Paweł Gładziejewski

Shared Representations, Perceptual Symbols, and the Vehicles of Mental Concepts

Abstract: *The main aim of this article is to present and defend a thesis according to which conceptual representations of some types of mental states are encoded in the same neural structures that underlie the first-personal experience of those states. To support this proposal here, I will put forth a novel account of the cognitive function played by 'shared representations' of emotions and bodily sensations, i.e. neural structures that are active when one experiences a mental state of a certain type as well as when one observes someone else experiencing a state of the same type. I will argue that shared representations in fact constitute vehicles of certain mental state concepts (more precisely, concepts of specific types of emotions and somatosensory states). The main line of arguing for this will consist in showing that shared representations exhibit specific, 'conceptual' functional properties: (1) causal effect on forming metacognitive judgments, (2) cognitive penetrability, (3) diversity of input types.*

Keywords: Emotion concepts, empathy, concepts, concept empiricism, concept vehicles, mental concepts, mental simulation, mindreading, mirror neurons, mirror systems, perceptual symbols theory, shared circuits, shared representations.

Correspondence:
Paweł Gładziejewski, Nicolaus Copernicus University, Institute of Philosophy, Szosa Staromiejska 1A, 87–100 Toruñ, Poland Email: pawel_gla@o2.pl

Journal of Consciousness Studies, **20**, No. 3–4, 2013, pp. ??–??

1. Introduction

The present article addresses the problem of how mental concepts are encoded in the brain, whereby concepts are broadly understood as mental representations of categories.[1] One of the crucial distinctions one has to keep in mind when dealing with conceptual representations is the distinction between the content of those representations and their neural vehicles. While the content of a conceptual representation consists in a default, readily available knowledge or set of information about a specific category, the physical (neural) structure that encodes this knowledge or information constitutes the vehicle of this representation. In order to specify the main goal of this article, then, it could be said that I will be defending in it a specific thesis regarding the nature of the *vehicles* of conceptual representations of at least some mental categories. According to the proposal that is presented and defended here, the vehicles of some mental state concepts — more precisely, concepts of certain types of emotions and somatosensory states[2] — are comprised of the same structures that underlie the first-personal experience of those states. Therefore, thinking about at least some types of mental states *conceptually* consists in mentally *simulating*[3] those very states. In order to rationalize this thesis, I will put forth a new theoretical interpretation of so-called 'shared representations' (SRs) of emotions and somatosensory states, i.e. the neural areas that contribute to generating the first-personal experience of some types of emotions and somatosensory states, but are also active when one observes other people experiencing those states. According to the proposal defended here, SRs are best explained by invoking the idea that undergoing certain mental states and representing those states conceptually share — at least partially — the neural substrate.

---

[1] I will therefore be using the term 'concept' as meaning 'conceptual representation' here.

[2] Based on the literature cited in this article, this idea might be applicable to concepts of disgust, fear, pain, and touch. However, there is no way to specify *a priori* the exact range of mental concepts that the main thesis of the present article can be successfully applied to. It seems to be a purely empirical matter.

[3] By 'simulating' I mean recreating or replicating certain (simulated) neural/mental processes by other neural/mental processes. This notion is basically identical to the one used by Goldman (2006), who also construes mental simulation as a sort of replication. More precisely, according to Goldman's definition: a (mental) process P is a simulation of another (mental) process P' = (def) (1) P duplicates, replicates, or resembles P' in some significant respects (significant relative to the purposes or function of the task), and (2) in its (significant) duplication of P', P fulfils one of its purposes or functions (see Goldman, 2006). Note that this definition doesn't specify what function the simulating process performs, only stating that it performs some function by being similar to some other process (see also note 11).

In the second section, I will briefly elaborate on SRs and the empirical data that support their existence. In the third section, I will lay out the reasoning behind the conceptual interpretation of SRs. In the fourth section, I will present some additional and independent theoretical ideas and empirical results that further support the thesis that representing some mental categories conceptually involves simulating them. I will close the paper by sketching some of the perspectives for future research that are inspired by the present proposal.

2. Shared Representations: A Short Characterization

One of the much debated discoveries made by neuroscientists in the last decade has been the discovery of so-called 'shared representations' (SRs) of emotions and somatosensory states.[4] It turns out that some neural structures that underlie first-personal experiences of some types of mental states also show increased activation when representing someone else undergoing states of the same type. According to some authors, the peculiarity of SRs consists in the fact that they are supposed to somehow 'match' first- and third-person perspectives or that they code certain information in a 'we-centric' way, without specifying *who* ('me' or 'someone else') is in a mental state of a given type, and thus they are 'shared' between self and other (Gallese, 2006). For present purposes, one can simply define SRs as neural systems or circuits that underlie *experiencing* a mental state, but that are *also* active when a person *observes a sign* that someone else is experiencing a mental state of the same type (see also Goldman, 2008; Goldman himself uses the term 'mirror systems' rather than 'SRs'). We have reasons to think that neural systems of this kind exist for emotions of disgust (Calder *et al.*, 2000; Jabbi *et al.*, 2008; Wicker *et al.*, 2003), fear (Sprengelmeyer *et al.*, 1999), and anger (Lawrence *et al.*, 2002), as well as for somatosensory states of pain (Jackson *et al.*, 2005; Lamm *et al.*, 2007; Singer and Frith, 2005) and touch (Blakemore *et al.*, 2005; Keysers *et al.*, 2004). The available data show that SR

[4] Note that I am restricting my discussion to SRs in emotional and somatosensory domains. Although the SRs discovered in the domain of *motor action* are perhaps the ones that have received the most attention from the researchers so far (see Rizzolatti and Craighero, 2004; Rizzolatti and Sinigaglia, 2010, for review), I will purposefully omit them when developing the argument for the main thesis of the present article. Generally speaking, this paper concentrates on studying the link between SRs and mindreading, a higher cognitive function that consists (among others) in the ability to attribute mental states to other persons. However, that the SRs discovered in the motor domain actually play a role in mindreading or understanding intentions of others is debatable on both empirical and theoretical grounds (Brass *et al.*, 2007; Jacob, 2008).

activation not only *co-occurs* with observing mental states of others, but also that it plays a crucial role in enabling one to *attribute* those states to others.

One could sort the empirical data that support the existence of SRs into two categories (see also Goldman, 2008). Firstly, there are studies that show that the brain areas active when a person is experiencing an emotion or a somatosensory state of a given type at least partially overlap with those that show increased activation when a person observes someone else experiencing an emotion or a somatosensory state of the same type. For example, (1) Wicker and colleagues (2003) observed that the anterior insula, a brain region activated when participants were exposed to disgusting stimuli, was active when the participants observed static and dynamic facial expressions of disgust; (2) Singer and colleagues (2004) found in that areas underlying the experience of physical pain in its affective aspect (anterior insula, rostral anterior cingulate cortex, brainstem, and cerebellum) were also active when participants received a signal indicating that a painful stimulus is applied to another person. Secondly, there are studies showing the existence of 'paired deficits', as Alvin Goldman (2008) calls them, whereby local brain damage that results in a selective impairment of the ability to experience certain emotion also results in a selective impairment of the ability to recognize this emotion in others. For example, (1) Calder and colleagues (2000) described the case of patient NK, who showed — due to insula damage — a selective impairment of the ability to experience disgust, but whose ability to recognize the disgust of others (from both visual and auditory stimuli) was also decreased; (2) an analogous case, but pertaining to the emotion of fear, was reported by Sprengelmeyer and colleagues (1999), as they described patient NM, who was unable to experience fear due to amygdala damage and who at the same time was deficient at recognizing the fear of others from visual and auditory stimuli.

It is important to explicitly note at this point that the term 'SRs' is employed here in as theoretically neutral a way as possible. It is supposed to *describe* or *pick out* a certain potentially interesting phenomenon — i.e. the existence of structures that get activated both when undergoing certain mental states and when representing them in others — rather than explain it.[5] All that is required to establish that SRs in this neutral sense exist is to show that the same neural/

[5] This is why I choose not to use possibly misleading terms 'mirror neurons' or 'mirror systems', which suggest that SRs are in some sense functionally involved in 'mirroring' the states of others. If the present proposal is on target, then understanding SRs in terms of mirroring is too restrictive and, overall, misguided. So-called 'mirroring' effects observed

representational resources are employed during feeling and observing (or otherwise representing) someone else's state. However, establishing this does not yet say anything about *why* we find SRs (i.e. overlapping activations and paired deficits) during experiments, or about the *function* SR activation has when in 'observation mode'. In other words, there is plenty of room to answer the explanatory question of what SRs actually do; the term 'SRs' refers to an *explanandum* rather than an *explanans*. One might even argue that SRs have no interesting cognitive functions and their existence is a natural by-product of associative properties of specific neural areas. However, in this article I will assume that a more substantive account of SRs is possible. The reason for making this assumption is the existence of paired deficits, which constitutes at least a *prima facie* reason for thinking that SRs actually play some *representational* role — they simply seem important in enabling the subject to represent certain categories of mental states. What needs to be answered then is the question about *how* precisely do SRs represent or enable one to represent mental states or what sorts of representational structures they are.

One of the most prevalent proposals as to how to interpret SRs theoretically is a thesis according to which SRs serve as a part of a larger neural mechanism (but do not constitute this mechanism by themselves) that underlies the ability to empathize with others (see, e.g. de Vignemont and Singer, 2006), whereby empathy is understood as the ability to share emotional states of other people, but while knowing that one's own state is the result of observing someone else. A similar proposal is developed by Goldman (2006; 2008), who interprets SRs from the point of view of his version of the simulation theory of mindreading.[6] Goldman equates SR activity with simulating mental states of others. According to this author, simulation enables one to

---

in laboratories are in fact results of conceptual processes. In other words, there is no 'mirroring' in any substantial or explanatorily interesting sense of the term.

[6] Goldman's theory is in some respects similar to the proposal that is being developed in the present article. However, there also exist some crucial differences between the two. According to Goldman's (2006; 2008) account of what he calls low-level mindreading, SR-based simulation constitutes a *process* by which one 'reads' or gets to know the mental state of some other person. For Goldman, this initial process doesn't involve conceptual knowledge about mental states. Mental concepts come into play only *subsequently*, when the (previously) simulated state is classified as belonging to a certain mental category. In other words, for Goldman, the mental state of the other person is first simulated (mimicked or replicated) and then categorized in light of mental concepts, so that a person can eventually form a belief about the mental state of someone else. However, from the point of view of the present proposal, simulation of a mental state should be equated with tokening of a mental concept. Thus, SRs are more directly involved in *categorizing* mental states and thus enabling one to *access her conceptual knowledge* about certain mental categories. So, although both accounts predict the effect of SR activity on mindreading, they do it for

attribute mental states to other people but does not constitute third-person mental attribution. Full-blown attribution, according to Goldman (2006; 2008), additionally requires an introspectively based classification of the simulated state. Lastly, some authors put forth a bold, albeit rather imprecisely formulated, theory according to which the activity of SRs is responsible for the ability to have a 'non-conceptual', direct experiential access to the mental states of others (see, e.g. Gallese *et al.*, 2008). None of those proposals explains SRs by linking them with strictly *conceptual* abilities.[7] According to each of them, SRs somehow contribute to representing mental states of others. However, none of them is based on or implies a thesis that SRs in fact encode *conceptual* knowledge about certain types of mental states. Therefore, each of those proposals substantially differs from the explanation of SRs that I attempt to develop here.

3. Shared Representations as the Vehicles of Mental State Concepts

The interpretation of SRs that I want to argue for here quite directly links SRs — and in this respect it resembles Goldman's theory — to mindreading, i.e. the ability to interpret, predict, and explain the behaviour of oneself and others in mental terms. The capacity to mindread essentially involves the ability to form metacognitive judgments like 'X believes that p' or 'X desires that p'.[8] What is important for the sake of present discussion is that this latter ability requires one

different reasons. For example, the present account predicts not that the damage of a specific SRs affects the (non-conceptual or non-inferential) process of reading mental states (whose results would be only subsequently interpreted using mental state concepts), but that it robs one of a specific mental concept as such, thereby disturbing the categorization process and making it impossible for someone to make any inferences based on categorical knowledge about a specific mental state. Furthermore, while Goldman's account of SRs ('mirror systems') is restricted to perception-based mindreading (see Goldman 2006; 2008), the present proposal is supposed to apply to conceptual thought (about certain kinds of mental states) *in general*.

[7] It needs to be noted though that Sperber (2004) and Jacob (2009) had already suggested that there is a connection between SRs and mental concepts. The present article may be viewed as an attempt to further develop and support ideas initially formulated by those authors.

[8] Although here I concentrate on first- and third-personal mindreading judgments, my proposal is not restricted to them. If the main thesis of this article is correct, then the activity of SRs is involved in any kind of mental state-related (more specifically, emotion-related or somatosensory-related) conceptual thought. This includes not only judgments or thoughts that attribute mental states to people, but also thoughts that express general categorical knowledge about mental states of specific kinds (e.g. 'People usually avoid pain', 'Smelly things evoke disgust'). Thus, the term 'metacognitive judgments' — as it is employed within the present article — should be interpreted broadly, rather than as being restricted to (first- or third-person) mindreading judgments.

to master concepts of mental states. This follows from a rather uncontroversial assumption that forming a judgment about another person as, for example, believing something to be true requires mastery of the concept BELIEF, the same way that forming judgments about rabbits requires mastery of the concept RABBIT. Although the examples of metacognitive judgments just mentioned involve attributing beliefs and desires, i.e. propositional attitudes, competent mindreaders can just as well attribute to others mental states that are not (at least *prima facie*) propositional attitudes, such as perceptual/sensory or emotional states. In other words, mindreaders are able to represent others not only as believing or desiring something, but also as experiencing pain, tickling, fear, disgust, or hope, and they use these attributions to predict and explain others' behaviour. Mindreading therefore involves not only 'reading' full-blown propositional attitudes, but also states of this latter kind. For this reason, mindreading requires one to master concepts of (somato)sensory and emotional states and the present article is devoted to the problem of what the vehicles of *those* concepts are.[9] All of the functional properties of SRs that will be discussed in this section are somehow related to SRs' involvement in attributing states of this kind, i.e. 'mindreading' those states.

According to the proposal defended here, what explains the properties of neural structures commonly named 'SRs' is that they in fact encode the conceptual knowledge of certain types of sensory and emotional mental states. This idea can be formulated in a different, broader way. If the argumentation that I develop below is valid, then the same neural structures that underlie the first-personal experience of certain types of mental states also serve as the vehicles of *conceptual representations* of those states. For example, the concept of disgust is encoded in the same structures that also underlie first-personal experience of disgust. This means that employing concepts of those states involves *simulating* those states. From this perspective, the fact that neural structures that underlie the first-personal experience of a certain type of mental state are also activated when someone observes someone else experiencing a state of the same kind should be interpreted as indicating that the observer is employing her *conceptual representation* of this state (or, simply, that she is *categorizing* the state of another person). The cases of paired deficits in experiencing and recognizing certain types of mental states should on the other hand be

---

[9] This is why it is so important that empirical studies discussed in Section 2 of this article confirm PST's predictions regarding mental concepts belonging precisely to *this* kind (i.e. concepts of *emotional* states).

interpreted as an evidence for the existence of *semantic* or *conceptual* deficits that result from the inability to *simulate* specific mental states.

Daniel Weiskopf (2007) observes that identifying certain neural structure with a vehicle of a certain concept is always based on the functional properties that this latter structure exhibits. In other words, we come to a conclusion that conceptual knowledge about a given category is encoded in a given neural structure whenever we have reasons to think that this structure realizes appropriate, conceptual *functions*, like categorizing objects as belonging to this category or making inferences about it. This observation is crucially important for the way I want to argue for the thesis that SRs in fact constitute vehicles of (certain) mental concepts. This is because my argumentation rests on the fact that the neural structures that are commonly dubbed 'SRs' exhibit specific *functional properties*. I have three properties in mind here: (i) the fact that the activity of SRs affects the ability to form certain metacognitive judgments; (ii) the fact that SRs are cognitively penetrable, i.e. their activity is modulated by the beliefs and other propositional attitudes that a person has; (iii) the fact that SRs are activated by a wide range of input types (for example, by stimuli from different sensory modalities) pertaining to a given category. I propose that the fact that SRs exhibit these three functional properties constitutes a strong reason to believe that conceptually representing (at least some) mental states is based on simulating those states. In the following part of this section, I will first discuss each of the three properties just mentioned, explain why I think we should expect vehicles of mental concepts to exhibit them, and present empirical data that serve as an evidence that SRs actually possess these properties.

(i) Affecting the Ability to Formulate Metacognitive Judgments

Distinguishing this property as characteristic of concept vehicles is based upon following assumption:

> If a neural structure N serves as a vehicle of a concept of a category C, then selectively damaging N will result in a selective impairment of an ability to represent C conceptually (to form judgments that pertain to C).

This idea should be regarded as rather uncontroversial. Damaging a physical structure that encodes a certain concept should result in, loosely speaking, a loss of this concept. In other words, it should result in a selective impairment of the ability to categorize or make inferences about the category represented by this concept. This sort of assumption is commonplace among psychologists who regard the fact that a local brain lesion results in a selective semantic or conceptual deficit as an evidence that the damaged neural structure encodes conceptual knowledge about a given category (see, e.g. Caramazza and Mahon, 2003; Martin, 2007). This bears rather straightforward consequences for the present discussion. If the 'conceptual' interpretation of SRs is valid, then damaging SRs should result in selective semantic/conceptual impairments. It should result in an inability to (or a reduced ability to) form judgments that are 'built' out of specific mental concepts ('Anne is disgusted by spiders', 'John felt a cold touch on his shoulder', etc.), including the inability to *categorize* and therefore attribute certain types of mental states (for example, to categorize specific types of facial expressions as expressions of disgust). Deficits of this sort should of course be selective relative to the domain of the damaged SRs. For example, damaging SRs in the domain of the emotion of disgust should result in a semantic deficit that affects conceptual processing of disgust and only disgust.

There is plenty of evidence that suggests that disrupting the activity of SRs results in significantly decreasing the ability to form judgments that attribute specific types of mental states to others. As has already been mentioned, one of the main arguments for the very existence of SRs is the existence of paired deficits, whereby damaging the neural substrate of an emotion of a given type results in the inability to *categorize* this type emotion and therefore represent another person as experiencing it. In the previous section I have described studies that show that paired deficits occur in the case of disgust (Calder *et al.*, 2000; see also Hayes *et al.*, 2007) and fear (Sprengelmeyer *et al.*, 1999). Results obtained in other studies reveal that conceptual deficits resulting from damage to SRs can go beyond disrupting the ability to simply categorize or recognize mental states. Hayes and colleagues (2007) found that Huntington's disease is regularly accompanied by a paired deficit in experiencing and recognizing disgust (which is possibly due to disturbances in insula activity in Huntington's patients, see Kipps *et al.*, 2007). However, they also found that patients with Huntington's disease show decreased ability to *list situations* that normally evoke disgust, without showing any signs of having similar problems when it comes to listing situations that evoke other emotions (anger, fear, happiness, sadness, surprise). Yet other studies reveal the existence of a relation between SRs and alexithymia, i.e. the deficiency in recognizing and describing one's *own* emotions. On the one hand, Silani and colleagues (2008) discovered that alexithymia is negatively correlated with activity in the anterior insula when the participants are asked to describe feelings they experience when watching unpleasant pictures. On the other hand, a study conducted by Bird *et al.* (2010) revealed that alexithymia is also related to decreased activation in the anterior insula when the participants observe a painful stimulus being received by another person.

(ii) Cognitive Penetrability

Distinguishing this property stems from following assumption:

> If a neural structure N serves as a vehicle of a concept of some category, then the activity of N should be cognitively penetrable.

The term 'cognitive penetrability' had been introduced by Pylyshyn (1984). Broadly defined, cognitive penetrability of a given mental process consists in its propensity to be modulated by beliefs or other propositional attitudes that a person has. Analogously, a process can be described as cognitively impenetrable when it cannot be affected by propositional attitudes a person has. For example, the famous Müller-Lyer illusion can be interpreted as evidence that perceptual experience is at least to some degree cognitively impenetrable, for even when we *know* that both lines we are seeing are of the same length, we still cannot help but perceive one of them as longer or shorter than the other. How is the notion of cognitive penetrability relevant for the present discussion? It seems natural to suppose that mental processes of a semantic or conceptual sort should be cognitively penetrable, since *how* and even *whether* a person applies a certain concept at a given moment is rather straightforwardly dependent on the content of beliefs and other propositional attitudes this person has at this moment. For example, I will not categorize something as a car when I *believe* that what I see is actually a suggestive dummy of a car. Similarly with metacognitive judgments, including those formed based on 'reading' the minds of others. Thus, we can predict that the activity of neural structures that serve as vehicles of conceptual representations should be cognitively penetrable. For instance, if we interpret SRs as vehicles of specific mental concepts, then we should expect their activation when in 'observation mode' to be cognitively penetrable. This means, among other things, that SRs should not be automatically activated by stimuli of some appropriate kind, but their activity should be somehow dependent on the propositional *knowledge* one has, of course as long as this knowledge is (subjectively) relevant to the application of a given mental concept.

Unfortunately, empirical data collected thus far that can be used as evidence for the cognitive penetrability of SRs are restricted to the domain of physical pain. However, albeit limited, these results seem to be quite conclusive. In a study conducted by Lamm and colleagues (2007), participants were presented with short video clips displaying persons expressing pain that (supposedly) resulted from hearing a painful auditory stimulus through headphones. The participants were informed by researchers that the painful stimulation they watched was a part of a novel therapy that was supposed to cure the persons displayed on the clips from a neurological disease. However, while in some cases the participants were informed that the person they watched was eventually cured, in other cases they received information that the procedure did not succeed and that the person died due to the disease. The experimenters found that the activation of certain brain areas that resulted from watching the painful facial expression of persons that (as the participants *believed*) eventually died was increased compared to the activation that accompanied watching persons for whom the treatment was (once again, as the participants believed) effective. Thus, one can conclude that *beliefs* — in this case, beliefs about the effectiveness of a medical therapy — affect the activity of the SRs of pain.[10]

The results of a study conducted by Singer and colleagues (2006) lend additional support for this conclusion. These authors measured neural activation of pain-related areas when the participants observed a painful stimulus being applied to a person with whom they had previously played a simple economic game. Depending on the instructions given by the experimenter, this competitor had played either in a fair or unfair way. The authors of this study observed that (although this result was obtained only with male and not with female participants) the activation was drastically decreased when the participants were watching pain of an unfair competitor compared to when they were watching pain of a competitor that was judged by them as fair. In other words, the way the (male) participants *judged* the fairness of another person affected the activity of SRs of pain.

One might reasonably argue that cognitive penetrability is a relatively weak evidence for the conceptual nature of a given process. The mere existence of top-down effects on a cognitive/neural process

[10] Importantly, results of this study revealed that beliefs about the effectiveness of the treatment also affected the way the participants rated or judged the intensity and unpleasantness of the perceived pain. Participants tended to rate the pain as more intense and unpleasant when they believed that the treatment was unsuccessful compared to when they thought that it was successful.

should not be regarded as a conclusive reason for thinking that this process is conceptual. There are two answers to this criticism. First, the part of the argument that pertains to cognitive penetrability should only be regarded within a larger context of the argument presented here. In other words, while admittedly inconclusive on its own, the point about cognitive penetrability of SRs at least strengthens the larger argument by being in line with the parts of the argument that concentrate on two other functional properties of SRs. Second, according to the empirical data, what we deal with in the case of SRs for pain is not just *any* top-down effect, but one in which there is a systematic correspondence between how being given specific background information modulates, on the one hand, *pain attribution* and, on the other hand, the level of *activity of SRs of pain*. Two studies revealed the existence of this sort of correlation (Jackson *et al.*, 2005; Lamm *et al.*, 2007). For example, the more intense the pain of another person is, as judged by the observer, the more activation there is in the anterior cingulate cortex (Jackson *et al.*, 2005). Thus, the effects of background information on a conceptual representation (judgment) of perceived pain are actually mirrored by the level of activity of SRs of pain. This is clearly in line with the thesis that the former is encoded in the latter.

(iii) Diversity of Input

The core idea that stands behind taking this property as characteristic of mental state concepts has been elegantly expressed by Sperber:

> Under what condition do we attribute to an organism possession of a concept rather than a mere perceptual discriminatory ability? When the organism is capable of integrating at some level information from different types of source (e.g. different sensory modalities) that pertain to a single type of phenomenon. If, for instance, an animal reacted in unrelated ways to, say, a dog seen, a dog heard barking, or a dog smelled, we would not attribute to it the concept of a dog. If on the other hand all three sources of information converged towards a common reaction to dogs, we would be more inclined to attribute to the animal the concept of a dog… (Sperber, 2004)

Having a concept of some category is essentially related to being able to apply this concept based on many different information types or sources that pertain to this category. For example, for a person to be truthfully described as having a concept CAR, she needs to be able to apply this concept when perceptually categorizing cars using different sensory modalities, when listening to or reading narratives about cars or when some endogenous thought process leads her somehow to bring about her conceptual knowledge about cars. If we apply this simple reasoning to the problem of concept vehicles, we come to following conclusion:

> If a neural structure N serves as a vehicle for a concept of a category C, then N should be activated whenever the concept of C is tokened, regardless of the input type or source that led to this tokening.

From the point of view of the idea that SRs encode mental state concepts, this conclusion can serve as a basis for making important predictions. Some authors have used the notion of SRs to explain how people can recognize emotions by visually perceiving specific facial expressions (Goldman, 2006; Goldman and Sripada, 2005). However, from the perspective of the proposal that is defended in the present article, SR activation should accompany visual *and* auditory perception of expressions of emotional states, as well as occur during reading or hearing *narratives* that pertain to certain mental categories. At the same time, we should expect that disrupting the activity of SRs should produce inability to recognize certain types of mental states not only regardless of the sensory modality used, but also when reading or hearing appropriate narratives.

That SRs are in fact activated by diverse inputs is especially evident when we attend to the research concerning the SRs that exists in the domain of disgust. There exists a wide range of studies showing the existence of paired deficits in experiencing and recognizing disgust. The authors of those studies have probed the occurrence of a disgust recognition impairment using a number of different experimental tasks. Some of them required the participants to recognize different emotions, including disgust, based on visual perception of static as well as dynamic facial expressions (Adolphs *et al.*, 2003; Calder *et al.*, 2000; Hayes *et al.*, 2007; Jabbi *et al.*, 2008; Wicker *et al.*, 2003). Some studies, however, used stimuli of different kinds: Hayes and colleagues (2007) used visual depictions of situations that evoke different emotions (including, of course, disgust) and other authors used vocal expressions (Calder *et al.*, 2000; Jabbi *et al.*, 2008). Moreover, Adolphs and colleagues (2003) and Jabbi and colleagues (2008) found that patients with a paired deficit for experiencing and recognizing disgust are also deficient at recognizing disgust from short *narratives* describing situations and actions that usually evoke this emotion. It might also be worth mentioning a case of patient B, described by Adolphs and colleagues (2003), who was often unable to recognize disgust during real social interactions.

Studies of SRs in different emotional and somatosensory domains bring similar results. In the case of the SRs of pain, we know that there is an overlapping neural activation for first-personally experiencing pain on one hand and, on the other, (1) observing a pain-invoking stimulus being applied to another person (see, e.g. Jackson *et al.*, 2005), (2) observing facial expressions of pain (see e.g. Botvinick *et al.*, 2005), and (3) observing an arbitrary cue indicating that pain will be inflicted on someone else (Singer *et al.*, 2006). The results of a study conducted by Sprengelmeyer and colleagues (1999) suggest that the input that activates the SRs of fear might be similarly diverse. The patient NM that these authors studied showed a paired deficit for experiencing fear and recognizing it from visual depictions of facial and postural expressions, as well as from vocal expressions.

At this point it needs to be stressed once again that the fact that SRs exhibit each of the three properties discussed above *separately* does not by itself constitute a good reason to believe that they encode conceptual representations of certain mental categories. However, I propose that if we take into account that they (as it seems) exhibit all three of them *at the same time*, it makes the conceptual interpretation of SRs very viable. It seems though that there also exist additional and independent reasons — both empirical and theoretical in nature — to hold that the same neural structures that underlie first-personal experience of (at least some) mental states also serve as vehicles of conceptual representations of those states. In the next section I will discuss them in order to, hopefully, persuade some of the sceptics.

4. Perceptual Symbols Theory and Mental Concepts

This above-mentioned additional support for my proposal is twofold. First, the thesis that I am arguing for here is closely related to a relatively novel, yet at the same time well-grounded, account of concept vehicles, namely the perceptual symbols theory (PST). To put it more precisely, if we apply PST to the problem of how mental concepts should be encoded in the brain, what we end up with is a conception that closely resembles the one I am arguing for in this article. Second, there are empirical data that support PST as applied to the problem of emotion concepts. Thus, at the same time, these data also lend additional support for the main points of the present article, including the explanation of SRs proposed in it.

PST has been formulated by Lawrence Barsalou (1999) and was further developed philosophically by Jesse Prinz (2002) soon after, but under the heading of ‘proxytype theory’. For the sake of the present discussion, I will summarize the core ideas of the original (Barsalou) version of this theory, leaving out some of its details.

Broadly speaking, PST is based on a thesis that conceptual representations of categories are stored in sensorimotor systems of the brain that subserve perception and action. In other words, sensorimotor areas of the brain serve as vehicles for *concepts*. Barsalou (1999) expands this general idea using technical notions of 'simulator' and 'simulation'. Every perception of or action performed on an object belonging to a certain category is accompanied by a widespread pattern of neural activity in sensorimotor systems. According to PST, patterns of this kind are partially registered and memorized in associative regions of the brain. At this level, the modally specific information becomes integrated, creating a multimodal profile of the object that includes its visual, tactual, or auditory properties as well as motor sequences that are used to interact with it. According to Barsalou (*ibid.*), since the perceptual and motor patterns of activation that result from interacting with instances of one category are similar, they are also stored in converging areas of the associative cortex. This way, with time emerges a multimodal representation of a *category*, called a 'simulator' (*ibid.*). For example, a simulator for the category of bicycles is a construct in long-term memory that stores information about how bicycles look, what sounds they make, how to ride them, etc. Using this simulator in order to perform cognitive tasks consists in partially reactivating modally specific perceptual and motor states that normally accompany seeing, hearing, or riding bicycles. Barsalou calls this kind of reactivation 'simulation' (*ibid.*).[11] According to his theory, simulations vary from context to context and never embody full categorical knowledge that is stored in simulators. For example,

[11] I assume here that the notion of simulation employed in Barsalou's PST is very close to the one employed by Goldman in his theory of mindreading, which is also the one I am using in this article (see also note 3). This assumption might be considered controversial since Barsalou never explicitly affirms that he accepts Goldman's definition of simulation. However, Barsalou uses the term 'simulation' as indicating re-experience, re-enactment, or reactivation of some perceptual/neural states or processes (see, e.g. Barsalou, 1999; 2009). These modality-specific re-enactments can play conceptual functions by being sufficiently similar to the original states so as to convey relevant categorical information required in conceptual tasks (Barsalou 1999; 2009). Therefore, it seems reasonable to think that simulating essentially means for Barsalou, as for Goldman, *replicating* some mental/neural processes in order to perform new cognitive *functions*. What is present in PST and not in Goldman's theory of mindreading is the thesis that simulation plays *conceptual* functions of categorization and inference. Although this thesis specifies the *function* of simulation, the very *notion* of simulation used in PST seems to be very similar, if not identical, with the one employed by Goldman. Similarly, although Goldman's account of SRs and the account that is advocated here share the basic notion of simulation, they view the functional role simulation plays in mindreading differently (see note 6).

the simulation used to answer the question about whether bicycles have pedals can differ from the one used in order to answer the question about whether bicycles have wheels (see also Barsalou, 2009). PST is based on the postulate that this sort of mechanism, composed of simulators and simulations, can subserve cognitive functions commonly regarded as conceptual. Barsalou (1999) and Prinz (2002) argue at length that this sort of theory can explain such cognitive functions as perceptual categorization or such crowning features of human thought as its compositionality and productivity.

The PST is well grounded empirically.[12] On one hand, it is supported by behavioural data (see e.g. Barsalou *et al.*, 1999; Pecher *et al.*, 2003; Solomon and Barsalou, 2004; Wu and Barsalou, 2009). For example, when participants are asked to list properties that are usually exhibited by the members of some category, they — as PST predicts — most often list properties that should be more available within perceptual simulation (Barsalou *et al.*, 1999). To illustrate, they mention 'red' and 'seeds' as characteristic of watermelons when they are asked to list properties of a 'half watermelon' more often than when they are simply asked to list properties of a 'watermelon'. On the other hand, PST is also supported by neuroimaging data (see e.g. Martin, 2007; Simmons *et al.*, 2007; 2008). For example, Simmons and colleagues (2007) observed neural activity when participants were performing a property verification task, in which they were asked to verify whether a given predicate is true of a given category. They found that verifying colour properties (e.g. whether milk is white) regularly activated the left fusiform gyrus, i.e. the area that is normally involved in processing colour perceptually. This result can be interpreted as evidence that the participants were using perceptual simulation to solve the experimental task.

As has already been stated, it seems reasonable to think that there is a direct connection between PST and the main thesis of this article, namely the thesis that neural systems underlying first-personal experience of some mental states also encode conceptual representations of those states. Applying the former theory to the problem of mental concepts results in formulating an account that closely resembles the one I am arguing for here. Take, for example, an account of emotional concepts (i.e. concepts of emotional states of specific types) that naturally follows from PST. Experiencing emotional and somatosensory states is based upon the activity of a whole range of interoceptive, somato-

[12] For a criticism of this claim, see Machery (2007). See also Prinz (2010) for an answer to this criticism.

sensory, and motor neural systems.[13,14] Such activation patterns can be partially memorized and stored as *simulators* in associative areas of the cortex, which can serve as generators of emotional/somatosensory *simulations* that subserve strictly conceptual functions. In other words, from the point of view of PST, processing mental states conceptually should be construed as simulating those states, i.e. (partially) replicating or recreating them. Thus, the same systems that underlie *first-personal experience* of at least some types of emotions also serve as *vehicles of concepts* of those types of emotions. This way, the 'conceptual' properties of neural systems commonly named 'SRs' which will be discussed at length in third section of this article become perfectly understandable. SR activity routinely observed in experimental settings results from the fact that participants employ their mental concepts in order to *categorize* the mental state of someone else. At the same time, since the simulation process underlies *conceptual* cognitive functions, it is natural that SRs exhibit 'conceptual' properties. In other words, SRs and the properties they have according to the present proposal are quite naturally predicted by PST.[15] Alto-

[13] This idea is of course directly inspired by a broadly Jamesian outlook on the nature of emotions, according to which emotions are constitutively and not merely causally related to perceptions of one's own bodily states. Because of the limited space, it is not possible to defend this theory here (for influential recent defences of this view, see Damasio, 2002; Prinz, 2002). However, it might turn out some weaker version of this theory is true, one according to which bodily perceptions are a crucial ingredient of emotions, but they do not by themselves exhaust emotions. Even if this was the case, we could still expect (in light of PST) conceptual processing of emotions to involve (but not necessarily be exhausted by) interoceptive simulation.

[14] If conceptual processing involves simulating perceptions and actions, then conceptual processing of emotions consists (among others) in simulating *interoceptions of one's own bodily states* (see note 13). For this reason, it is also important to note that neural structures called SRs (like the insula, anterior cingulate cortex, or somatosensory cortex) are in fact involved in perceiving bodily states (the amygdala is the more problematic case, since it is involved in bodily regulation rather than bodily perception).

[15] To clarify this, it must be noted that the idea that there exist SRs for emotions and bodily sensations is *not* a theoretical claim that could be seen as separate from the claims made by PST. The relation between the two should rather be seen as one that holds between an *explanans* (PST) and a description of an *explanandum* (SRs). The existence of overlapping activations (as well as paired deficits) is in need of theoretical explanation. *This* is where PST comes into play, since the simulator-simulation mechanism it postulates naturally explains *why* we find that the same structures that are involved in experiencing mental states are also active when representing those states. PST provides an explanation for SRs, an explanation that goes hand-in-hand with the proposal developed in Section 3 of the present article. It also explains why SRs are structures normally involved in bodily perception (see note 14).

gether, PST and the thesis that SRs in fact constitute vehicles of mental concepts are closely related.[16]

Importantly, some authors have additionally attempted to *directly* verify PST as applied to emotion concepts (Niedenthal *et al.*, 2009; Halberstadt *et al.*, 2009; Vermeulen *et al.*, 2007). These studies are based on the observation that — in line with the main point of the present paper — if we accept PST, we should conclude that:

> …knowledge of an emotion concept is not reducible to an abstract, language-like description, but involves simulation of experienced emotional states relevant to the concept. Although these simulations may not constitute full-blown emotions, and may not even be conscious, they nevertheless can contain enough information about the original states to support conceptual processing. (Niedenthal *et al.*, 2009, p. 1121)

The authors of the study cited above predicted that simulations of emotional states should involve, among others, a *motor* component of emotions — more precisely, they should involve simulation of *facial expressions* associated with emotional states (*ibid.*; see also Halberstadt *et al.*, 2009). One could derive further, more specific, predictions from this general assumption. First, one might predict that performing experimental tasks probing conceptual knowledge of emotions would be accompanied with simulation of emotion-specific facial expressions. Second, one might also expect that disabling this kind of motor simulation would result in participants performing worse in those tasks. A study conducted by Niedenthal and colleagues (2009) confirmed both these predictions. In this study, the authors have used a property verification task, in which the participants were presented with pairs of words and asked to determine whether their designates are associated. These included pairs of a noun phrase or an adjective (some of the words, like ‘vomit’ or ‘delighted’, were emotionally charged, while some, like ‘chair’ or ‘quantified’, were not) and an emotion term (‘anger’, ‘disgust’, or ‘joy’). Activity of facial musculature was measured using electromyography in order verify whether

---

[16] Weiskopf (2007) observes that PST — although he himself uses the term ‘concept empiricism’ — can come in various strengths. He proposes that we distinguish strong global empiricism (‘all thoughts are entirely composed of percepts’) from weak global empiricism (‘all thoughts are partially composed of percepts’), strong local empiricism (‘some thoughts are entirely composed of percepts’) and weak local empiricism (‘some thoughts are partially composed of percepts’). *Prima facie*, nothing seems to stand in the way of saying that the account of mental state concept vehicles that I argue for in this article is compatible with all of those versions of PST (concept empiricism), including the weak local one. After all, the research on SRs does not fully rule out the possibility that concepts of emotions and somatosensory states might be *partially* encoded in a modal way.

performing the experimental task was accompanied by tacit activation of muscles that are normally involved in expressing emotions. The authors found that this was actually the case. Furthermore, these facial simulations (1) were observed only when participants were actually judging the connection between designates of words, and not when they were asked to judge perceptual features of words themselves; (2) were emotion-specific, for example activity of a muscle normally involved in expressing joy (*zygomaticus major*) increased when the participants were verifying pairs of words in which one word designated joy. Lastly, the group also found that temporarily immobilizing musculature responsible for emotion expression impairs the ability to perform the experimental task. Once again, this phenomenon turned out to be emotion-specific. For example, the ability to verify pairs of words that included 'happiness' or 'anger' decreased significantly when muscles for expressing those particular emotions (*zygomaticus major* and *levator labii superioris* respectively) were immobilized. Altogether, these results support the thesis that representing emotions conceptually involves simulating them. Although I will not describe them here, results of other studies point in the same direction (Halberstadt *et al.*, 2009; Vermeulen *et al.*, 2007).

The above-mentioned results are very important from the perspective of the present discussion since they all seem to confirm the main thesis of the present article, albeit on grounds different than examining the properties of SRs. To sum up, when we take in mind the theoretical connection between PST and the present account of SRs, we can say that the probability of the former (in its weaker or stronger versions, see note 16) being true increases the viability of the latter. In other words, the research on PST as applied to emotion concepts makes the 'conceptual' explanation of SRs significantly more credible. At the same time, the present account of SRs seems to bring some additional empirical support for PST itself, especially PST as applied to mental (emotional and somatosensory) concepts. Regardless of what one wishes to focus attention on, what remains is the fact that PST (along with the empirical work that speaks in favour of it) and the present account of SRs support each other.

5. Conclusions

The goal of this article has been to put forth an explanation of shared representations (SRs) of emotions and somatosensory states that differs substantially from those that are currently most prevalent within the literature. According to this alternative proposal defended here, we should explain SRs by adopting a thesis according to which the same structures that underlie experience of certain mental states also constitute the vehicles of concepts of those states, and so thinking about those states consists in mentally simulating them. The main strategy of arguing for such proposal consisted in performing an abductive inference that aimed at providing an explanation for specifically 'conceptual' functional properties exhibited by SRs: (1) the fact that the activity of SRs affects the ability to form metacognitive judgments that pertain to some mental categories; (2) cognitive penetrability of SRs; (3) the diversity of input types or sources that activate SRs. This reasoning has also been supplemented by presenting empirical data that support perceptual symbols theory (PST) as applied to the problem of emotions concepts. As has been shown, these data — as well as PST itself — can be easily construed as providing additional support for the main point of this article.

It seems that treating the proposal defended here seriously might open new perspectives for research on SRs and point it into new, previously unexplored directions. It might be beneficial to close this article by briefly indicating what some of them might be. Firstly, it needs to be conceded that the evidence I have presented here as supporting the idea that SRs exemplify 'conceptual' functional properties was in a certain way selective. For example, the data collected so far that could serve as evidence for the cognitive penetrability of SRs are restricted to the research concerning only one of those systems, namely the one that exists in the domain of physical pain. It would be fruitful, then, to verify whether SRs in *each* emotional or somatosensory domain in fact exhibit *each* of the three functional properties that have been discussed in this article. Secondly, it would be worth asking whether the general idea presented here — i.e. that the neural structures that underlie experience of some types of mental states also encode conceptual knowledge about those types of states — applies to mental categories different than those that have been studied so far. For example, we have reasons to believe that the brain areas that underlie physical pain partially overlap with those that underlie psychic or social pain (Eisenberger and Lieberman, 2004). It would be interesting to verify experimentally whether those same structures are activated when a person *represents* the social/psychic pain of someone else, as well as whether this activation exhibits the three 'conceptual' properties that have been discussed here. Thirdly, the 'conceptual' interpretation of the nature of SRs naturally raises the question of how persons with paired deficits in experiencing and recognizing certain mental states would cope with experimental tasks designed specifically to probe the ability to process these states *semantically*. For example, one might ask whether persons with a paired deficit in experiencing and recognizing disgust would also show selective impairment of the ability to solve a property verification task designed to probe conceptual knowledge of emotions, including disgust. It seems that conducting a series of studies inspired by these three basic ideas could enable us to directly verify the main thesis of this article, rather than by *post hoc* analysis of already existing data. Nonetheless, as I have been attempting to show, the results collected thus far make this thesis perfectly viable, and therefore worthy of further study.

References

Adolphs, A., Tranel, D. & Damasio, A. (2003) Dissociable neural systems for recognizing emotions, *Brain and Cognition*, **52** (1), pp. 61–69.

Barsalou, L.W. (1999) Perceptual symbol systems, *Behavioral and Brain Sciences*, **22**, pp. 577–609.

Barsalou, L.W. (2009) Simulation, situated conceptualization and prediction, *Philosophical Transactions of Royal Society B*, **364**, pp. 1281–1289.

Barsalou, L.W., Solomon, K.O. & Wu, L.L. (1999) Perceptual simulation in conceptual tasks, in Hiraga, M.K., Sinha, C. & Wilcox, S. (eds.) *Cultural, Typological, and Psychological Perspectives in Cognitive Linguistics: The Proceedings of the 4th Conference of the International Cognitive Linguistics Association, Vol. 3*, pp. 209–228, Amsterdam: John Benjamins.

Bastiaansen, J.A.C.J., Thioux, M. & Keysers, C. (2009) Evidence for mirror systems in emotions, *Philosophical Transactions of Royal Society B*, **364**, pp. 2391–2404.

Bird, G., Silani, G., Brindley, R., White, S., Frith, U. & Singer, T. (2010) Empathic brain responses in insula are modulated by levels of alexithymia but not autism, *Brain*, **133**, pp. 1515–1525.

Blakemore, S.J., Bristow, D., Bird, G., Frith, C. & Ward, J. (2005) Somatosensory activations during the observation of touch and a case of vision-touch synaesthesia, *Brain*, **128**, pp. 1571–1583.

Botvinick, M., Jha, A.P., Bylsma, L.M., Fabian, S.A., Solomon, P.E. & Prkachin, K.M. (2005) Viewing facial expressions of pain engages cortical areas involved in the direct experience of pain, *NeuroImage*, **25**, pp. 312–319.

Brass, M., Schmitt, R.M., Spengler, S. & Gergely, G. (2007) Investigating action understanding: Inferential process versus action simulation, *Current Biology*, **17**, pp. 2117–2121.

Calder, A.J., Keane, J., Manes, F., Nagui, A. & Young, A.W. (2000) Impaired recognition and experience of disgust following brain injury, *Nature Neuroscience*, **3**, pp. 1077–1078.

Caramazza, A. & Mahon, B.Z. (2003) The organization of conceptual knowledge: The evidence from category-specific semantic deficits, *Trends in Cognitive Sciences*, **7**, 354–361.

Damasio, A. (2002) *Descartes Error: Emotion, Reason and the Human Brain*, New York: G.P. Putnam.

de Vignemnot, F. & Singer, T. (2006) The empathic brain — when and why, *Trends in Cognitive Sciences*, **10**, pp. 435–441.

Eisenberger, N.I. & Lieberman, M.D. (2004) Why rejection hurts: A common neural alarm system for physical and social pain, *Trends in Cognitive Sciences*, **8**, pp. 294–300.

Gallese, V. (2006) Intentional attunement: A neurophysiological perspective on social cognition and its disruption in autism, *Brain Research*, **1079**, pp. 15–24.

Gallese, V., Keysers, C. & Rizzolatti, G. (2004) A unifying view of social cognition, *Trends in Cognitive Sciences*, **8**, pp. 396–403.

Goldman, A. (2006) *Simulating Minds: The Philosophy, Psychology and Neuroscience of Mindreading*, Oxford: Oxford University Press.

Goldman, A. (2008) Mirroring, mindreading and simulation, in Pineda, J. (ed.) *Mirror Neuron Systems: The Role of Mirroring Processes in Social Cognition*, New York: Humana Press.

Goldman, A. & Sripada, C.S. (2005) Simulationist models of face-based emotion recognition, *Cognition*, **94**, pp. 193–213.

Halberstadt, J., Winkielman, P., Niedenthal, P.M. & Dalle, N. (2009) Emotional conception: How embodied emotion concepts guide perception and facial action, *Psychological Science*, **20**, pp. 1254–1261.

Hayes, C.L., Stevenson, R.J. & Coltheart, M. (2007) Disgust and Huntington's disease, *Neurospychologia*, **45**, pp. 1135–1151.

Jabbi, M., Bastiaansen, J. & Keysers, C. (2008) A common anterior insula representation of disgust observation, experience and imagination shows divergent functional connectivity pathways, *PLOS One*, **3**, e2939.

Jackson, P.L., Meltzoff, A.N. & Decety, J. (2005) How do we perceive the pain of others? A window into the neural processes involved in empathy, *NeuroImage*, **24**, pp. 771–779.

Jacob, P. (2008) What do mirror neurons contribute to human social cognition?, *Mind & Language*, **23**, pp. 190–223.

Jacob, P. (2009) The turning-fork model of social cognition: A critique, *Consciousness and Cognition*, **18**, pp. 229–243.

James, W. (1884) What is an emotion?, *Mind*, **9**, pp. 188–205.

Keysers, C., Wicker, B., Gazzola, V., Anton, J.-L., Fogassi, L. & Gallese, V. (2004) A touching sight: SII/PV activation during the observation of touch, *Neuron*, **42**, pp. 335–346.

Kipps, C.M., Duggins A.J., McCusker, E.A. & Calder, A.J. (2007) Disgust and happiness recognition correlate with anteroventral insula and amygdala volume respectively in preclinical Huntington's disease, *Journal of Cognitive Neuroscience*, **19**, pp. 1206–1217.

Lamm, C., Batson, C.D. & Decety, J. (2007) The neural substrate of human empathy: Effects of perspective-taking and cognitive appraisal, *Journal of Cognitive Neuroscience*, **19**, pp. 42–58.

Lawrence, A.D., Calder, A.J., McGowan, S.M. & Grasby, P.M. (2002) Selective disruption of the recognition of facial expressions of anger, *NeuroReport*, **13**, pp. 881–884.

Machery, E. (2007) Concept empiricism: A methodological critique, *Cognition*, **104**, pp. 19–46.

Martin, A. (2007) The representation of object concepts in the brain, *Annual Review of Psychology*, **58**, pp. 25–45.

Niedenthal, P.M., Winkielman, P., Mondillon, L. & Vermeulen, N. (2009) Embodiment of emotional concepts: Evidence from EMG measures, *Journal of Personality and Social Psychology*, **96**, pp. 1120–1136.

Pecher, D., Zeelenberg, R. & Barsalou, L.W. (2003) Verifying properties from different modalities for concepts produces switching costs, *Psychological Science*, **14**, pp. 119–124.

Prinz, J. (2002) *Furnishing the Mind: Concepts and Their Perceptual Basis*, Cambridge, MA: MIT Press.

Prinz, J. (2006) *Gut Reactions: A Perceptual Theory of Emotion*, Oxford: Oxford University Press.

Prinz, J. (2010) Can concept empiricism forestall concept eliminativism?, *Mind & Language*, **25**, pp. 612–621.

Pylyshyn, Z. (1984) *Computation and Cognition*, Cambridge, MA: MIT Press.

Rizzolatti, G. & Sinigaglia, C. (2010) The functional role of the parieto-frontal mirror circuit: Interpretations and misinterpretations, *Nature Reviews Neuroscience*, **11** (4), pp. 264–274.

Silani, G., Bird, G., Brindley, R., Singer, T., Frith, C. & Fith, U. (2008) Levels of emotional awareness and autism: An fMRI study, *Social Neuroscience*, **3**, pp. 97–112.

Simmons, W.K., Ramjee, V., Beauchamp, M.S., McRae, K., Martin, A. & Barsalou, L.W. (2007) A common neural substrate for perceiving and knowing about color, *Neuropsychologia*, **45**, pp. 2802–2810.

Simmons, W.K., Hamann, S.B., Harenski, C.N., Hu, X.P. & Barsalou, L.W. (2008) fMRI evidence for word association and situated simulation in conceptual processing, *Journal of Physiology — Paris*, **102**, pp. 106–119.

Singer, T., Seymour, B., O'Doherty, J.P., Kaube, H., Dolan, R.J. & Frith, C. (2004) Empathy for pain includes the affective but not sensory aspects of pain, *Science*, **303**, pp. 1157–1162.

Singer, T., Seymour, B., O'Doherty, J.P., Klaas, S.E., Dolan, R.J. & Frith, C. (2006) Empathic neural responses are modulated by the perceived fairness of others, *Nature*, **439**, pp. 466–469.

Singer, T. & Frith, C. (2005) The painful side of empathy, *Nature Neuroscience*, **8**, pp. 845–846.

Solomon, K.O. & Barsalou, L. (2004) Perceptual simulation in property verification, *Memory & Cognition*, **32**, pp. 244–259.

Sperber, D. (2004) 'Mirror neurons' or 'concept neurons'?, [Online], http://www.interdisciplines.org/medias/confs/archives/archive_8.pdf [13 Dec 2011].

Sprengelmeyer, R., Young, A.W., Schroeder, U., Grossenbacher, P.G., Federlein, J., Buttner, T. & Prztunek, H. (1999) Knowing no fear, *Proceedings of the Royal Society B*, **266**, pp. 2451–2456.

Vermeulen, N., Niedenthal, P.M. & Luminet, O. (2007) Switching between sensory and affective systems incurs processing costs, *Cognitive Science*, **31**, pp. 183–192.

Weiskopf, D. (2007) Concept empiricism and the vehicles of thought, *Journal of Consciousness Studies*, **14** (9–10), pp. 156–183.

Wicker, B., Keysers, C., Plailly, J., Royet, J.-P., Gallese, V. & Rizzolatti, G. (2003) Both of us disgusted in my insula: The common neural basis of seeing and feeling disgust, *Neuron*, **40**, pp. 655–664.

Wu, L.L. & Barsalou, L. (2009) Perceptual simulation in conceptual combination: Evidence from property generation, *Acta Psychologica*, **132**, pp. 173–189.

Paper received December 2011; revised May 2012.