A scientific case for conceptual dualism: The problem of consciousness and the opposing domains hypothesis.

Anthony I. Jack

Department of Cognitive Science, Case Western Reserve University

To appear in:

J. Knobe, T. Lombrozo & S. Nichols (Eds.)

Oxford Studies in Experimental Philosophy (Vol. 1)

Oxford University Press.

Abstract

In recent years, a number of scientists and philosophers have suggested that the psychological and neural sciences provide support for, and are committed to, reductive physicalism – the view that all aspects of the mental are best explained by the physical processes of the brain. Here I suggest a different view. Emerging research in neuroscience and psychology suggests a dualism in human understanding. Our capacity for understanding physical processes appears to be in fundamental tension with our capacity for thinking about the inner mental states of others. In this essay, I first review evidence for a divide in our neural structure which maps onto thinking about minds versus thinking about the mechanical properties of bodies. This divide is intriguing; however it falls short of actually explaining why we perceive difficulties for integrating these two types of understanding. I then introduce a bold hypothesis – that our neural structure constrains our thinking in a way that limits our ability to integrate these two types of understanding. This hypothesis was generated to explain one perceived problem, the apparent existence of an explanatory gap, and makes novel and falsifiable predictions. I then review behavioral and neuroscientific evidence which confirms these predictions and extends the model to address other related issues, including motivational factors associated with belief in ontological dualism. By demonstrating that this theoretical framework yields testable predictions, these findings lend support to the bold hypothesis. I conclude by exploring some theoretical and practical implications of the hypothesized dualism in human understanding.

1. Introduction

> *Like a handful of others (Bloom, 2005; Dawkins, 2006; Dennett, 2006; Harris, 2004), I believe that our intuitive dualism causes a lot of problems… [T]he mission of social neuroscience, as the offspring of social psychology and neuroscience, is to understand all of human subjective experience in physical terms. The rise of social neuroscience is the demise of the soul.* (Greene, 2011)

Greene (2011) argues that one of the most exciting and socially significant products of neuroscience will be a philosophical conclusion. A scientific reductionist approach will be seen to triumph, banishing the misbegotten concept of the soul. This view is certainly tempting to those of us who are enamored with the power and promise of science. But is it right? I don't think so. Philosophy has long suggested difficulties for this brand of scientific triumphalism (Jack and Robbins, 2004). But the account I offer here does not rest on philosophical argument. I am going to show you that the science itself suggests reductionism will fail in its attempts to elucidate all aspects of the mental.

In forwarding this account, I don't disagree with Greene's claim that intuitive dualism creates problems. I just don't think these problems can be avoided. I think intuitive dualism tells us something quite fundamental about the nature of human understanding, a feature of our psychology which has important implications for how we should best approach the science of the mind. One reason for supposing this is the long history of influential thinkers who have arrived at the conclusion that scientific thinking fails, in some way or another, to capture important aspects of the mental.

> *Moreover, we must confess that the perception, and that which depends on it, is inexplicable in terms of mechanical reasons, that is, through shapes and motions. If we imagine that there is a machine whose structure makes it think, sense, and have perceptions, we could conceive it enlarged, keeping the same proportions, so that we could enter into it, as one enters into a mill. Assuming that, when inspecting its interior, we will only find parts that push one another, and we will never find anything to explain a perception. (Leibniz, 1714)*

As the quote from Leibniz illustrates, the sense of a disconnect between experiential aspects of the mind and mechanical explanations has been present ever since science, as we currently understand it, took shape in the scientific revolution. In the meantime, scientific explanation has advanced considerably – we now appeal to electrochemical mechanics rather than classical mechanics to explain neural function. Philosophical conceptions of the mind-body problem have also become considerably more nuanced, yielding a variety of modern forms most frequently referred to as the problem of consciousness. The core of this problem is widely acknowledged to be captured by the existence of a (real or apparent) 'explanatory gap': a disconnect between our understanding of neural mechanisms and our understanding of experiential mental states (Levine, 2000). The explanatory gap has been made vivid by a variety of a thought experiments (e.g. Nagel, 1974, Jackson, 1986, Chalmers, 1996). However, I won't dwell on the details of these arguments, which are in many cases highly nuanced and technical in nature. I think it is tremendously important that we take the problem of consciousness seriously – that we consider the problem in a way that Greene and many other scientists eschew. But the way I want to take it seriously here is not the way that may be most familiar – the philosopher's approach of carefully unpacking the arguments and their consequences.

For many scientists, the idea of focusing on an apparently intractable philosophical problem looks like a waste of time – science is about empirical investigations that deliver the answer to a question. This sense of impatience has encouraged many to recast the philosophical problem of consciousness into a more empirically tractable form. Hence, we may aim to use neuroscience to illuminate the biological basis of consciousness (as suggested by a Science magazine editorial, Miller, 2005), or "understand all of human subjective experience in physical terms" (Greene, 2011). The danger here is that, in the rush to make the problem tractable, we may lose sight of the real problem altogether.

I see the scientific significance of philosophy in a different way. On this view, the most scientifically valuable philosophical problems are exactly the ones which appear most intractable. Such problems likely reflect a fundamental tension between two ways of understanding. Cognitive science studies how we understand the world. So, if we have located a genuinely insoluble philosophical problem, we can expect to see the tension reflected in our cognitive and neural structure. The problem of consciousness may be such a genuinely insoluble problem. So here I will resist trying to make the problem tractable, and instead entertain the

hypothesis that it isn't. Instead of trying to find the biological basis of consciousness, I aim to identify the biological basis for the *problem* of consciousness.

The view that emerges will, I suggest, transform our understanding of the problem and how to deal with it. It will cast the problem of consciousness in a broader context. According to this view, our neural structure constrains our thinking, giving rise to a fundamental division in human understanding. I will suggest that this divide is relevant to a variety of philosophical issues, and to methodological issues that have long plagued the history of psychology. The brain is, of course, a mere organ, a product of evolution. Your arm can move in many ways that allow it to do many different sorts of things. Evolution has created a remarkably flexible and effective structure. But that structure still has constraints. If you have an itchy back, your arm is poorly designed to help with that. You do much better if you give up on contorting your arm, go make nice with someone and ask them to scratch your back instead. Similarly, I believe that when it comes to understanding the mind, the reductionist approach needs help. The reductionist approach is tremendously flexible and effective. It has solved many problems and it will solve many more. Yet, according to this view, it just isn't suited to scratching the itch of consciousness - mechanistic explanations will never enable us to fully understand human experience.

2 The divided mind

2.1 Systems for social cognition

Work in social cognitive neuroscience suggests three major categories of regions involved in social information processing:

First, there are systems which respond to social stimuli, regardless of the task (Wiggett et al., 2009). These lie close to occipital and temporal areas involved in basic visual and auditory processing, and form parts of the ventral visual stream (Goodale and Milner, 1992). They include some well delineated functionally specific areas such as the fusiform face area on the ventral surface (Kanwisher et al., 1997) and the extra-striate body area on the lateral surface (Downing, 2005); as well as a complex of regions centered on the posterior portion of superior temporal sulcus. The latter regions are not, as yet, so tightly defined, but they are reliably associated with gaze processing, action observation and decoding of emotional expressions (e.g. Pelphrey, 2005).

Second, there is a system involved in understanding the transitive actions of others (e.g. manual manipulations of objects). This system includes inferior parietal sulcus (including the inferior and anterior portion of intra-parietal sulcus) and frontal premotor cortices (including inferior precentral sulcus). This system is known as the mirror neuron system since it is presumed to be homologous to regions which contain mirror neurons in the monkey. This system includes parts of the dorsal visual stream (Goodale and Milner, 1992) and areas involved in motor planning and visuo-motor transformation (Corbetta et al., 1998). Resting connectivity is increasingly used to identify broad functional networks in the brain. The activity of regions in such a network is highly correlated when participants are not engaged in any task, suggesting that they form part of functionally coherent network (e.g. Honey et al., 2009). The mirror neuron system substantially overlaps a network defined by resting connectivity known as the dorsal attention network (Fox et al., 2006).

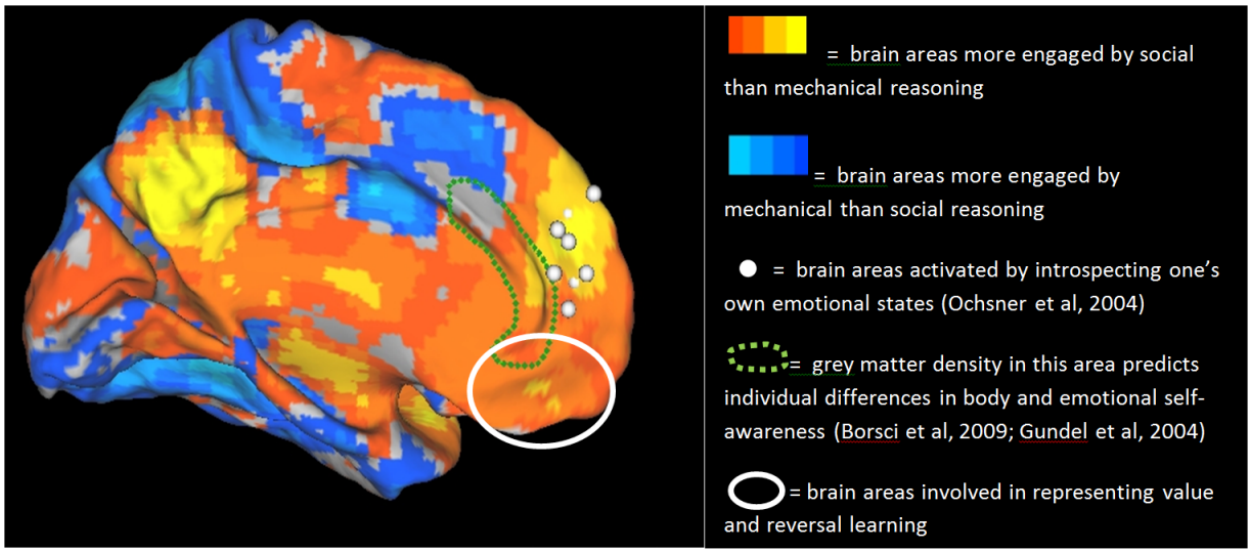

Figure 1: Illustration of the relationship between social cognition and other cognitive processes in medial prefrontal cortex. The figure depicts the medial surface of the left hemisphere. The red and blue coloring shows the contrast between social and mechanical reasoning tasks taken from (Jack et al., 2012).

Third, there is a network of regions known as the mentalizing system. This system is clearly distinct from the mirror neuron system (Van Overwalle and Baetens, 2009). The mirror neuron system is most engaged by tasks that involve either executing or watching transitive actions, in many cases in the absence of any larger social context and/or without being able to

see anything more than the hand/arm of the individual performance the action. In contrast, mentalizing tasks typically involve a richer social context or narrative and are frequently (but not always) textual in nature. Their defining feature is that they encourage the participant to think about the internal mental states of others (beliefs, desires, emotions). The mentalizing system includes two pronounced midline regions (Figure 1): dorsal MPFC and medial parietal / posterior cingulate (MP/PC); as well as a right lateralized region near the junction of the temporal and parietal cortices (rTPJ). The latter region lies adjacent to the complex of regions in superior temporal cortex described as part of the first system. In contrast to this first system, engagement of the mentalizing system is associated with the attribution of mental states regardless of whether the target/stimulus is animate (e.g. humans, animals, faces) or inanimate (e.g. robots) (Waytz et al., 2010). Hence, the first system described is more bottom-up - its activity is relatively insensitive to the task but highly sensitive to the type of stimulus - whereas the mentalizing system is more top-down – its activity is influenced heavily by the cognitive context and much less by surface characteristics of the stimuli.

The functions of the mentalizing network appear to be quite general across different types of thinking about minds, as co-activation of these regions is commonly observed during a variety of tasks involving social cognition (Amodio and Frith, 2006, Mitchell, 2009, Van Overwalle, 2009). These mentalizing regions occupy the majority of a second network that can be defined by resting connectivity, which is known as the default mode network (DMN) (Schilbach et al., 2008, Mars et al., 2012). Although this network is often engaged as a unit, there is evidence for some functional specialization of its regions. The midline regions (figure 1) are of particular significance for understanding the problem of consciousness, since standard formulations of the problem focus predominantly on consideration of our own internal phenomenally conscious states, rather than our understanding of the mental states of others. Many studies and meta-analyses have now solidly established that these midline regions are recruited both when we introspect our own current mental states (e.g. emotions) and when we attribute mental states to others (Ochsner et al., 2004, Amodio and Frith, 2006, Saxe et al., 2006, Denny et al., 2012, Schilbach et al., 2012). In contrast, the evidence does not support the view that either of the other two systems described above are involved in thinking about experience.

2.2 Systems for physical reasoning

How do these systems involved in social cognition relate to those involved in thinking about inanimate objects and physical properties/processes? There is good evidence for a broad cognitive division. In some cases, this has been established by experiments that use closely matched stimuli designed to examine exactly this distinction (Mitchell, 2002, Martin and Weisberg, 2003, Blos et al., 2012). In other cases, this has been established by meta-analysis (Van Overwalle, 2010). Just as with social reasoning, care must be taken to distinguish different types of physical reasoning and to consider the specificity/generality of function of the brain areas involved by reference to a larger literature.

First, with regard to more sensory areas, it is well established that there is an animate/inanimate distinction on the ventral surface, such that lateral areas (adjacent to and including the fusiform face area) are more engaged by animate / biological stimuli (including animals as well as human faces and bodies), whereas medial areas (adjacent to and including the parahippocampal place area) respond preferentially to inanimate objects. Competing theories of the functional specialization of the ventral surface also exist (e.g. Tarr and Gauthier, 2000). These predominantly focus on the role of these regions in specific perceptual processes, and hence undermine the notion of a processing distinction that relates to physical versus social reasoning in these areas. Nonetheless, the animate vs. inanimate view remains a dominant hypothesis (Grill-Spector et al., 2004). In particular, more recent findings suggest the animate/inanimate distinction on the ventral surface is innate and is not due to visual processing demands, because it can be observed using auditory stimuli in individuals born blind (Mahon et al., 2009).

Second, there is evidence for a domain general non-social reasoning system which comprises regions in lateral parietal and frontal cortex. This was evident even in early meta-analyses (Duncan and Owen, 2000). Later reviews extend the picture, demonstrating considerable overlap between regions which are recruited by a variety of non-social tasks, including visual attention, working memory, language, logical reasoning, mathematical reasoning, general problem solving and causal/mechanical reasoning tasks (Van Overwalle, 2010). Since this network is recruited by such a broad variety of tasks, and because so many of the tasks historically used by neuroscientists were non-social in nature, it came to be labeled the ‘task positive network’ or

TPN (Fox et al., 2005). The TPN overlaps two networks which can be identified by resting connectivity: the dorsal attention system and an adjacent network in lateral parietal and frontal cortices called the frontoparietal control network (Vincent et al., 2008). The task positive network is clearly distinct from the mentalizing system of the default network both in terms of anatomical location and in terms of the tasks which preferentially recruit them (Van Overwalle, 2010). The task positive network substantially overlaps the mirror neuron system (Van Overwalle and Baetens, 2009).

2.3 Philosophical brain mapping

If we wish to map the problem of consciousness onto the brain, then it appears that current work in neuroscience presents us with two promising candidates. However, neither mapping is without complication. Let us consider them in turn.

First, we might seek to map a perceived incongruity in thinking about minded (animate) versus purely physical (inanimate) objects onto the neural division seen in occipital/temporal cortex. This division in processing has already been suggested as accounting for some aspects of the explanatory gap and fueling belief in dualism (Fiala et al., 2012). This account helps to make sense of various low-level processing effects which clearly influence our attributions of mindedness, especially in infancy. For instance, increased attributions of mindedness plausibly explain why infants to show a greater interest in objects when they have face-like structures or appear self-animated (Johnson et al., 1998). This account also fits well with the observation that infants appear to have a profoundly dualistic view of the world, such that they are not surprised when humans break some laws of physics, whereas they are when inanimate objects do (Bloom, 2004).

On the other hand, it isn't clear how well this account fairs when it comes to explaining the more nuanced dualistic beliefs that are evidenced by adults. Adult reaction times, when making judgments denying experience, are still influenced by cues related to agency (Arico et al., 2011). However, in the actual choices they make, adults show less sensitivity to these surface features than to cognitive context (e.g. cultural beliefs about whether that kind of object has a mind). Further, philosophical arguments concerning the explanatory gap rely predominantly on

considerations of one's own experience[1]. However, as previously noted, the sensory-driven regions for agency detection in the occipital and temporal lobes are not implicated in introspection.

Fiala et al's account seems to represent part of the story, but falls short of offering a full account of the explanatory gap. This neural divide likely plays a role in the tendency to intuitive dualism present in infancy. It is plausible that it also contributes to the intuitions that reinforce intuitive dualism in adulthood.

It looks like we can get further with the second mapping (i.e. with the mentalizing and task positive networks) because in this case the social regions are clearly involved in thinking about experiential mental states, both our own and other people's. Nonetheless, there are a number of potential concerns with this mapping. Let's begin by considering some which I think we can dispose of with relative ease.

First, there may be a concern that the TPN is engaged in variety of forms of reasoning. Hence, like Van Overwalle (2010), we might label it as a system for general reasoning rather than a system for physical reasoning. Undoubtedly the TPN is involved in forms of reasoning other than physical reasoning; however this doesn't present a critical objection to our mapping. The critical point is just that reasoning about physical properties and mechanical processes is one of the functions of this network. At the same time, I believe we can be more specific than 'general reasoning' when characterizing this network. Anatomical location, overlap in task-related activation, and resting connectivity all suggest the TPN is built upon and re-uses primitive systems for perceptual processing and motor control, i.e. systems which originally evolved to perform spatial visual functions such as: moving the eyes, allocating attention, visual guidance of action, and motor planning. The frontal regions of the TPN are also heavily engaged by working memory tasks and their activity increases with the cognitive load of the task. I believe a good characterization of this network is that it is involved in *analytic-empirical-critical* thinking. This contrasts with the mentalizing network of the DMN which, while most reliably recruited by social tasks, is also implicated in some more *synthetic* forms of non-social

[1] This hasn't always been the case. It appears the tension with experience is currently perceived as the most compelling; however Brentano's problem concerns the naturalization of the Intentional. We touch on this briefly again later in the manuscript.

reasoning, such as insight problem solving (Subramaniam et al., 2009) and detecting broader patterns (Kroger et al., 2002). Of course, I am pointing here to very broad characterizations of function that describe wide-scale cortical networks. Future research will guide much finer subdivisions of the networks that can guide a more nuanced picture. My claim is merely that this is some coherence to these broad characterizations.

Second, a surprising feature of the neural divide we see between the mentalizing network and the TPN is that at least one type of social reasoning appears to fall within the domain of the TPN. There is evidence that mirror neurons specifically code the intentions of actions (Iacoboni et al., 2005). Hence it appears that certain aspects of reasoning about intentional actions may occur in the TPN rather than the DMN. But if they do, then it appears the 'intention' the mirror neuron represents is, in effect, a description of the mechanical manipulation which is to be performed on an object. This fits well with the characterization of the TPN presented above, and is consistent with a broad division between neural mechanisms for thinking about minds vs. mechanical processes.

Third, there may be a concern about where thoughts about the body should be situated relative to this neural divide. On the one hand, Descartes' characterization of the problem as the mind-body problem would seem to suggest thoughts about our body should be situated with physical reasoning. On the other hand, a great deal of recent research has focused on the notion of embodiment, and it is highly plausible that our interoceptive understanding of our own bodies plays an important role in social cognition. It is important to remember that Descartes was very interested in biology; a discipline which he argued should view the body first and foremost as a mechanism. Hence Descartes' formulation of the philosophical problem involved a quite specific conception of the body. This is quite distinct from our first-person understanding of our own bodies. Anatomical location, overlap in task-related activation, and resting connectivity all suggest the DMN is built upon and re-uses primitive systems for visceral/emotional self-awareness and self-regulation. The clearest story can be told about the MPFC (figure 1). Anatomical studies in the monkey show this region lies adjacent to areas which receive visceral inputs from the body (Ongur and Price, 2000). Correspondingly, the density of grey matter in areas of anterior cingulate cortex immediately adjacent to mentalizing regions predicts bodily and emotional self-awareness, depicted in Figure 1 (Gundel, 2004, Borsci et al., 2009).

2.4 Preparing for the leap

Several researchers have drawn a close analogy between folk psychology and folk physics because they both involve high-level processes such as abstraction, inference, model building, prediction, and the postulation of unobservable processes or states (Lewis, 1972, Gopnik, 1996, Saxe, 2005). These types of high level reasoning capacities are far more developed in humans than other animals. Correspondingly, there is good evidence that both the TPN and the mentalizing system are highly evolved (Semendeferi et al., 2001, Schoenemann, 2006, Van Essen and Dierker, 2007, Rakic, 2009). Although folk psychology and folk physics may be analogous in this regard, it is evident that the folk use very different neural systems to engage in these high level processes, depending on whether they are building a model of a mind or a mechanism. Hence it would appear that the critical question for the problem of consciousness is: how should we understand the relationship between folk and scientific understanding? For mathematics and the hard sciences, the evidence suggests that scientific understanding represents a refinement, or cultural recycling (Dehaene and Cohen, 2007), of our folk capacities. But how does it work for the sciences of the mind? Is scientific psychology a refinement of folk psychology? Some schools of psychology, such as the Introspectionists and some therapeutic approaches, certainly appear to have this flavor. Yet this hardly seems a plausible model for neuroscience, which shares with biology and the other hard sciences a paradigmatic focus on understanding mechanism (Craver, 2007). Neuroscience looks more like a refinement of folk physics, not folk psychology. Might this be the source of the philosophical problem? Perhaps emphasizing mechanistic explanation promotes a switch in the faculties we use to guide our understanding of the mind, and so generates the sense that more mechanistic accounts are 'leaving out experience'?

However, there is a problem with this account. Why, when we switch to thinking about the mechanical properties of a brain, don't we also continue think of it as an experiencing mind? Why should the use of one faculty preclude the use of the other? Indeed, the very fact these faculties are supported by different neural networks might suggest they are well suited to operating in concert, without interference. An analogy may help here. We know that color and motion processing are associated with distinct brain areas, yet these representational properties

are seamlessly integrated in normal perception. Shouldn't it be the same for the experiential and mechanical properties of minds?

A better perceptual analogy for the explanatory gap is the rivalry generated by ambiguous figures such as the duck-rabbit illusion. We know both things are there, but focusing on one causes us to lose sight of the other. Hence, to have a satisfactory account of the explanatory gap, we need more than mere division. Perceptual rivalry is the result of a reciprocal inhibitory relationship between the two competing representations. Similarly, to explain the explanatory gap, we want evidence for reciprocal inhibition between experiential and mechanical representations.

2.5 Opposing brain networks

The basic methodology for cognitive brain mapping is straightforward. It relies on a principle called cognitive subtraction. The typical approach is to give participants one more cognitively complex task and a second control task. The control task may simply be rest - looking at a fixation point. Alternatively, it might be a task which is nearly as complex as the experimental task. Many experiments seek to isolate a single cognitive component and hence employ a control task which involves many of same processes as the experimental task. Measurements of brain activity associated with the control task are then subtracted away from the experimental task, and the standard inference is that any brain areas which are significantly activated are engaged in the additional cognitive processes associated with the experimental task.

However, some time ago experimenters began to notice that some brain areas were breaking the rules of cognitive subtraction. In particular, the DMN (including the mentalizing system) tends to be deactivated during the same broad range of tasks that activate the TPN – i.e. these regions are more active at rest than in many experimental conditions (Shulman et al., 1997). Later, it was discovered that the TPN and DMN are in tension even when participants are at rest (i.e. when they are not given any task). In other words, there is a 'spontaneous' or 'natural' tendency for the DMN to be suppressed when the TPN is activated, and vice versa (Fox et al., 2005). The term 'anticorrelated networks' was coined to describe this phenomenon. Since then, the DMN and anticorrelations have become a major topic of investigation in systems neuroscience, with hundreds of publications describing various aspects of their anatomy, function, relationship to other networks, and how they are affected by neuropsychiatric

conditions (Buckner et al., 2008, Broyd et al., 2009b, Andrews-Hanna, 2011). This phenomenon represents what may be the hottest topic, in terms of citations and studies, in the field of brain imaging today. So, what does this work suggest is the cognitive significance of the reciprocal inhibitory relationship between these networks? Two major hypotheses have dominated the literature.

The first hypothesis gave the networks their names. The idea was that the TPN ("task positive") is activated and the DMN ("default mode") is deactivated whenever the participant is engaged in a goal-directed task. According to this hypothesis, the function of the DMN is spontaneous cognition, which is in tension with goal directed cognition. There is little reason to dwell on this hypothesis, first because this characterization of the function of the DMN arguably never made much sense, but more critically because we know the DMN can be activated above resting levels by goal directed tasks (Iacoboni, 2004).

The second major hypothesis is currently dominant in the literature. It is the idea that the TPN is involved in externally directed cognition whereas the DMN is involved in internally directed cognition, including introspection, mental time travel (including episodic memory recall), and conceptual association. This hypothesis views the tension between the networks as a type of attentional competition. On the face of it, this seems plausible. However, there is a problem with this account. It doesn't look like this processing divide would yield the problem of consciousness as a primary concern. A division between internally and externally directed thinking might fit well with the belief "I am not a physical mechanism." Yet, it is much harder to see how we would ever extrapolate this problem to others. On a simple reading of this view we would never attribute the kinds of internal mental states we are acquainted with through introspection to others, because other people are, obviously, external to us. So the most straightforward reading of this account suggests the most salient disjoin would not be between minds and machines, but rather between one's own mind and everything external. This would incline us not to dualism but to a type of solipsism where we are convinced that we are the only conscious being. But this is clearly not the view to which most healthy people are inclined[2]. In short, the problem with a straightforward reading of the internal vs. external attention account is

[2] Not that I think the view is impossible. My bet is that a tendency to this view is associated with narcissism.

that it fails to take account of our prolific tendency to attribute experiential mental states to others; a process which we know recruits the same brain regions as introspection.

In conclusion, neither of the two major hypotheses that have been put forward to provide a cognitive characterization of the tension between the TPN and DMN appears wholly satisfactory, either as accounts of the imaging data or the problem of consciousness. If we buy the idea that the tension between these two opposing brain networks accounts for the explanatory gap, then we need to show that there is a tension between thinking about physical mechanisms and thinking about internal mental states, using goal directed tasks that don't confound the issue of internal vs. external attention.

2.6 The opposing domains hypothesis

The opposing domains hypothesis holds that the neural antagonism between the TPN and DMN reflects a fundamental cognitive tension between thinking about internal mental states and physical mechanisms. We recently published evidence which provides strong support for this hypothesis (Jack et al., 2012). In order to distinguish this hypothesis from competing accounts, we compared tasks matched in terms of sensorimotor processing demands and external focus, but which differed in terms of whether participants are thinking about mental states or mechanical processes. Further, to establish generality, we used two different types of social and mechanical reasoning tasks – one presented as short textual narratives, the other as video with sound. We found that both the social tasks (soap opera-like stories, and videos of two people conversing and misunderstanding each other) activated DMN regions, and deactivated TPN regions. Conversely, both the mechanical reasoning tasks (physics problems taken from puzzle books, and clips from the video encyclopedia of physics) activated TPN regions and deactivated DMN regions. In other words, TPN and DMN regions were pushed up and down like a see-saw, depending on whether the task involved thinking about physical mechanisms or internal mental states.

Our study revealed that a very large proportion (54%) of the cortical surface is sensitive to the domain of reasoning (i.e. social vs. mechanical), regardless of the modality of presentation (i.e. texts vs. videos). This supports the view there is a major division in cortical organization between social and mechanical reasoning. However, only more circumscribed regions were pushed all the way above and below resting levels by the different tasks, providing clear evidence for mutual suppression. Why might this be? All tasks recruit a variety of cognitive

processes – for instance, our externally focused tasks all involved visual attention, and they all required an occasional manual response (stimuli lasted 20s, followed by a 7s response window to a yes/no textual question). The sum of activity which is observed will depend on all the processes engaged. So it is hardly surprising that some DMN and TPN regions never fell below resting levels, and that some never rose above resting levels. These regions were likely recruited by processes which were common to all the tasks (e.g. external attention), or processes which were not engaged by any of them (e.g. internal attention, strong emotions). The critical question is what is the functional role of the regions which were pushed up and down? First, the DMN regions which were pushed up and down matched very closely to the classic mentalizing regions (MPFC, MP/PC, rTPJ) which numerous studies have found to be associated with thinking about the internal mental states of ourselves and others. Second, the TPN regions which were pushed up and down matched regions involved in analytic reasoning and working memory, as well as the mirror neuron system. They did not match so well with areas specifically involved in external attention. A third important question is how these DMN and TPN regions line up with regions which are found to be anti-correlated at rest. This question speaks to a central and theoretically significant aspect of the account I offer, namely the idea that our neural structure constrains our cognition and gives rise to the perceived problem of consciousness.

Suppose participants were given two visual attention tasks in alternation: the first task requires them to attend to the right visual field, while ignoring stimuli in the left visual field; the second requires them to attend to the left while ignoring the right. It is plausible that, given the right conditions, this pair of tasks would generate a pattern of reciprocal inhibition between regions similar to that which we observed. However, in this case the positive and negative constraints on cognition are built into the tasks themselves, and do not correspond to cognitive processes which are fundamentally opposed; e.g. it is also possible to spread attention between both visual fields. While such built-in constraints were not evident in the tasks we used, it could be that our culture and education causes us to engage in social and mechanical reasoning in such a way that we inhibit the other type of reasoning. This would be consistent with the suggestion that the problem of consciousness represents a purely cultural construction (Wilkes, 1988). However, if the activations and deactivations match regions which are fundamentally opposed, i.e. which are anti-correlated during spontaneous cognition, then this would suggest that the tension between social and mechanical reasoning arises as a product of our neurobiology. To

help establish the cross-cultural validity of our findings, we tested this hypothesis using a large sample in which half the participants were from the USA, and the other half from China. As predicted, there was a very close match between the brain regions which see-sawed during our tasks, and areas of maximal anti-correlation.

These findings strongly support our hypothesis concerning the nature of the cognitive tension that is reflected by the neural antagonism between these networks. It is highly implausible that our findings can be better explained by the internal vs. external attention hypothesis. Our social tasks were no less demanding of external attention than our mechanical reasoning tasks, and all the tasks were much more demanding of external attention than the resting baseline condition to which they were compared. At rest, we may presume that the participant predominantly attends to internal information (e.g. memories, imaginings, thoughts, internal sensations, conceptual associations) because their external environment is barren and unchanging. At least, it is just this assumption which has fueled the internal vs. external hypothesis. Yet if that account were true, our externally oriented social tasks should not have activated maximally anti-correlated DMN regions any more than being at rest, and TPN regions should have been more active than they are at rest.

Because our tasks were externally focused, philosophers may be concerned that our findings do not speak directly to standard formulations of the problem of consciousness, where the primary focus is on our own experiential mental states. Fortunately, an earlier study by Goldberg et al (2006) speaks to this lacuna. They demonstrate deactivation during sensorimotor processing of a DMN region which was activated when participants introspected their perceptual experience while being presented the same stimuli. The region involved, in superior frontal gyrus, is anatomically very close to the maximally anti-correlated MPFC region in the DMN. For methodological reasons, this study fails to speak decisively about which broad hypothesis best accounts for the tension between the DMN and TPN. However, it does provide empirical support for the claim that the tension extends to the case where participants are focused on their own perceptual experiences.

These findings provide strong positive support for my hypothesis concerning the neural origins of the perceived explanatory gap. They do not provide evidence that the mind is non-physical. Thus, they do not challenge physicalism (Jack and Shallice, 2001). But they do suggest

that reductive physicalism, the explanatory strategy, is misconceived: our neural structure seems to present a barrier to understanding experience in physical terms. According to this view the explanatory gap is genuine, but it isn't a feature of the world, it lies in our heads.

However, an important question remains. Why does there seem to be a particularly potent philosophical concern about scientific accounts of experiential mental states as opposed, e.g., to accounts of intentional action? The mentalizing system is implicated in a broad array of social cognitive tasks, so shouldn't we see a similar tension for all varieties of mental state attribution? Although Descartes's formulation of the mind-body problem saw human reason as the central aspect that lay beyond mechanical explanation, the march of progress in both psychology and computer science appears to have softened that perception, yielding modern forms of the problem which follow Nagel's view that "consciousness is what makes the mind-body problem really intractable" (Nagel, 1974). To properly address this issue, we should take a look at the theory which guided this investigation.

3 The broader picture

3.1 The phenomenal stance

Robbins & Jack (2006) take the view that research in social cognition has been heavily influenced by a model which views social cognition as, first and foremost, a tool for predicting and manipulating others. In our view, this approach fails to take proper account of some very important functions of social cognition, namely its role in social bonding/affiliation, moral cognition, and pro-social behaviors such as child-rearing. To remedy this omission, we developed a model which built on Dennett's work. Using his language and framing, we introduced a new construct, the phenomenal stance (Figure 2). The phenomenal stance is the stance we adopt when we reflect upon our own experiences and feelings, as well as the experiences and feelings of others. It is a stance that we tend to avoid if we feel no compassion for the other – in which case the tendency is to deny or dismiss their feelings. Entering into this stance towards another deepens our social connection with them, and our sense of moral commitment to them. When this stance successfully guides another's interactions with us, it makes us feel *understood*. This notion of intimate interpersonal understanding is, we contend,

quite distinct from other ways of understanding the mind. It is not the same as "I understand what you think and I can guess what you are going to do" (the intentional stance), nor "I understand the neurobiological processes occurring in your brain" (the physical stance). The following quote gives an intuitive sense of it:

> *"[W]hen a person realizes he has been deeply heard, his eyes moisten. I think in some real sense he is weeping for joy. It is as though he were saying, "Thank God, somebody heard me. Someone knows what it's like to be me." In such moments I have had the fantasy of a prisoner in a dungeon, tapping out day after day a Morse code message, "Does anybody hear me? Is anybody there?" And finally one day he hears some faint tappings which spell out "Yes." By that one simple response he is released from his loneliness; he has become a human being again." (Rogers, 1980)*

Robbins & Jack (2006) put forward two key claims. The first was that thinking about experience (the phenomenal stance) is tightly linked to moral sentiments, in particular feelings of moral concern. We suggest this linkage isn't present for types of social cognition better characterized by Dennett's intentional stance. Second, our central hypothesis was that the problem of consciousness is generated by a tension between the phenomenal stance and the physical stance, a tension which either isn't present or is considerably less powerful between the physical and the intentional stances. Linking these claims together we predicted that psychopaths would not be able to perceive the problem of consciousness.

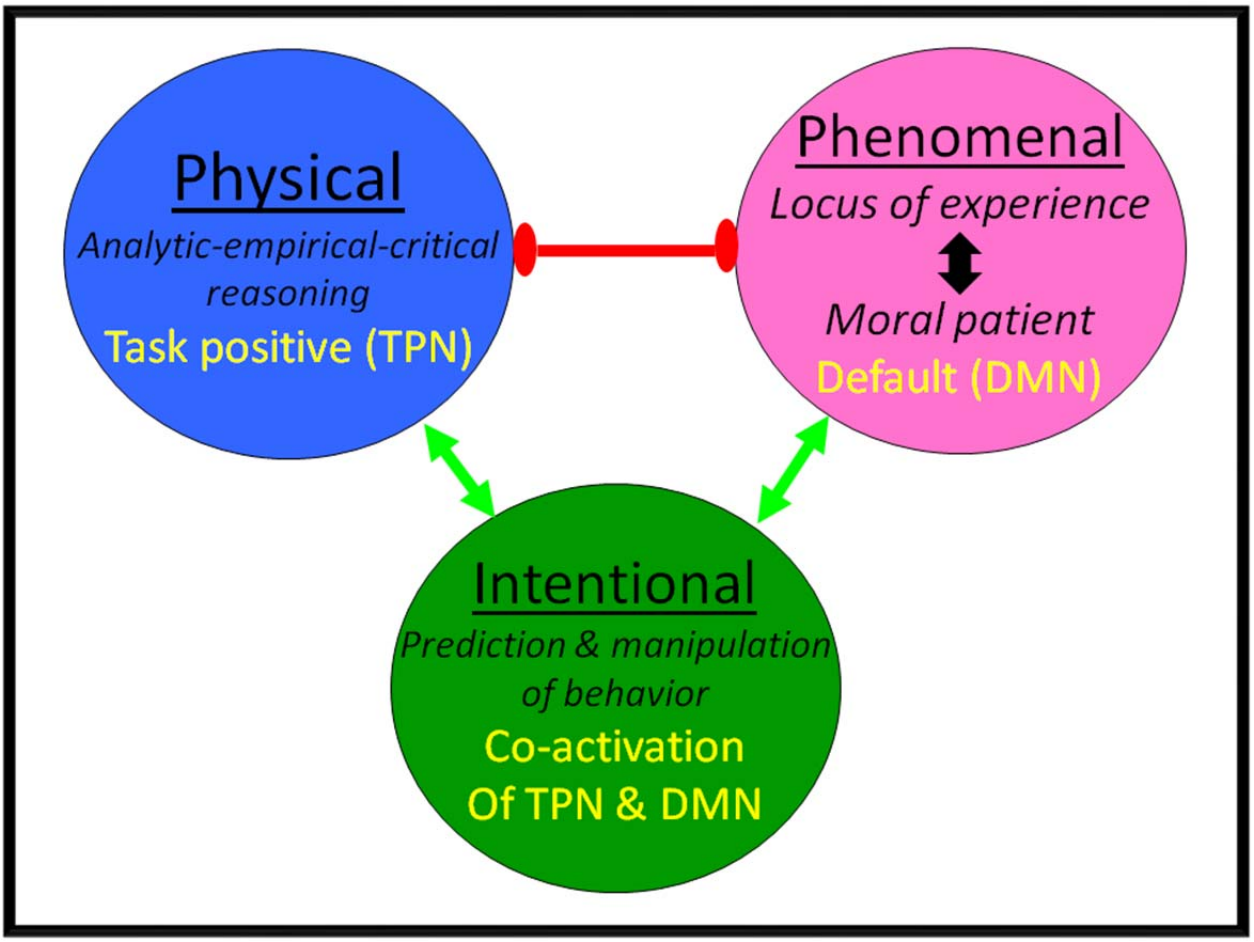

Figure 2: Three cognitive stances, their relationships to each other, and the brain networks involved. Bidirectional arrows indicate mutual compatibility; barbell indicates mutual antagonism. Adopting the Intentional stance corresponds to a focus on the behavior of an intentional agent. This stance bridges between the Phenomenal and Physical stances, and involves co-activation of default and task positive regions. Nonetheless, there remains a fundamental tension between representing experiential mental states and representing mechanical processes. The model depicted represents a synthesis of the Phenomenal Stance model (Robbins & Jack, 2006; Jack & Robbins, 2012) and the opposing domains hypothesis (Jack et al., 2012).

3.2 Opposing and blended cognitive modes

We know beyond any reasonable doubt that there is a major tension in the brain. Aside from the evidence already presented (reviewed at much greater length in the citations that follow), a large number of studies also show that this tension is strongly associated with healthy mental function: the tendency for these networks to suppress each other is diminished in virtually

every major mental disorder, including disorders which impact social function such as autism and schizophrenia (Buckner et al., 2008, Broyd et al., 2009a, Andrews-Hanna, 2011).

Nonetheless, even in the healthy brain this tension is far from absolute. Parts of these networks do co-activate both during spontaneous cognition and during certain tasks. How should we understand this phenomenon? I believe this occurs when the brain is supporting blended cognitive modes. The notion is that these modes aren't identical with either the full blown phenomenal stance or with full blown varieties of analytical-empirical-critical reasoning, such as mechanistic reasoning. But these blended cognitive modes do borrow aspects from each of the two pure modes. One such mode is creative thinking, or insight problem solving, which involves a combination of logical thought and a more intuitive mode of thinking. We see co-activation of parts of both networks at the moment of insight (Subramaniam et al., 2009). More creative individuals also demonstrate less tension between the networks (Takeuchi et al., 2011) – the only clearly desirable individual difference which appears to be associated with a lessening of this tension. The problem with thinking creatively is that we sometimes come up with flatly foolish ideas. In other words, our critical faculties are partially suspended during creative thinking. I hypothesize that this is an inevitable trade-off of breaching the neural divide to recruit the DMN alongside the TPN.

According to my model, the intentional stance also represents such a blended cognitive mode. Both behavioral and neuroimaging evidence is consistent with this view (I won't re-review the evidence here for reasons of space, but see discussion near the end of Jack et al., 2012). According to this view, the intentional stance involves a (limited) appreciation of the internal mental states of others blended with analytic-empirical-critical thinking. This blend both enables us to make predictions about others' actions and causes us to adopt a more emotionally detached view of them. This blending of the two cognitive modes is reflected in ordinary language: when we refer to someone as "calculating" or "manipulative," we do not literally mean that they are doing sums or using their fine motor skills. We are, of course, referring to an emotionally distanced and somewhat anti-social mode of social cognition. We likely use these terms because this mode of social cognition involves brain areas associated with mathematical calculation and transitive action. Hence, for instance, when conditions involving deception are compared with non-deceptive conditions, the most consistent differences in brain activity are seen in the TPN (Christ, Van Essen, Watson, Brubaker, and McDermott, 2009). Individuals with

a tendency to Machiavellian thinking also show greater activation of TPN regions during social cognition (Bagozzi et al., 2013).

3.3 Incommensurable understanding

My laboratory is engaged in testing and extending this model using both neuroimaging and behavioral measures. A central theme that runs through this work is that our model specifically links the phenomenal stance, and the tension with the physical stance, to moral sentiments[3]. Hence, for us: (1) Moral sentiments are specifically tied to the attribution of states and properties which escape mechanical explanation, in particular experiential states. (2) Both moral sentiments and the attribution of these seemingly non-physical states and properties are associated with activation of the DMN accompanied by deactivation of the TPN.

To behaviorally assess the tendency for individuals to adopt the pure phenomenal stance, over the blended intentional stance, we use standard well-validated measures of moral concern, the empathetic concern subscale of the Interpersonal Reactivity Index (IRI-EC, Davis, 1980), and the callous affect subscale of the Self-Report Psychopathy scale (SRP-CA, Paulhus et al., in press). To assess the tendency for individuals to adopt pure analytic-empirical-critical reasoning we use two measures: the cognitive reflections task (CRT, Frederick, 2005) primarily measures critical reasoning; and the Intuitive Physics Test (IPT, Baron-Cohen, 2001) measures basic mechanical reasoning. We have found a small (r~-0.2) but highly robust and consistent negative correlation between the IRI-EC and the IPT (Jack and Gabriel, in preparation)[4]. In contrast, negative correlations were not observed with measures of the intentional stance (e.g. theory of mind accuracy). We do not know of any other theory which would predict this specific negative relationship[5].

---

[3] Note the claim is specific to moral sentiments, such as compassion, altruism, moral approbation and outrage, and thoughts that direct relate to these sentiments. The TPN also clearly plays a role in moral deliberation. I suspect the TPN is key to moralizing, hence the dissonance this provokes in those who possess genuine moral sensibility.

[4] There are also robust gender differences associated with each measure (females higher on IRI-EC, lower on IPT), however it is notable that the negative correlation between these measures is still present even when gender and other demographic variables including education are partialled out. Consistent negative correlations are also observed between IRI-EC and CRT, but are smaller and not significant all the individual samples.

[5] We focus on psychopathy because it represents a different primary deficit (the phenomenal stance) from autism (the intentional stance). The theories of Simon Baron-Cohen and Francesca Happe, both of which are focused on autism, might be translated as hypothesizing a tension between the intentional and the physical stances. James Blair's work helped to characterize the differences between psychopathy and autism. All of their work was

There is also clear evidence that supports our view that moral concern is more strongly linked to attributions of experience than attributions of intentional states (Gray et al., 2007, Knobe and Prinz, 2008, Gray et al., 2011, Jack and Robbins, 2012). Work on the phenomenon of dehumanization, in which individuals are viewed as less than human and anti-social behavior towards them is sanctioned, similarly indicates a link between moral concern and the attribution of experiential states. Dehumanization has been shown to be linked to a belief that target individuals are less capable of experiencing certain kinds of more sophisticated emotional states (Leyens et al., 2001, Castano and Giner-Sorolla, 2006, Čehajić et al., 2009). A number of accounts also highlight the important role that essentialist thinking plays in dehumanization. A human essence, or soul, tends to be attributed to in-group members, but denied to dehumanized out-groups (Smith, 2011). This work has overlap with the observation that undergraduates presented with authoritative scientific denials of the existence of the human soul and free will demonstrate more anti-social behavior (Vohs and Schooler, 2008, Baumeister et al., 2009).

We (Jack et al., under review-a) have conducted a neuroimaging investigation which distinguishes two forms of dehumanization: animal and machine (Haslam, 2006). This work supports our view that the tension between the DMN and TPN reflects a distinction between thinking about conscious agents for whom we feel moral concern and inanimate objects for which we do not. We found that machine dehumanizing (aka 'objectifying') involves a lessening of activity in the DMN (corresponding to social indifference), whereas animal dehumanizing involves co-activation of the DMN and the TPN (the signature of the blended cognitive mode of the intentional stance). However, we did find that one region was consistently associated with seeing human: the MP/PC region which can be seen to the left of Figure 1. The MP/PC region is a central node for the DMN, and demonstrates the strongest anti-correlations with TPN regions. The same region activates when we look at pictures of the faces of people we personally know and are close to (i.e. in-group members), but deactivates when we view pictures of famous or unfamiliar faces. I hypothesize that this region plays a key role in generating the belief that the perceived individual possesses a human essence or 'soul'.

---

influenced by the work of their shared mentor/collaborator, Uta Frith. The work of all these individuals was influential in the development of our theory.

Our tendency to essentialism raises few philosophical qualms when it is applied to the physical stance. Endorsement of essentialism about physical properties motivates physicalism. However, the philosophical problems multiply when essentialist thinking is applied to other modes, such as the intentional stance (Brentano's problem), or the phenomenal stance (metaphysical dualism). Philosophers continue to debate whether the explanatory gap poses a real challenge to physicalism. I don't believe that it does, provided we suspend our tendency to essentialism for this stance. I endorse this approach to metaphysics[6] and certain aspects of scientific understanding (more on this shortly). However, I have serious concerns about applying it as a general prescription. In everyday life, I suspect that suspending essentialist thinking represents an unnatural, ineffective, and undesirable way of thinking.

We have developed a measure of belief in metaphysical dualism which comprises five items, including "Humans have a soul", "The mind can be understood completely by thinking of it as like a very complicated computer", and "Thoughts and feelings are nothing more than the activity of neurons" (last two reverse coded). In a series of five experiments (Jack, in preparation), we found a highly replicable and robust negative correlation (r~-0.34) between belief in dualism and the primary psychopathic trait of callous affect[7]. This negative correlation survived after partialling out demographic variables, cognitive measures (e.g. the IPT & CRT) and measures of religious belief. In contrast, correlations between dualism and measures of physical (e.g. IPT) and intentional (e.g. ToM) reasoning were weak or absent.

Clearly these findings fit well with the hypothesis (Robbins and Jack, 2006) that psychopaths can't see the problem of consciousness[8]. Taking these finding together with other work on dehumanization and the anti-social effects of denying the soul and free will, they present a powerful picture. When we see persons, that is, when we see others as fellow humans,

---

[6] I view metaphysics as making claims about the structure of the physical world. Hence, I regard metaphysical dualism as a category error. It is driven by a mode of understanding which is not suited to understanding the physical world. I suggest we ought to recognize the need for distinct ontologies associated with the different stances, and that only one such ontology should be seen as relevant to metaphysics. Beyond these casual remarks, I leave the details to philosophers.

[7] Similar but positive correlations exist between dualism and IRI empathetic concern

[8] The effect is more specific, to this aspect of the problem of consciousness, than we originally anticipated. Psychopaths don't seem to have any problem perceiving an explanatory, epistemological or conceivability gap. In one experiment we presented scenarios based on Nagel's bat (1974), Jackson's Mary (1986) and Chalmers's philosophical zombie (1996). Virtually everyone perceived these gaps, and there was no clear correlation with psychopathy.

then our percept is of something essentially non-physical nature. This feature of our psychology appears to be relevant to a number of other philosophical issues, including the tension between utilitarian principles and deontological concerns about harming persons (Jack et al., accepted), the question of whether God exists (Jack et al., under review-b), and the problem of free will[9]. So, should we regard our tendency to intuitive dualism as a problem, as Greene (2011) suggests? I believe it does create a problem, because it requires our moral, scientific and philosophical thinking to be more nuanced and complex. However, I believe we will face larger problems than a mere increase in intellectual demands if we ignore or dismiss this fundamental feature of our psychology. It is tempting to suppose that our different ways of understanding the world can be rationalized into a single coherent world view; hence it is a reasonable hypothesis that a more considered approach might allow for greater reconciliation. However, this hypothesis isn't supported by the data. A striking finding that emerges from our work on dualism is that individuals who score higher on reflective reasoning (CRT) don't escape the effect that empathetic concern has on creating a divided metaphysical world view (figure 3). If anything the effect of empathy on the tendency to believe in dualism is stronger for individuals high in critical thinking than for those who tend to go with intuition.

[9] An ongoing project in collaboration with Joshua Knobe and others suggests that anti-physicalist metaphysical beliefs about free will are associated with having stronger moral sentiments about the need for just punishment.

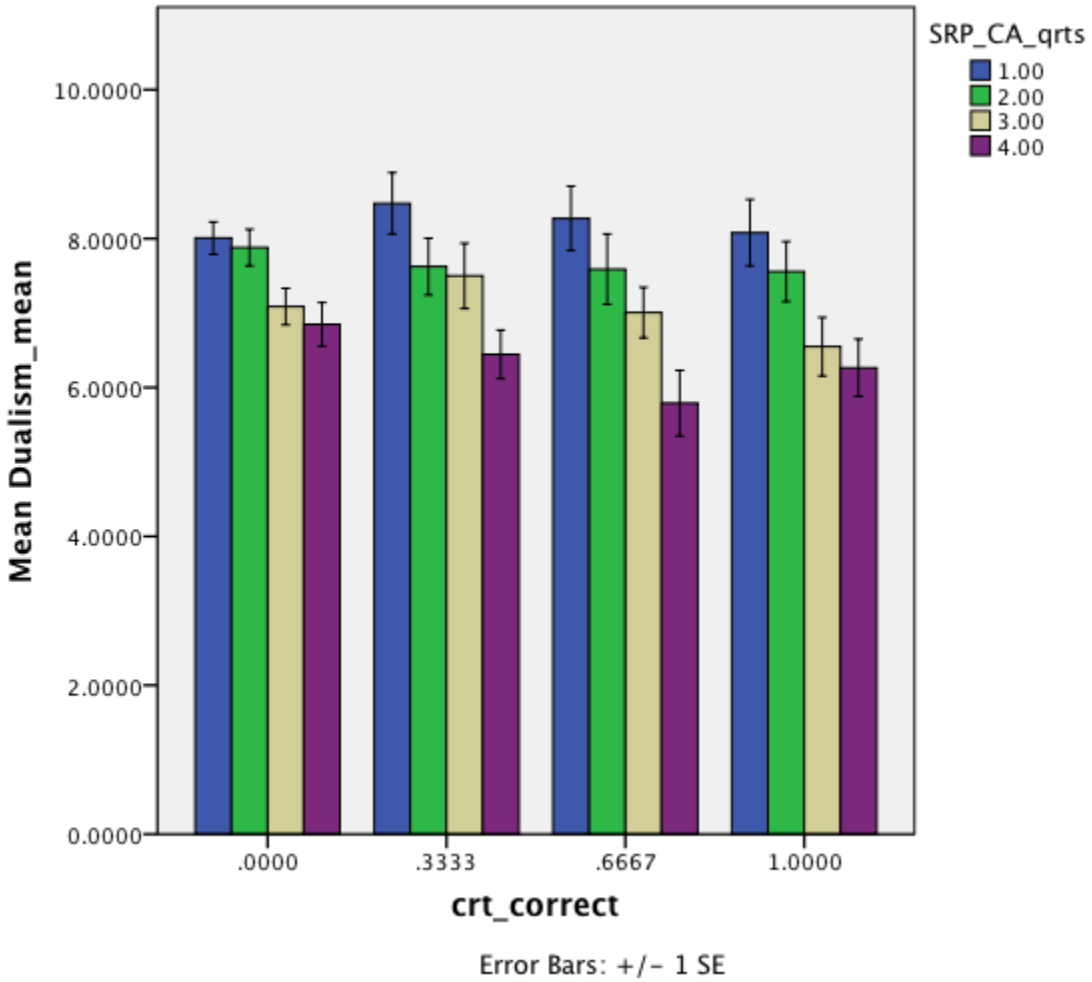

Figure 3: Graph depicting tendency to believe in dualism as a function of psychopathic callous affect (SRP-CA, split into quartiles) and raw score on the cognitive reflections task (CRT). Data is pooled from mainly online studies, totaling 874 participants. The CRT measures critical thinking by requiring participants to avoid giving an intuitively appealing but incorrect answer. Note that the tendency for less psychopathic individuals to believe in dualism holds regardless of score on the CRT. Hence, the general tendency to inhibit intuitive but incorrect responses does not prevent empathetic concern from influencing one's metaphysical world view.

3.4 Concluding remarks

The model I present here is a type of dual-process theory. However, it characterizes a very different cognitive divide from classic dual-process theory, which is best known from the work of the Nobel prize-winning psychologist Daniel Kahneman. According to classic dual-process theory (Kahneman, 2003), numerous decision-making phenomena may be seen as reflecting a fight between evolutionarily primitive, unconscious and automatic processes on the

one hand, and on the other a conscious deliberative general-purpose type of reasoning which is more amenable to education. Unlike classic dual-process theory, my model is informed primarily by what cognitive neuroscience tells us about the types of reasoning supported by the networks we see to be in tension in the brain, as well as by some philosophical considerations. According to my view, we possess two mutually exclusive faculties, both of which are conscious, deliberative and highly evolved, and each of which may be cultivated through distinct cultural learning traditions (Snow, 1959). Yet, each is substantially incomplete: one is incapable of comprehending human experience and essential aspects of morality, while the other is incapable of comprehending the mechanical and mathematical structure of the physical world. While we can blend these cognitive modes, our neural structure creates interference between them. As a result, blended cognitive modes fail to capture insights that emerge only when each of the pure opposing cognitive modes operates in isolation. According to this view, there is no faculty which can be properly called 'general reasoning', because we lack a single integrated capacity capable of generating the full range of human insight.

It follows from this view that progress in psychology will not be best achieved by adopting a blended cognitive mode, i.e. the intentional stance, to the exclusion of other perspectives[10]. Instead, it appears that a complete understanding requires something more like juggling: we must fully immerse ourselves in distinct perspectives and only then seek to build bridges between the incommensurable conceptual frameworks that emerge. To briefly flesh this out, my view is that our social cognitive machinery creates a virtual world of experiences and persons which we imbue with meaning (Jack, 2011). From the impersonal point of view of the physical stance, this virtual world only exists as a figment of our imagination, and its coherence between individuals is only partial. To us, on the other hand, this virtual world is often more 'real', in the sense of being genuinely important and meaningful, than anything else. The only way to explore this virtual world is to adopt the phenomenal stance, for example using the first and second person approaches initially developed by the Introspectionists (Jack and Roepstorff, 2002). We can't properly explore this world using objective measures, because when we come to interpret the data using either the physical or the intentional stance the very act of scientific interpretation moves us away from providing a description of experience, and blinds us to its

[10] For a directly parallel argument about moral thought, concerning the inadequacy of a utilitarian perspective (corresponding to the intentional stance) by itself, see Jack et al (in press).

nuances[11]. Hence, contra Dennett's heterophenomological approach, objective measures cannot directly speak to the world of experience, and gainsay our introspective enquiries. Similarly, introspective enquires fail to provide direct evidence concerning the mechanical structure of the mind[12]. Nonetheless, these two incommensurable perspectives can be partially reconciled by adopting a more removed perspective that bridges between the personal world of experience and the impersonal world of mechanism. No account of our neural workings will ever tell us what it is like to be in love or see red. But they can tell us what is happening in someone that makes them have the experience of being in love or seeing red. If we want to offer such enlightening scientific accounts, we must not only develop our accounts of mechanism but also more nuanced accounts of experience. For instance, "being in love" is surely too broad and crude a description of experience to find traction with a sophisticated scientific account of the processes involved in human romantic attachment. There is simply no substitute for introspective methods for improving our accounts of experiential phenomena, yet systematic explorations (e.g. Hurlburt et al., 1994) are extremely rare. Without making use of these methods, our understanding of experience will be too crude to be worthy of our scientific efforts, and the scientific accounts we generate won't be able to indirectly inform the alternate non-physical reality that we can only directly perceive through the phenomenal stance: a world constituted by irreducible persons, experiences and moral truths.

Some psychologists have vehemently resisted the rise of cognitive neuroscience, arguing the workings of the brain are irrelevant to understanding the mind. These authors have sought to privilege a conceptual framework which emerges from the intentional stance over one that derives from the physical stance. I have often felt embarrassed on behalf of these authors. They fail to realize that their writings are little more than manifestos which promulgate their prejudiced belief in a highly limited conception of the mind (Jack et al., 2006). In this case,

[11] To illustrate this, the history of failed equations between objective measures and subjective states is long and has generated much controversy. Examples can be found in my cited papers. I suggest the solution is to concede that the notion of an 'objective measure of awareness' is an oxymoron. We can make actual progress by carefully establishing convergence between subjective measures which speak to experience and objective measures which speak to mechanism (Jack & Shallice, 2001).

[12] The error here is analogous to imagining that the structure of the filing system apparent from your computer's user interface is informative about how your files are physically encoded on the hard drive, and vice versa. Just as with experiential and physical perspectives on the mind, the two can be related, but it takes a lot of work to make the link. Is the filing structure you know from everyday experience 'real'? In one sense no, in another it is much more real and certainly more useful than knowledge of the physical encoding.

embarrassment seems an appropriate emotion, since this prejudice has been in the minority and continues to erode. However, there is another prejudice which still remains in the majority. In scientific psychology, there remains a strong inclination to dismiss an experiential perspective on the mind, acquired through introspective methods. This dismissal of the insights that psychology can glean from the phenomenal stance is a more troubling concern. Like other prejudices whose effect is to demean or ignore the humanity of others, the emotion it triggers in those who recognize it is not embarrassment but outrage (Jack, in press). I have written about this issue, and how we might resolve it, in the past (Jack and Shallice, 2001, Jack and Roepstorff, 2002, 2003, Jack, 2011). However, I have come to understand that the main barrier to progress has been that most psychologists, and some philosophers, simply have not seen the problem (Greene, 2011). I hope this essay helps to make it more apparent[13].

Acknowledgments

This work was supported in part by NSF grant 0841313 and by a grant from the University Hospitals Case Medical Center Spitz Brain Health fund, both awarded to the author. The author would like to thank Jared Friedman for help with analysis, and Philip Robbins, Shaun Nichols, Stuart Youngner and two anonymous reviewers for comments on an earlier draft.

References

Amodio DM, Frith CD (2006) Meeting of minds: the medial frontal cortex and social cognition. Nat Rev Neurosci 7:268-277.

Andrews-Hanna JR (2011) The Brain's Default Network and Its Adaptive Role in Internal Mentation. Neuroscientist.

Arico A, Fiala B, Goldberg RF, Nichols S (2011) The Folk Psychology of Consciousness. Mind Lang 26:327-352.

Bagozzi RP, Verbeke WJMI, Dietvorst RC, Belschak FD, van den Berg WE, Rietdijk WJR (2013) Theory of Mind and Empathic Explanations of Machiavellianism: A Neuroscience Perspective. Journal of Management.

[13] I assume the conflicts that arise between our incommensurable perspectives on the world will always tend to create confusion. Another way to put my take-home message is that neither physicalism nor metaphysical dualism do anything to resolve this. I endorse physicalism, but note that it ignores the problem. Metaphysical dualism acknowledges the problem but responds by setting our confusion in stone. I suggest we accept we have these incommensurable perspectives, that they are compatible with physicalism, and adopt a methodological dualism in order to chip away at the links between them.

Baron-Cohen S, Sally Wheelwegith, Amanda Spong, victoria Scahill, John Lawson (2001) Are intuitive physics and intuitive psychology independent? A test with childeren with Asperger Syndrome. Journal of Developmental and Learning Disorders 5:47 - 78.

Baumeister RF, Masicampo EJ, DeWall CN (2009) Prosocial Benefits of Feeling Free: Disbelief in Free Will Increases Aggression and Reduces Helpfulness. Personality and Social Psychology Bulletin 35:260-268.

Bloom P (2004) Descartes' baby : how the science of child development explains what makes us human. New York: Basic Books.

Blos J, Chatterjee A, Kircher T, Straube B (2012) Neural correlates of causality judgment in physical and social context-The reversed effects of space and time. Neuroimage 63:882-893.

Borsci G, Boccardi M, Rossi R, Rossi G, Perez J, Bonetti M, Frisoni GB (2009) Alexithymia in healthy women: A brain morphology study. Journal of Affective Disorders 114:208-215.

Broyd SJ, Demanuele C, Debener S, Helps SK, James CJ, Sonuga-Barke EJ (2009a) Default-mode brain dysfunction in mental disorders: a systematic review. Neurosci Biobehav Rev 33:279-296.

Broyd SJ, Demanuele C, Debener S, Helps SK, James CJ, Sonuga-Barke EJS (2009b) Default-mode brain dysfunction in mental disorders: A systematic review. Neuroscience & Biobehavioral Reviews 33:279-296.

Buckner RL, Andrews-Hanna JR, Schacter DL (2008) The Brain's Default Network: Anatomy, Function, and Relevance to Disease. Annals of the New York Academy of Sciences 1124:1-38.

Castano E, Giner-Sorolla R (2006) Not quite human: infrahumanization in response to collective responsibility for intergroup killing. J Pers Soc Psychol 90:804-818.

Čehajić S, Brown R, González R (2009) What do I Care? Perceived Ingroup Responsibility and Dehumanization as Predictors of Empathy Felt for the Victim Group. Group Processes & Intergroup Relations 12:715-729.

Chalmers DJ (1996) Facing up to the problem of consciousness. Com Adap Sy 5-28.

Corbetta M, Akbudak E, Conturo TE, Snyder AZ, Ollinger JM, Drury HA, Linenweber MR, Petersen SE, Raichle ME, Van Essen DC, Shulman GL (1998) A common network of functional areas for attention and eye movements. Neuron 21:761-773.

Craver CF (2007) Explaining the brain : mechanisms and the mosaic unity of neuroscience. Oxford

New York Oxford University Press,: Clarendon Press ;.

Davis MH (1980) A multidimensional approach to individual differences in empathy. Psychology 10:85 - 114.

Dehaene S, Cohen L (2007) Cultural recycling of cortical maps. Neuron 56:384-398.

Denny BT, Kober H, Wager TD, Ochsner KN (2012) A Meta-analysis of Functional Neuroimaging Studies of Self- and Other Judgments Reveals a Spatial Gradient for Mentalizing in Medial Prefrontal Cortex. J Cogn Neurosci 24:1742-1752.

Downing PE (2005) Domain Specificity in Visual Cortex. Cerebral Cortex 16:1453-1461.

Duncan J, Owen AM (2000) Common regions of the human frontal lobe recruited by diverse cognitive demands. Trends in Neuroscience 23:475-483.

Fiala B, Arico A, Nichols S (2012) On the psychological origins of dualism: Dual-process cognition and the explanatory gap. In: Creating Consilience: Integrating the Sciences and the Humanities(Slingerland, E. and Collard, M., eds) New York: Oxford University Press.

Fox MD, Corbetta M, Snyder AZ, Vincent JL, Raichle ME (2006) Spontaneous neuronal activity distinguishes human dorsal and ventral attention systems. Proc Natl Acad Sci U S A 103:10046-10051.

Fox MD, Snyder AZ, Vincent JL, Corbetta M, Van Essen DC, Raichle ME (2005) The human brain is intrinsically organized into dynamic, anticorrelated functional networks. Proc Natl Acad Sci U S A 102:9673-9678.

Frederick S (2005) Cognitive reflection and decision making. Journal of Economic Perspectives 19:25-42.

Goldberg I, Harel M, Malach R (2006) When the Brain Loses Its Self: Prefrontal Inactivation during Sensorimotor Processing. Neuron 50:329-339.

Goodale MA, Milner AD (1992) Separate visual pathways for perception and action. Trends Neurosci 15:20-25.

Gopnik A (1996) The scientist as child. Philos Sci 63:485-514.

Gray HM, Gray K, Wegner DM (2007) Dimensions of Mind Perception. Science 315:619-619.

Gray K, Jenkins AC, Heberlein AS, Wegner DM (2011) Distortions of mind perception in psychopathology. Proc Natl Acad Sci U S A 108:477-479.

Greene JD (2011) Social Neuroscience and the Soul's Last Stand. In: Social Neuroscience: Toward Understanding the Underpinnings of the Social Mind(Todorov, A. et al., eds) New York: Oxford University Press.

Grill-Spector K, Knouf N, Kanwisher N (2004) The fusiform face area subserves face perception, not generic within-category identification. Nat Neurosci 7:555-562.

Gundel H (2004) Alexithymia Correlates With the Size of the Right Anterior Cingulate. Psychosomatic Medicine 66:132-140.

Haslam N (2006) Dehumanization: an integrative review. Pers Soc Psychol Rev 10:252-264.

Honey CJ, Sporns O, Cammoun L, Gigandet X, Thiran JP, Meuli R, Hagmann P (2009) Predicting human resting-state functional connectivity from structural connectivity. Proc Natl Acad Sci U S A 106:2035-2040.

Hurlburt RT, Happe F, Frith U (1994) Sampling the form of inner experience in three adults with Asperger syndrome. Psychol Med 24:385-395.

Iacoboni M (2004) Watching social interactions produces dorsomedial prefrontal and medial parietal BOLD fMRI signal increases compared to a resting baseline. NeuroImage 21:1167-1173.

Iacoboni M, Molnar-Szakacs I, Gallese V, Buccino G, Mazziotta JC, Rizzolatti G (2005) Grasping the Intentions of Others with One's Own Mirror Neuron System. PLoS Biol 3:e79.

Jack AI (2011) Describing Inner Experience? Proponent Meets Skeptic. Philos Psychol 24:283-287.

Jack AI (in preparation) Consciousness and Callousness: Empathetic concern distinguishes folk from scientific conceptions of the mind.

Jack AI (in press) Introspection: the tipping point. Consciousness & Cognition.

Jack AI, Dawson AJ, Begany KL, Leckie RL, Barry KP, Ciccia AH, Snyder AZ (2012) fMRI reveals reciprocal inhibition between social and physical cognitive domains. Neuroimage 66C:385-401.

Jack AI, Dawson AJ, Norr M (under review-a) Seeing Human: distinct and overlapping neural signatures associated with two forms of dehumanization.

Jack AI, Gabriel JJ (in preparation) Individual and gender differences demonstrate a trade-off between empathetic concern and physical reasoning ability.

Jack AI, Nolan SN, Boyatzis RE, Friedman JP (under review-b) A novel dual-process account of religious belief.

Jack AI, Robbins P (2004) The illusory triumph of machine over mind: Wegner's eliminativism and the real promise of psychology. Behavioral and Brain Sciences 27:665-+.

Jack AI, Robbins P (2012) The Phenomenal Stance Revisited. Review of Philosophy and Psychology 3:383-403.

Jack AI, Robbins PA, Friedman JP, Meyers CD (accepted) More than a feeling: Counterintuitive effects of compassion on moral judgment.

Jack AI, Roepstorff A (2002) Introspection and cognitive brain mapping: from stimulus-response to script-report. Trends Cogn Sci 6:333-339.
Jack AI, Roepstorff A (2003) Why trust the subject? Journal of Consciousness Studies 10:v-xx.
Jack AI, Shallice T (2001) Introspective physicalism as an approach to the science of consciousness. Cognition 79:161-196.
Jack AI, Sylvester CM, Corbetta M (2006) Losing our brainless minds: how neuroimaging informs cognition. Cortex 42:418-421; discussion 422-417.
Jackson F (1986) What Mary Didnt Know + Knowledge Argument against Physicalism. J Philos 83:291-295.
Johnson S, Slaughter V, Carey S (1998) Whose gaze will infants follow? The elicitation of gaze-following in 12-month-olds. Developmental Science 1:233-238.
Kanwisher N, McDermott J, Chun MM (1997) The fusiform face area: a module in human extrastriate cortex specialized for face perception. J Neurosci 17:4302-4311.
Knobe J, Prinz JJ (2008) Intuitions About Consciousness: Experimental Studies. Phenomenology and the Cognitive Sciences 7:67-83.
Kroger JK, Sabb FW, Fales CL, Bookheimer SY, Cohen MS, Holyoak KJ (2002) Recruitment of Anterior Dorsolateral Prefrontal Cortex in Human Reasoning: a Parametric Study of Relational Complexity. Cerebral Cortex 12:477-485.
Leibniz G (1714) The Mondadology.
Levine J (2000) Conceivability, identity, and the explanatory gap. From Anim Animat 3-12.
Lewis D (1972) Psychophysical and theoretical identifications. Australasian Journal of Philosophy 50:249 - 258.
Leyens J-P, Rodriguez-Perez A, Rodriguez-Torres R, Gaunt R, Paladino M-P, Vaes J, Demoulin S (2001) Psychological essentialism and the differential attribution of uniquely human emotions to ingroups and outgroups. European Journal of Social Psychology 31:395-411.
Mahon BZ, Anzellotti S, Schwarzbach J, Zampini M, Caramazza A (2009) Category-Specific Organization in the Human Brain Does Not Require Visual Experience. Neuron 63:397-405.
Mars RB, Neubert FX, Noonan MP, Sallet J, Toni I, Rushworth MF (2012) On the relationship between the "default mode network" and the "social brain". Front Hum Neurosci 6:189.
Martin A, Weisberg J (2003) Neural Foundations for Understanding Social and Mechanical Concepts. Cognitive Neuropsychology 20:575-587.
Miller G (2005) What is the biological basis of consciousness? Science 309:79.
Mitchell JP (2002) Distinct neural systems subserve person and object knowledge. Proceedings of the National Academy of Sciences 99:15238-15243.
Mitchell JP (2009) Social psychology as a natural kind. Trends Cogn Sci 13:246-251.
Nagel T (1974) What Is It Like to Be a Bat. Philos Rev 83:435-450.
Ochsner KN, Knierim K, Ludlow DH, Hanelin J, Ramachandran T, Glover G, Mackey SC (2004) Reflecting upon feelings: an fMRI study of neural systems supporting the attribution of emotion to self and other. Journal of Cognitive Neuroscience 16:1746-1772.
Ongur D, Price JL (2000) The organization of networks within the orbital and medial prefrontal cortex of rats, monkeys and humans. Cereb Cortex 10:206-219.
Paulhus DL, Neumann CS, Hare RD (in press) Manual for the Self-Report Psychopathy scale. Toronto: Multi-Health Systems.
Pelphrey KA (2005) Neural basis of eye gaze processing deficits in autism. Brain 128:1038-1048.
Rakic P (2009) Evolution of the neocortex: a perspective from developmental biology. Nat Rev Neurosci 10:724-735.
Robbins P, Jack AI (2006) The phenomenal stance. Philosophical Studies 127:59-85.
Rogers CR (1980) A way of being. Boston: Houghton Mifflin.

Saxe R (2005) Against simulation: the argument from error. Trends Cogn Sci 9:174-179.
Saxe R, Moran JM, Scholz J, Gabrieli J (2006) Overlapping and non-overlapping brain regions for theory of mind and self reflection in individual subjects. Social Cognitive and Affective Neuroscience 1:229-234.
Schilbach L, Bzdok D, Timmermans B, Fox PT, Laird AR, Vogeley K, Eickhoff SB (2012) Introspective minds: using ALE meta-analyses to study commonalities in the neural correlates of emotional processing, social & unconstrained cognition. PLoS One 7:e30920.
Schilbach L, Eickhoff S, Rotarskajagiela A, Fink G, Vogeley K (2008) Minds at rest? Social cognition as the default mode of cognizing and its putative relationship to the "default system" of the brain. Consciousness and Cognition 17:457-467.
Schoenemann PT (2006) Evolution of the Size and Functional Areas of the Human Brain. Annual Review of Anthropology 35:379-406.
Semendeferi K, Armstrong E, Schleicher A, Zilles K, Van Hoesen GW (2001) Prefrontal cortex in humans and apes: a comparative study of area 10. Am J Phys Anthropol 114:224-241.
Shulman GL, Fiez JA, Corbetta M, Buckner RL, Miezin FM, Raichle ME, Petersen SE (1997) Common blood flow changes across visual tasks: II. Decreases in cerebral cortex. Journal of Cognitive Neuroscience 9:648-663.
Smith DL (2011) Less than human : why we demean, enslave, and exterminate others. New York: St. Martin's Press.
Snow CP (1959) The two cultures and the scientific revolution. New York,: Cambridge University Press.
Subramaniam K, Kounios J, Parrish TB, Jung-Beeman M (2009) A brain mechanism for facilitation of insight by positive affect. J Cogn Neurosci 21:415-432.
Takeuchi H, Taki Y, Hashizume H, Sassa Y, Nagase T, Nouchi R, Kawashima R (2011) Failing to deactivate: the association between brain activity during a working memory task and creativity. Neuroimage 55:681-687.
Tarr MJ, Gauthier I (2000) FFA: a flexible fusiform area for subordinate-level visual processing automatized by expertise. Nat Neurosci 3:764-769.
Van Essen DC, Dierker DL (2007) Surface-Based and Probabilistic Atlases of Primate Cerebral Cortex. Neuron 56:209-225.
Van Overwalle F (2009) Social cognition and the brain: a meta-analysis. Hum Brain Mapp 30:829-858.
Van Overwalle F (2010) A dissociation between social mentalizing and general reasoning. Neuroimage.
Van Overwalle F, Baetens K (2009) Understanding others' actions and goals by mirror and mentalizing systems: A meta-analysis. NeuroImage 48:564-584.
Vincent JL, Kahn I, Snyder AZ, Raichle ME, Buckner RL (2008) Evidence for a frontoparietal control system revealed by intrinsic functional connectivity. J Neurophysiol 100:3328-3342.
Vohs KD, Schooler JW (2008) The value of believing in free will: encouraging a belief in determinism increases cheating. Psychol Sci 19:49-54.
Waytz A, Morewedge CK, Epley N, Monteleone G, Gao JH, Cacioppo JT (2010) Making sense by making sentient: effectance motivation increases anthropomorphism. J Pers Soc Psychol 99:410-435.
Wiggett AJ, Pritchard IC, Downing PE (2009) Animate and inanimate objects in human visual cortex: Evidence for task-independent category effects. Neuropsychologia 47:3111-3117.
Wilkes KV (1988) Yishi, duh, um and consciousness. In: Consciousness in Contemporary Science(Marcel, A. J. and Bisiach, E., eds): Oxford University Press.