characterisation is wholly understandable; for example, Henry Wellman, an advocate of scientific-theory-theory, writes that 'mental states, such as beliefs and desires, are private, internal and not observable in others.' (1992, p. 107).<sup>1</sup> The most natural interpretation of this claim is that the content of our perceptual experiences does not include other people's psychological states. To use Gallagher's terminology, it implies that our perception of others is 'not-so-smart'. A direct perception experience which is 'not-so-smart' is one where very little information about the object is available in the

---

<sup>1</sup> Such claims are common-place; for more examples see Lavelle (2012).

content of the perceptual experience. Gallagher offers the example of opening his eyes to see 'a certain unrecognized red mass with a specific shape just in front of me' (2008a, p.536). A smart direct perception is one where Gallagher opens his eyes and sees his car (ibid). One can thus interpret Wellman as saying that our perceptions of others are 'not-so-smart', in that while they can provide an experience of a person behaving, they are unable to provide information about the other's psychological states. And this is indeed how Gallagher and Varga (hence G&V) interpret such claims:

[...] both Theory-theory and simulation theory subscribe to the 'principle of imperceptibility' and hold that when we perceive others, we perceive mere bodily activity, patterns of mechanical movements that warrant the inference or suggest the correct simulation to the other's intentional states. I may register what the other person does, but until I call forth some theory, or until I run a simulation routine, I seem not to have any sense of what that person is up to or what the behavior means. (2014, p.186)

If the imperceptibility principle means that other people's mental states cannot be a part of our perceptual experience of them, and if theory-theory and simulation are committed to this principle, then theory-theory and simulation face a significant problem. For this is one case in cognitive science where experience counts for something, and our experiences of other people are not such that we see patterns of behaviour and then engage in a simulation or theoretical process to bring forth their psychological states. If simulation- and theory-theory maintain that the processes in question happen on a personal level then they are scuppered. While it may be the case that we sometimes have to consciously engage in a process of deduction or simulation because the other's behaviour appears odd to us, this does not characterise the majority of our social interactions with others. If mindreading accounts are to maintain that mental states are unobservable, then they need to explain how this can be so while mental states remain part of our phenomenological experience of others.

In response, advocates of information rich<sup>2</sup> (broadly theory-like) accounts of mindreading have claimed that they are not committed to the imperceptibility principle if it is to be interpreted as a phenomenological claim.<sup>3</sup> For example, Peter Carruthers writes,

‘We don’t just see someone’s arm moving in the direction of a transparent object, we see her as *reaching for a drink*; and we don’t just hear a stream of phonemes when someone talks, but hear him as *wanting to know the way to the church*; and so on and so forth.’ (2011, p.48)

On this view other people’s psychological states can be a part of our perceptual experience of them, be that hearing or seeing their intent. Our perceptual experiences of others are ‘smart’, to use Gallagher’s phrase.

If both proponents of the DSPH and of theory-based mindreading views agree that other people’s mental states can be part of our perceptual experience of them, then are they still in conflict? G&V maintain that they are, due to differences in how the sub-personal processes facilitating these experiences are characterised. Critically, they claim that mindreading accounts are committed to ‘extra-perceptual inferences’ at a sub-personal level, in contrast to the DSPH which posits no such inferences. This paper challenges G&V’s presentation of the distinction between the DSPH and mindreading views. In particular, it will focus on the contrast between the DSPH and the modular mindreading account as presented by Peter Carruthers and H. Clark Barrett. Although advocates of the DSPH consider their target to be much broader than this, the modular mindreading account serves as an appropriate foil for two reasons: first, it is one of the most detailed accounts of mindreading available (see Carruthers 2006, 2011, 2013); and secondly, it captures all those features DSPH advocates wish to challenge with their ‘smart perception’ account; in other words,

---

<sup>2</sup> Information rich accounts of mindreading are those which maintain that the cognitive system contains representations of many inferential rules which describe how psychological states relate to each other and behaviours. There is a further debate between scientific theory-theorists and modularity theorists about how this information is acquired. The relevant contrast is with information poor accounts, like the simulation theory, which claims that few rules need be represented. For more on this distinction see Goldman (1992).

<sup>3</sup> Similar claims may also have been made by advocates of the simulation theory. The focus of this paper is, however, on information rich accounts.

modular mindreading epitomises the kind of theory of social cognition that the DSPH is purported to oppose.

## **2. DSPH vs. extra-perceptual inferences**

Both the DSPH and theory-based views agree that other's psychological states can be part of the contents of our perceptual experience of them; the disagreement now lies in whether the sub-personal processes that facilitate such experiences draw on any cognitive processes that are not perceptual ones. I will not be addressing the question of whether 'inference' properly describes any of our sub-personal cognitive processes: this debate merits more attention than it can be given here. Instead, the focus will be on the coherence of G&V's concept of an 'extra-perceptual' cognitive process.

### **2.1. Sub-personal smart perception**

G&V characterise smart perception as follows:

In the case of smart social perception, the brain actively contributes – more precisely, the organism, including the brain, is engaged and has something to contribute to the shaping of perception. Perception involves complex, dynamic processes at a sensory-motor level – but these processes are part of an enactive engagement or response of the whole organism, rather than additional, extra-perceptual, inferential or simulative processes. (G&V 2014, p.192)

The implicit foil here is a view of perception which does insist that an 'extra-perceptual, inferential or simulative process' is required for psychological states to be part of the perceptual experience of the other. Such a view claims that...

...perception of an intention is underpinned by both perceptual and extra-perceptual sub-personal processes. On some versions of TT and ST something like an extra-perceptual inference (or simulation) is added to the perception because the perception itself is characterised as an impoverished form of observation, detached from action (or interaction). (ibid p.191)

Although they claim that theory-theory and simulation characterise perception as 'an impoverished form of observation', G&V are not implying that these views entail a

perceptual experience that is impoverished (see above). Rather, their view is that in terms of sub-personal processing theory-theory and simulation are committed to a two stage process: the first is the processing of the perceptual process and the second is the addition of mental state information to the outcome of that process. The initial perceptual process only yields ‘not-so-smart’ perception, a perception with crucial information missing and which therefore needs to be supplemented by theory or simulation. They describe ‘not-so-smart perception’ of another person as seeing ‘meaningless behaviour... my eyes are working fine; my visual cortex is processing all of the visual information, but what vision delivers is relatively meaningless “thin” behaviour, which I then have to interpret in some further cognitive steps that involve inference’ (ibid, p.192). The ‘not-so-smart’ perception remains sub-personal while the theory/simulation processes do their work before the phenomenal experience of ‘seeing someone doing yoga’ is produced.<sup>4</sup>

For G&V the ‘perception-plus-inference’ story fundamentally mischaracterises the nature of perceptual processes because they already incorporate processes which generate mental state information. They use the example of the fusiform area to illustrate this point, the area of the brain which is active when we see faces, but which is also active when we see face-like things, e.g. the lights and grill on the front of a car. They maintain that the activation of the fusiform area is partially constitutive of the process facilitating perception: ‘I do not perceive and then go through some process that correlates to the activation of the fusiform area; rather, fusiform activation helps to constitute the way that I perceive the car, or the other person’s face etc.’ (ibid). G&V are advocating a liberal account of perceptual processes, one that can incorporate processes from diverse parts of the cognitive system, like the fusiform ‘face recognition’ area, and their concern is that accounts which see mental state information as a theory or simulation process separate from perception do not sufficiently respect the richness of the perceptual process.

---

<sup>4</sup> Earlier papers by Gallagher suggested that theory-theory and simulation were committed to our person-level perceptual experiences as being ‘not-so-smart’. He has since adjusted his position to target sub-personal processes in response to theory-theorists’ protestations that his earlier characterisation did not properly describe their position (see Carruthers 2011).

The debate here is essentially over what constitutes the perceptual process, and whether the processes that facilitate the perception of mental states can be a part of it. In many respects it's unclear what transpires if a process *P*, formerly thought to be distinct from the perceptual process but nevertheless contributing to the phenomenal perceptual experience, should turn out to be constitutive of the perceptual process after all. But even if we grant that this is an important distinction to make, it's not clear that a theory-theory or simulation view need commit to the 'perceptual processes plus extra inferences' story.

## 2.2. The 'bulletin board' model of perception

Modularity theory has evolved since Fodor's classic 1983 exposition, and in particular the concepts of 'encapsulation' and 'domain specificity' have become more detailed. Recent advocates of massive modularity individuate modules by their functional role, describing modules as *functionally* encapsulated as each module performs only one function and its internal processes cannot be co-opted to perform a different functional role (Carruthers 2006, p.xii); e.g. the module that processes and marks particular visual representations as 'faces' cannot also be used to recognise snakes. This concept of a module is much closer to how the term is used in biology and does not share many of the features of the Fodorian modular system. Most importantly, the modularity advocated by Peter Carruthers (2004; 2006; 2011) and H. Clark Barrett (2005; 2006) does not make the same commitments to innateness, input limitations, and non-conceptual output that Fodor considers to be typical features of modules.

It is the difference in how one understands the inputs to cognitive modules that is most salient to this debate. On the Barrett/Carruthers conception, a modular cognitive architecture has the feature of 'access generality with processing specificity' (Barrett, 2005, p. 270). This means that while a cognitive module can only process specific representations, it nevertheless has access to a wide pool of representations. This is a feature of Global Workspace theories of consciousness, which posit a central 'bulletin board' where sensory information is made available to multiple processing systems. These systems scan the bulletin board for information they can process, and when they come across such information they process it, altering it slightly, then post

it back to the board to be 'rescanned' by the rest of the systems. Global Workspace theories of consciousness have been discussed at length elsewhere. (Baars 1988, 1994; Carruthers 2006, 2012). While some have worried that it perpetrates Dennett's myth of the Cartesian theatre, proponents of the view counter that not all theatres are Cartesian, and a central workspace which broadcasts information to a host of processing systems is as un-Cartesian as a real theatre where audience members watch a play (Baars, 1994, p. 167). The bulletin board metaphor continues to be a fruitful one within cognitive science, and can be realised through enzymatic processing systems. While Barrett and other advocates of the view do not think modules are enzymes, the enzymatic process is used as evidence that such a bulletin board model is an ecologically viable form of information processing.

In brief, the bulletin board model of vision is as follows (for the full account see Barrett, 2005). First, 'raw data' is detected by the retina etc., but this data is not in a format that can be used by other cognitive processes. So it must be transformed by being processed by 'sensory transducers' (Fodor, 1983), during the course of which information about movement, texture, shape, distance etc. is extracted from the data and an 'object representation' is created (Barrett 2005, p. 264). The function of this process is to present the information from the retina etc. in such a way that it can be used by the modular cognitive processes, and the resulting 'object representation' meets this criterion. Barrett describes this process as an 'object parser', as the raw retinal data is transformed into information about the colour, movement, and other properties of the object that can be extracted from the retinal data. It is these representations that are posted to the bulletin board to be 'scanned' by the cognitive modules. Each module scans the board for representations it can process, but each system is itself capable of processing very specific representations.

One might think of this analogy in terms of a 'lock and key' system. A module is locked until it finds a representation with the right structural properties to unlock it. This is how enzymatic processing works: enzymes combine substrates to produce new molecules. The substrates are all in a common space, or bulletin board, where they float around bumping into enzymes. When a substrate has the physical features that can 'unlock' an enzyme, that enzyme processes the substrate and produces a new



substrate with slightly different physical properties. This new substrate is released back into the common pool until it bumps into an enzyme that it can 'unlock' so it can be altered again, etc.

Turning back to cognitive architecture, one can posit biological structures with physical features that correlate with particular visual properties that have been detected by the retina, e.g. orange-ness, motion, roundness. The physical properties have the function of representing visual properties, and can 'unlock' a cognitive module. The cognitive module evolved to detect structures with specific physical features and to alter them in a particular way; an alteration is made to the physical structure and it is placed back onto the bulletin-board. In making its alteration, the module has added a 'tag' to the structure. One can posit a module that scans the visual bulletin-board for those representations that have a physical feature that represents 'orange'. When it finds a representation with this feature it adds something to it, creating a structure that represents 'orange & danger'. This is a danger detection module. Other modules scanning the bulletin board which have evolved to process those representations with the physical features correlating with 'danger' are unlocked by this representation, and change it accordingly.

The enzymatic model of computation is borne out of the need to explain how biological system can process information about the world in a way that is ecologically and evolutionarily viable. Modules co-evolve with the object representations that are created from the retinal data: however the object parser adds the information 'is orange' to a particular object representation, this information must be in a format that can be recognised by the modules which process representations of orange objects. The biological structures that represent the environment do so through a historical correlation which need not be perfect, but simply good enough, for an organism's survival. If a particular neural structure historically correlates with the presence of orange things in the environment, then modules which evolved to recognise second-order properties of the environment like 'danger' can co-evolve with this representation so that they can be unlocked by it (Barrett 2005, p.264).

This bulletin-board organisation makes for a modular system which does not face the input limitations of Fodorian modules. Many modules have access to one bulletin board, with the result that they can act on the contents of the board in parallel. In this way one gets access generality with processing specificity, as the modules all have access to the same pool of representations, but they can only process a strictly limited subset of those representations.

Barrett suggests that each new alteration to a representation in the bulletin board constitutes a 'semantic tag' (ibid, pp.276 – 8). Imagine that there is a 'face detecting module' scanning the bulletin board for specific representations it can process.<sup>5</sup> It finds some with the right properties, processes them and adds the tag 'face'. This changes the representation slightly, such that it can be picked up and processed by another cognitive system which scans for representations tagged 'face' whose function it is to discriminate familiar faces from unfamiliar ones. It processes and changes the representation putting the new representation back into the pool tagged 'face I know'. This in turn allows the representation to be detected and processed by a more specific face detecting mechanism, perhaps adding the tag 'your mother's face'. The eventual result is a perceptual experience such that you see your mother, not just an unknown shape or person. This is precisely what grounds Carruthers' claim that mental states can be part of our perceptual experience:

Conceptual information of varying degrees of abstractness is generally bound into the content of any given sensory states and broadcast along with it. Thus Kosslyn (1994), for example, characterizes the early stages of visual processing as a continual "questioning" of non-conceptual visual input by conceptual systems, which seek a "best match" with their representations of what objects and events of the relevant kind should look like. When a match is found, it is bound into the content of the visual percept to be broadcast along with it for yet other conceptual systems to consume and draw inferences from. In this way there can be a *cascade of increasingly abstract concepts bound into any perceptual state*, as successive conceptual systems receive the product of earlier

---

<sup>5</sup> This example is purely for exposition, and is not intended to make empirical claims about how a face detection module may work.

systems' work, and categorize the input accordingly (Barrett 2005). As a result we don't just see someone's arm moving in the direction of a transparent object, we see her as 'reaching for a drink'; and we don't just hear a string of phonemes when someone talks, but we hear him as 'wanting to know the way to the church'. (2011, p.48, my emphasis).

To take a more mindreading-oriented example, one can imagine a representation of 'moving shape' being pinned onto the visual system's bulletin board. There is ample evidence to show that humans, even very young humans, are highly sensitive to movements that signal agency, e.g. motion that is contingent on the environment, like moving around an obstacle (for reviews see Johnson 2000, 2005). One can imagine an 'agency detecting' module which scans the bulletin board for movements which exhibit those features that signal agency, and which tags certain object representations as 'agents'. A representation tagged as 'agent' can then unlock the mindreading processing system, which can adjust the representation by adding its own tags, e.g. 'reaching' or 'hiding'. The tagging process can be thought of as a reaction which affects the constitution of the representation, which in turn affects whether it can unlock other processing systems (Barrett, *ibid.* 276). A representation can carry many tags, resulting in the 'cascade of increasingly abstract concepts' being bound into perceptual experiences.

This modular account can explain why the fusiform area is active when we perceive the front of a car. The 'face detecting' modular system is perhaps realised in the fusiform temporal gyrus, and one part of this system scans the visual bulletin board for representations which fit a particular template, e.g. two darker patches above a line. The object representation of a car has features which match this template, and the tag is added to it, a process which involves the fusiform area. However, other tags relevant to face recognition are not added by other components of the face-detection system, with the consequence that the representation does not end up carrying a definitive 'face' tag. As a result that 'face' may not appear in our perceptual experience of the car, despite parts of this modular system having been unlocked.

Modular accounts of the mind have been associated with Fodorian encapsulation, where each cognitive system has access only to limited inputs (Fodor 1983, 1984), with the result that perception has been described as a cognitively impenetrable process. The modular architecture proposed by Carruthers and Barrett does not entail this kind of cognitive impenetrability. Modular systems can only operate on particular representations, but they have access to a large pool of representations. This view respects the modular encapsulation of processing, while opening the way to cognitively penetrable perception, because many cognitive processes can affect the contents of perception. Moreover, the modular systems are learning systems, with the result that their processing is sensitive to environmental parameters (see below). It is possible to argue that perception has a modular architecture without the commitment to the cognitive impenetrability of perception.

### **2.3. Binding concepts to perceptual experiences**

What does it mean for a concept to be 'bound' into a perceptual experience, and does this amount to that concept being seen? Answering these questions first requires addressing another, namely, what is the relation between our phenomenal perceptual experience and the sub-personal 'tagged' representations?

Carruthers, along with many other theorists, maintains that conscious experiences are the outputs of cognitive systems that are 'globally broadcast' (Baars, 1988; Carruthers, 2006; 2011). Global broadcasting is the ultimate bulletin board: information that is globally broadcast is available to a wide range of processing systems, and is available to verbal report and to guide decision-making and reasoning. A distinction needs to be made between Barrett's 'bulletin board' system and the 'global broadcast' concept used by Carruthers. Bulletin-boards are intended to describe a particular kind of cognitive architecture, and there are likely multiple systems organised in this way within cognition. Object representations that are placed on the visual bulletin board to be accessed by a sub-set of cognitive systems (those evolved to process representations of that kind) are unavailable to the gustatory systems, which will have their own bulletin boards. Those representations that are globally broadcast are the culmination of multiple bulletin-board structured

processes, each of which will have tagged the representations and passed it onto another pool to be examined by other processes. The tagging processes are not conscious: what is conscious is the percept that makes it into the global broadcast when all the modular tagging has taken place.

We are now in a position to answer the first question, namely, when a percept is globally broadcast, how is all the information provided by the tagging process presented in experience? Both Carruthers and Barrett accept that the concepts that are bound into a visual perceptual experience need not themselves be visual. Barrett carefully works through how the tag 'predator' can be attached to a perceptual experience of a lion (2005, pp.278 – 80); and Carruthers uses a similar example in detecting cheaters (free-riders) (2005, p.221), but neither 'predator' nor 'cheater' is a visual category. The claim is that the percept can be conjoined with certain thoughts that come as a result of sub-conscious processing, with the result that the perceptual experience is inseparable from these additional thoughts:

'... *cheat* isn't a perceptual category. But why shouldn't that concept nevertheless become attached to a perceptually represented and globally broadcast item? In effect, what would be globally broadcast would be a perceptual item conjoined with the thought, *that is a cheat*, where the indexical 'that' refers to the perceived person in question.' (Carruthers 2005, p.220).

One might read this as an experience containing a percept with some additional thoughts attached. However, I do not think this does justice to Carruthers' view. Terms like 'bound' and 'conjoined' are, I think, intended to convey the strength of the connection between the perceptual experience and its conceptual contents, while nevertheless keeping the concepts and percepts constitutively distinct. It is misleading to construe this relation chronologically, as a perception preceding the conjoined thoughts. The two can co-occur, and result in a unified perceptual experience, without being constitutively the same. As a result, we perceive disappointment, agency, desire, and other psychological states.

#### **2.4. Extra-perceptual inferences?**

On this modular account of vision the distinction between ‘perceptual’ and ‘non-perceptual’ processes is not clear cut, and this causes problems for G&V’s ‘perception-plus-extra-perceptual-inference’ criticism. One might argue that, if the mindreading system is one of the conceptual systems which has access to the perceptual bulletin board and can ‘bind’ concepts into representations, then there is an ‘extra-perceptual step’ of the kind G&V warn against. But in many ways the bulletin board model looks very much like smart perception, where the process of producing a visual percept involves a number of different cognitive systems. With so many systems involved, what is the value of saying that any one of them is ‘extra-perceptual’?

G&V may reply that any process that adds conceptual information to the initial ‘object representation’ is extra-perceptual. For this to be the case, the object representation must come about when ‘my eyes are working fine; my visual cortex is processing all of the visual information, but what vision delivers is relatively meaningless “thin” behaviour’ (ibid p.192). The object representation is thin to be sure; so thin that it’s unlikely to even deliver ‘behaviour’. However, one might wonder whether the ‘object representation’ is a product of a fully-functioning visual cortex as it’s not obvious that someone with only the apparatus to get to this point would be considered sighted. The systems that process the object representation evolved to deal with visual information: to scan a specific bulletin board – the visual object representation board – and to process all and only those representations that fit their processing system. In this way they are as much a part of the visual processing system as the retina and the processes detecting texture, colour, movement etc. that yield ‘object representations’. There is insufficient information in the bare object representation to facilitate an appropriate response to the environment; and a creature incapable of generating an appropriate response to its environment based on visual information about that environment might reasonably be described as blind.

Alternatively, one might argue that as the thought ‘that is a cheat’ is conjoined with a perceptual experience, the thought is ‘extra-perceptual’, and that therefore the DSPH is distinct from a modular account. But this runs into the same problems rehearsed above. If one has a perceptual experience unaccompanied by any thought at all about what was perceived, then one would be incapable of responding to it in an

appropriate way. While such experiences are conceivable, it is not obvious what visual information about the environment they would convey to the organism, and some justification is required for why such an experience would count as perception. The DSPH and modular mindreading accounts are agreed that perceptual experiences involve quite abstract information about what is perceived, and they are agreed that this information is incorporated into the visual process at a sub-personal level. The difference boils down to whether a given cognitive process is perceptual or extra-perceptual, and I'm not sure what hangs on this distinction.

If something like the above modular account works, then, as G&V observe, 'mindreading just is the perception of mental states' (ibid p.9). They imply that this is a position which mindreading theorists would not wish to be aligned with, although it's not clear that Carruthers, the epitome of a mindreading modular theorist would object. There are questions to be asked about the kinds of mental state that can be perceived which I will return to presently, but 'a cascade of increasingly abstract concepts' suggests the scope of perceivable mental states could be quite broad. However, G&V try to block this move for the modularity theorist by pointing to the plasticity of neural structures and cross-cultural differences in mental state recognition, arguing as follows:

1. If the perceptual process includes systems that allow mental state concepts to be bound into the perception, then mindreading just is the perception of mental states.
2. Mindreading uses folk psychological concepts of mental states.
3. Sub-personal perceptual processes cannot use folk psychological mental state concepts ('it's not clear how such [sub-personal] processes would be related to folk psychology' 2014, p.192).
4. Therefore, a mindreading module cannot be a part of the perceptual process.

In what follows, I question the strength of premises 2 and 3 on two fronts. The first relates to G&V's concerns that a modular mindreading view does not sufficiently account for learning and cross-cultural differences in mindreading ability. I explore this in section three. The second pertains to the scope of role of mindreading, and whether its explanandum is solely folk psychological states. This opens up more

interesting questions about the scope of mindreading views to date and whether we can perceive non-folk psychological mental states. This will be the focus of section four.

### **3. Cultural variations in mindreading**

#### **3.1. Modularity and cultural variation**

It is well-documented that your experiences and culture can affect your perceptions of others, and G&V maintain that the DSPH explains this better than a modular account of mindreading. They point to two phenomena: that there are ‘cultural variations in brain mechanisms specifically underlying person perception’ (p. 193); and the ‘dehumanisation’ phenomenon, describing how we find it easier to empathise with and recognise the intentions of those similar to us than those who are not. Their claim is that mindreading views ‘ignore the possibility that sensory-motor areas have undergone plastic modification in prior experience’ (ibid) and so cannot account for these phenomena.

Modules, as conceived in contemporary accounts of a modular architecture, are learning systems (Carruthers 2005; 2011). They function so as to extract and process information of a certain type from the environment, but the token information they store will be dependent on the organism’s environment. A modular system which evolved to detect kinship will clearly pick out different people as kin for each individual. Such a module will be sensitive to cues that indicate kinship, e.g. presence in early childhood, similarity in features, even hormonal features, so that the token individuals that are picked out as fitting this type will vary for each person. Modular systems have particular types of outcome as their developmental target, but the token outcomes for each individual will be subject to that individual’s environment (Barrett 2006). By distinguishing developmental type- outcomes from token manifestations of those types in individuals, one can see how cognitive systems of a particular type can be universal while their outcomes are unique to the individual.

One can therefore posit a cognitive system with the function of detecting in- and out- group members (type of outcome) and which passes the information on to a



mindreading cognitive system which uses this information to determine how one should interact with the perceived person. If the person is tagged as 'out-group', then the mindreading system may alter how much one empathises with them. It is plausible that the two systems, the 'out-group detection' system and the 'mindreading' system should evolve to interact with each other, as there are evolutionarily significant advantages in avoiding empathising with out-group members. This is supported by the large body of data concerning the phenomenon of 'dehumanisation' cited by G&V. Modular systems may be functionally discrete, but the information they process affects how other systems process a representation, and one would expect out-group member representations to be tagged in a way that affects how the mindreading module works on them. This is another example of a non-perceptual concept, 'out-group', being bound into a perceptual representation, which affects how one experiences and responds to a perception of a particular person.

The module that functions to detect out-group members will be realised in a similar brain area across humans, but again the token instantiation will be different for individuals. The cues that pick out 'out-group members' will pick out different people depending on your environment and culture. Modules are learning systems; it is through interaction with the environment that the 'out-group detecting' modules can become more precise and form more detailed templates for out-group members in a particular environment. Finally, note that these modules need not rely on folk psychological concepts. The cues that the modules are sensitive to, e.g. slightly different facial structure, may have no folk- analogue. Uncovering these cues is an empirical issue to be settled through controlled experiments. One cannot simply introspect them based on common-sense generalisations.

### **3.2. Cultural similarities in mindreading**

While token developmental outcomes of particular cognitive systems can vary across individuals, there appear to be some outcomes which remain the same across cultures. Specifically, infant behaviours in violation of expectation tasks that involve mindreading have been shown to be robust across a variety of cultures. Tara Callaghan and colleagues (2005) conducted verbal false belief tasks equivalent to the

Sally-Ann task (Baron-Cohen & Leslie, 1985) in five different cultures: Canada, India, Peru, Samoa and Thailand. Their findings were comparable to those of the original verbal false belief tasks: children aged three years or under failed the task while children of five years and over reliably passed it. Callaghan and her colleagues conclude:

The fundamental shift in understanding the impact of a false belief on behavior appears to be a universal milestone of development that occurs between 3 to 5 years of age. Synchrony in the age at which children of diverse cultures pass the false belief task undermines the claim that particular cultural views, such as a Western concept of mind, profoundly influence this very basic aspect of early mental state reasoning, and strengthens a claim of universality. (2005, p. 381)

More recently, H. Clark Barrett and colleagues (2013a) took three types of false belief tasks to a variety of field sites. Two differences between the chosen sites and North American/European (Western) cultures are salient. First, in two of the populations, the Shuar of Ecuador and the Yasawan of Fiji, open talk about psychological states is not common. For instance, the authors note that the Yasawan are not unwilling to talk about inner mental states, but rarely do so and 'indeed regard very young children as incapable of thought or feeling' (Barrett, et al., 2013b, p. 5). Children in these communities are therefore less exposed to explanations of behaviour that reference psychological states than Western children. Second, the role of children in these communities and their interactions with adults, are quite different from those of Western children. The Salar in Qinghai Province, China, 'do not encourage young children's independence... Children are typically not asked their views or opinions for decision making' (ibid p.6); and children in the Shuar community have much less contact with adults than Western children, as they are typically cared for by older siblings. Barrett and colleagues examined whether children in these traditional non-Western communities passed false belief tasks at a comparable age to Western children, despite the clear differences in experience they had with regards to mental state explanations and interactions with adults.

Barrett and colleagues tested infants and children on a variety of location-change verbal and non-verbal false belief tasks, but importantly each task measured implicit rather than elicited responses from the participants (for details see Barrett et al. 2013b). The verbal false belief tasks measured where children (average age 40 months) looked when the protagonist returned to retrieve the object they had left, and found that their looking behaviour matched that expected of a someone who can attribute false beliefs to others (i.e. they look to where the protagonist originally left the object rather than to where it has been moved to in her absence). They also conducted a non-verbal violation-of-expectation task, similar to Onishi and Baillargeon's landmark 2005 study, with Shuar and Salar infants (average age 22.5 months). The results were comparable to those presented by Onishi and Baillargeon: infants looked longer when the protagonist acted in a way that did not conform with her beliefs, be those beliefs true or false.

Although there are significant differences in how infants are cared for, these infants are still in environments where people typically look for objects where they last left them, and where reaching for an object is perceived as wanting that object. So a cognitive system which evolved to tag certain movements with information about another's psychological states has the same 'raw data' to work from in these different environments, which explains why the token developmental outcome in this limited subset of mental state attributions is the same for infants across the world (contrary to the token developmental outcome for an 'out-group detector').<sup>6</sup>

There are now two claims on the table. The first is that there are social cognitive capacities that we expect to manifest quite differently across groups as a result of an individual's environment. Examples include who is recognised as kin, or an out-group member, or a predator, where the developmental outcome type is stable across cultures, but where the interaction between the environment and the cognitive systems in question shape quite different token outcomes. So while one would expect individuals across the world to distinguish in- from out- group members, the token

---

<sup>6</sup> It really is a limited subset of mental state attributions. When it comes to verbal measures, all manner of variables can affect performance, from whether one has older siblings (Ruffman et al. 1998) to how frequently mental state vocabulary is used around a child (Rakoczy, 2010).

manifestation of that ability, including the neural realisation, will be unique to each individual. The second is that there are other abilities, such as the ability to predict how another's false belief will affect her behaviour, where the difference between token outcomes is slight: infants across the world respond in the same way. These data are not in conflict, contrary to G&V's construal:

Others [...] maintain that the pattern of ToM development is identical across the species, which is in marked contrast to the uneven and culturally dependent development of many other capacities. Evidence from studies of dehumanisation, however, is inconsistent with these expectations, and shows that mechanisms of social cognition are constitutively dependent upon historical-cultural situatedness and group membership. (p.194)

The development of the majority of cognitive modular systems is fundamentally dependent on the environment, as explained above. However, there are some whose manifestation appears to be robust across quite different environments, such as the ability to respond appropriately to another's false belief about an object's location. This is not a difficult outcome for modular views to accommodate: it simply mirrors the fact that certain psychological states affect behaviours in the same way across cultures, resulting in similar token-outcomes. We expect to see quite different token-outcomes for other modular systems though, like predator detectors and out-group detectors, as environments have different predators and different individuals marked as 'out-group'.

### **3.4. Interim conclusions**

The DSPH has been promoted as a better explanation of cross-cultural differences in social cognition than traditional mindreading accounts. This section has argued otherwise, maintaining that the data can easily be accommodated by a modular account of mindreading, provided one properly understands the difference between type developmental outcomes and token instantiations in individuals. Learning and experience are central to contemporary modular views, as modular systems build their data-sets and templates from the organism's interactions with the environment. This contrasts against traditional Fodorian modules where there is no role for learning in a

modular architecture. Advocates of the DSPH cannot assume that a challenge against Fodorian modules is a strike against all modular accounts, and need to tailor their arguments accordingly. I have not done very much to defend the modular mindreading approach independently of the DSPH's challenges; my aim was simply to show that proponents of the DSPH need to do more to show that their explanation of cross-cultural phenomena is better than that offered by contemporary modular approaches.

#### **4. Perceiving mental states**

In this last section of the paper I turn to the more open question of what kind of information about another's mental state can be 'bound' to our perceptions of them.

##### **4.1. Perceiving non-folk psychological mental states**

If one accepts that the behaviour of pre-linguistic infants in non-verbal false belief tasks is best explained by claiming that they attribute psychological states to others, then one must be clear on the nature of psychological states being attributed.<sup>7</sup>

Kristine Onishi and Renée Baillargeon maintain that infants 'possess (at least in a rudimentary and implicit form) a representational theory of mind' (2005, p. 257). This has been challenged by Ian Apperly and Stephen Butterfill (2009; 2013; Apperly, 2010) who suggest that infants attribute relational rather than representational states to others. Onishi and Baillargeon's account is closer to Western folk psychology, where beliefs and desires are understood as propositional, representational states; whereas there is no clear folk psychological analogue to the relational states proposed by Apperly and Butterfill. This raises a question: could such non-folk psychological mental state concepts be part of one's perceptual experience of another person?

Non-folk psychological information plays an important role in generating our perceptual experiences of other's psychological states, just as information we have no sense of at a personal level of cognition shapes our perceptions of the world. This was demonstrated in Hess's famous study demonstrating that when men were shown two photos of a woman, where the only difference between the photos was that her pupils

---

<sup>7</sup> Not everyone accepts this, see e.g. Heyes, forthcoming; Perner and Ruffman, 2005. Their concerns, while legitimate, are tangential to the aim of this paper.

were slightly larger in one picture than in the other, men found the photo with the larger pupils more attractive (Hess, 1965). The perception of attractiveness was affected by information that the men had no introspective access to. Similarly, one's perception of something as an agent can be facilitated by sub-personal information about psychological states which bears little or no resemblance to folk psychological concepts. This means that one can see something as an agent, and respond to it appropriately, without having an understanding of psychological states that is consciously accessible. And this is where Apperly and Butterfill's non-folk psychological concepts may contribute to our perception of others. The infant's mindreading system may use relational psychological state information to mark a particular representation on the visual-system's bulletin board, and this in turn affects how the infant perceives the other and how she responds to them. The information can affect a perception without being consciously accessible, with the consequence that non-folk psychological information contributes to the perception of another as an agent even though such information can never be introspected.

#### **4.2. Responding vs. perceiving**

When a thought about another's psychological state becomes bound to our perception of them, does this amount to seeing the psychological state? The binding process certainly affects the phenomenology of the perception, and it affects how one responds to what is perceived. But does it make the other person's mental state observable?

There are situations where an organism responds appropriately to some feature of the environment without seeing that feature. For example, imagine a vertical flag with the bottom half occluded (I owe this example to Till Vierkant). The occluded end of the flag can be moved by two forces: wind or water. A trained observer can tell from the movement of the visible part of the flag whether it is being moved by water or wind (let's assume she is always accurate in her judgment). She has distinct phenomenological experiences when she sees the flag move by wind or water, and she doesn't experience any inferential process when she judges that there is wind or water present. One cannot tell from her behaviour whether her response to the flag stems

from her seeing only the flag, or the flag and the wind/water. But, it does not seem correct to say that she sees the water or the wind, and there is a clear distinction between the cause of the flag's movement and the flag itself. This seems to be a case where there is distinct phenomenology to seeing the flag *as* moved by wind or water, and where the perceptual experience yields an appropriate response to the wind or water, but where the wind/water is not seen.

On a modular view, the perception of the flag is conjoined with a sub-personal thought 'moved by water'. This conjunction, while contributing to the phenomenal perception of the flag, does not amount to seeing the water. A similar story runs for psychological states: when the percept of behaviour is bound to a particular psychological state concept, e.g. 'she wants the cake', the psychological state concept affects the phenomenology of the perception of the behaviour, but it does not amount to seeing the state. Information about another's psychological state affects one's perception of them, which in turn affects how one responds to that perception. But one need not interpret this as seeing the psychological state.

On the other hand, it's not clear what the analogue to removing the flag's occluder would be in the case of psychological states. We only have first-personal access to our psychological states, and it's not obvious what more one could reveal to an external observer so as to match the analogy with the flag. While there may be arguments about whether another person can see our psychological states in our behaviour, it seems as though what is provided in perceptual experience is an extremely reliable source of knowledge about psychological states.

## **5. Conclusions**

The DSPH was introduced as an alternative to simulation and information rich, theory-based accounts of mindreading. This paper focussed on the contrast between the DSPH and one particular information-rich account of mindreading, namely, the Barrett and Carruthers' modular approach. The accounts share more common-ground than advocates of the DSPH have previously admitted. First, they both accept that other people's mental states are part of our experiences of them. Second, they both maintain that our perceptions of other people are not cognitively impenetrable. That

the modular approach has been associated with this claim is unsurprising, as the cognitive impenetrability of perception was a central thesis in Fodor's modular architecture (see also Fodor 1984). However, subscribing to a modular architecture does not entail subscribing to Fodor's modular architecture, as evidenced by the account discussed here. On this account, perception can be understood as a modular process, but one which has access generality with processing specificity. That is, while there may be discrete processing systems for recognising certain aspects of the environment from visual information, these systems have access to a large 'pool' of representations. Each system can process and 'tag' relevant representations, changing their structure slightly which then affects how they are processed by other cognitive systems. This is a parallel processing exercise, with representations being placed back into the pool multiple times after each transformation for the cognitive systems to operate on. The visual percept that is globally broadcast, that is, the percept that is conscious, is conjoined with multiple information tags, including information about another's mental states. This modular architecture allows for numerous cognitive processes to contribute to the perceptual experience. This is the exact advantage that advocates for the DSPH maintained their view had over traditional accounts. A modular architecture does not preclude many cognitive processes from informing perception, and it is misleading to present it as a 'foil' to the DSPH.

One must therefore look elsewhere for a distinction between modularity theory and the DSPH. The DSPH maintains that it does not involve any 'extra-perceptual inferences' at a sub-personal level, in contrast to its opponents. This opens up two questions: first, whether any sub-personal cognitive process can properly be described as inferential; and second, the criteria for describing a sub-personal process as a perceptual one. Answering the first question lies beyond the scope of this paper; in response to the second, I have argued that the cognitive processes which 'tag' the initial representations of visual information should be considered perceptual. Without these processes one does not have a perception that is of any use. Advocates of the DSPH would at this point maintain that the modular view is committed to perception being 'not-so-smart', requiring extra perceptual supplements. However, those processes that the DSPH cites as contributing to perception, e.g. activity in the



fusiform, are exactly the same as those the modularity theorist draws on to explain how perceptual experiences gain their content. Both accounts accept that fusiform activity is essential to the perception of faces.

The second part of the paper addressed whether folk psychological states can be part of the perceptual experience. On Carruthers and Barrett's view of modularity, modules are learning systems. They can therefore explain phenomena like dehumanisation and cross-cultural differences in certain social cognitive practices. Importantly, they can also account for important similarities in social cognitive skills. Modular views do not preclude folk psychological concepts from being part of a perceptual experience, nor do they preclude the effects of culture and environment on an organism's perception of others.

However, it is important to note that the sub-personal processes that facilitate experiences of others as agents may not always use folk psychological states. It may be the case that these processes use relational psychological states as proposed by Apperly and Butterfill. Testing this hypothesis is an empirical task, however, the fact that relational states do not appear in our perceptual experiences of others does not mean that they do not contribute to the perception of the other person as an agent. An organism with folk psychology may be able to describe her experiences to herself and others using a particular framework, and this framework may in turn affect her later perceptions of other's behaviours. However, such a grasp of folk psychology may not be necessary to perceive others as psychological agents, and for these perceptions to guide our behaviours in ways that are sensitive to the other's psychological states. This is supported by the uniform response of infants from quite different folk psychological cultures in non-verbal false belief tasks. Responding appropriately to another's intentions does not require a folk psychological grasp of their psychological states.

Finally, it is possible for another's intention to be part of one's perceptual experience of them without that intention being seen. The modular account presented keeps psychological states distinct from behaviours while claiming that they can be 'bound into' perceptual experiences. There are further questions to explore

about the nature of this ‘binding’, and about the perception of mental states more generally. The nature of ‘goal states’ is particularly interesting, as there is currently some debate about whether goal states are psychological states (Butterfill & Apperly, 2013, p. 614), which in turn affects how they may be perceived. Whatever your preferred theory of social cognition, it must explain our perceptual experiences of other people’s psychological states: the aim of this paper was simply to show that the DSPH is not the only theory capable of meeting this explanatory desideratum.

Word count 8,830

## **Acknowledgments**

This paper benefitted from discussion at the 2013 workshop ‘The scope and limits of direct perception’ hosted by the Center for Subjectivity Research at the University of Copenhagen. I’m also grateful to Till Vierkant and Alistair Isaac for patient and helpful discussions.

## **Bibliography**

- Apperly, I. (2010). *Mindreading*. Psychology Press.
- Apperly, I., & Butterfill, S. (2009). Do humans have two systems to track beliefs and belief-like states? *Psychological Review*, 116, 953 - 970.
- Baars, B. (1988). *A Cognitive Theory of Consciousness*. Cambridge University Press.
- Baars, B. (1994). A Global Workspace Theory of Conscious Experience. In A. Revensuo, & M. Kamppinen (Eds.), *Consciousness in Philosophy and Cognitive Neuroscience* (pp. 149 - 172). Lawrence Erlbaum.
- Baron-Cohen, S., & Leslie, A. (1985). Does the autistic child have a theory of mind. *Cognition*, 21, 37 - 46.
- Barrett, H. (2005). Enzymatic computation and cognitive modularity. *Mind and Language*, 20, 259 - 287.
- Barrett, H. (2006). Modularity and design reincarnation. In P. Carruthers, S. Stich, & S. Laurence (Eds.), *The Innate Mind vol. 2* (pp. 199 - 217). O.U.P.
- Barrett, H., Broesch, T., Scott, R., He, Z., Baillargeon, R., Wu, D., et al. (2013). Early false-belief understanding in traditional non-Western societies. *Proceedings of the Royal Society B*, 280, 1-5.

- Barrett, H., Broesch, T., Scott, R., He, Z., Baillargeon, R., Wu, D., et al. (2013b). Early false-belief understanding in traditional non-Western societies [Supplementary material]. *Proceedings of the Royal Society B: Biological Sciences*, 280, 1-29.
- Butterfill, S., & Apperly, I. (2013). How to construct a minimal theory of mind. *Mind and Language*, 28.
- Callaghan, T., Rochat, P., Lillard, A., Claux, M., Odden, H., Itakura, S., et al. (2005). Synchrony in the onset of mental state reasoning. *Psychological Science*, 16, 378 - 384.
- Carruthers, P. (2004). The mind is a system of modules shaped by natural selection. In C. Hitchcock (Ed.), *Contemporary Debates in Philosophy of Science* (pp. 293 - 311). Blackwell.
- Carruthers, P. (2006). *The Architecture of Mind*. Oxford University Press.
- Carruthers, P. (2011). *The Opacity of Mind*. O.U.P.
- Carruthers, P. (2013). Mindreading in Infancy. *Mind and Language*, 28, 141 - 172.
- Davies, M., & Stone, T. (2001). Mental Simulation, Tacit Theory and the Threat of Collapse. *Philosophical Topics*, 29, 127 - 173.
- Fodor, J. (1983). *The Modularity of Mind*. Bradford Books/M.I.T. Press.
- Fodor, J. (1984). Observation reconsidered. *Philosophy of Science*, 51, 23-43.
- Gallagher, S. (2005). *How the Body Shapes the Mind*. Oxford: Clarendon.
- Gallagher, S. (2008a). Direct perception in an intersubjective context. *Consciousness and Cognition*, 17, 535-543.
- Gallagher, S. (2008b). Inference or interaction: social cognition without precursors. *Philosophical Explorations*, 11, 163 - 174.
- Gallagher, S., & Varga, S. (2014). Social constraints on the direct perception of emotions and intentions. *Topoi*, 33, 185 - 199.
- Gallagher, S., & Zahavi, D. (2008). *The Phenomenological Mind*. Oxford: Routledge.
- Goldman, A. (1992). In defense of the simulation theory. *Mind and Language*, 7, 104 - 119.
- Hess, E. H. (1965). Attitude and pupil size. *Scientific American*, 212, 46-54.
- Heyes, C. (Forthcoming). False belief in infancy: a fresh look. *Developmental Science*.
- Johnson, S. (2000). The recognition of mentalistic agents in infancy. *Trends in Cognitive Sciences*, 4.
- Johnson, S. (2005). Reasoning about intentionality in preverbal infants. In P. Carruthers, S. Laurence, & S. Stich (Eds.), *The Innate Mind vol.1: Structure and Contents* (pp. 254 - 271). O.U.P.

- Lavelle, J. S. (2012). Theory-theory and the direct perception of mental states. *Review of Philosophy and Psychology*, 3, 212-230.
- Onishi, K., & Baillargeon, R. (2005). Do 15 month-old infants understand false beliefs? *Science*, 308, 255-58.
- Perner, J., & Ruffman, T. (2005). Infants' insight into the mind: how deep? *Science*, 308, 214-216.
- Rakoczy, H. (2010). From Thought to Language to Thought: Towards a Dialectical Picture of the Development of Thinking and Speaking. *Grazer Philosophische Studien*, 81, 77 - 103.
- Ruffman, T., Perner, J., Naito, M., Parkin, L., & Clements, W. (1998). Older (but not younger) siblings facilitate false belief understanding. *Developmental Psychology*, 34, 161 - 174.
- Wellman, H. (1992). *The Child's Theory of Mind*. Cambridge, Mass: M.I.T. Press.